



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 1: Introduction to Design of Experiments

Lesson 1: Introduction to Design of Experiments

Introduction

In this course we will pretty much cover the textbook - all of the concepts and designs included. I think we will have plenty of examples to look at and experience to draw from.

Please note: the main topics listed in the syllabus follow the chapters in the book.

A word of advice regarding the analyses. The prerequisite for this course is STAT 501 - Regression and STAT 502 - Analysis of Variance. However, the focus of the course is on the design and not on the analysis. Thus, one can successfully complete this course without these prerequisites, with just STAT 500 - Applied Statistics for instance, but it will require much more work, and for the analysis less appreciation of the subtleties involved. You might say it is more conceptual than it is math oriented.

Learning objectives & outcomes

Upon completion of this lesson, you should be able to do the following:

- understand the issues and principles of Design of Experiments (DOE),
- understand experimentation is a process,
- list the guidelines for designing experiments, and
- recognize the key historical figures in DOE.

Let's start with a simple question...

What is the Scientific Method?

Do you remember learning about this back in high school or junior high even? What were those steps again?

Decide what phenomenon you wish to investigate. Specify how you can manipulate the factor and hold all other conditions fixed, to insure that these extraneous conditions aren't influencing the response you plan to measure.

Then measure your chosen response variable at several (at least two) settings of the factor under study. If changing the factor causes the phenomenon to change, then you conclude that there is indeed a cause-and-effect relationship at work.

How many factors are involved when you do an experiment? Some say two - perhaps this is a comparative experiment? Perhaps there is a treatment group and a control group? If you have a treatment group and a control group then in this case you probably only have one factor with two levels.

How many of you have baked a cake? What are the factors involved to ensure a successful cake? Factors might include preheating the oven, baking time, ingredients, amount of moisture, baking temperature, etc.-- what else? You probably follow a recipe so there are many additional factors that control the ingredients - i.e., a mixture. In other words, someone did the experiment in advance! What parts of the recipe did they vary to make the recipe a success? Probably many factors, temperature and moisture, various ratios of ingredients, and presence or absence of many additives. Now, should one keep all the factors involved in the experiment at a constant level and just vary one to see what would happen? This is a strategy that works but is not very efficient. This is one of the concepts that we will address in this course.

Tasks for this Lesson

- Read Chapter 1 of the textbook
- Read the [Online supplement \[1\]](#) for Chapter 1
- Complete the homework for this lesson.

1.1 - A Quick History of the Design of Experiments (DOE)

The textbook we are using brings an engineering perspective to the design of experiments. We will bring in other contexts and examples from other fields of study including agriculture (where much of the early research was done) education and nutrition. Surprisingly the service industry has begun using design of experiments as well.

"All experiments are designed experiments, it is just that some are poorly designed and some are well-designed."

Engineering Experiments

If we had infinite time and resource budgets there probably wouldn't be a big fuss made over designing experiments. In production and quality control we want to control the error and learn as much as we can about the process or the underlying theory with the resources at hand. From an engineering perspective we're trying to use experimentation for the following purposes:

- reduce **time** to design/develop new products & processes
- improve **performance** of existing processes

- improve **reliability** and performance of products
- achieve product & process **robustness**
- perform **evaluation** of materials, design alternatives, **setting** component & system tolerances, etc.

We always want to fine tune or improve the process. In today's global world this drive for competitiveness affects all of us both as consumers and producers.

Robustness is a concept that enters into statistics at several points. At the analysis stage robustness refers to a technique that isn't overly influenced by bad data. Even if there is an outlier or bad data you still want to get the right answer. Regardless of who or what is involved in the process - it is still going to work. We will come back to this notion of robustness later in the course (Lesson 12).

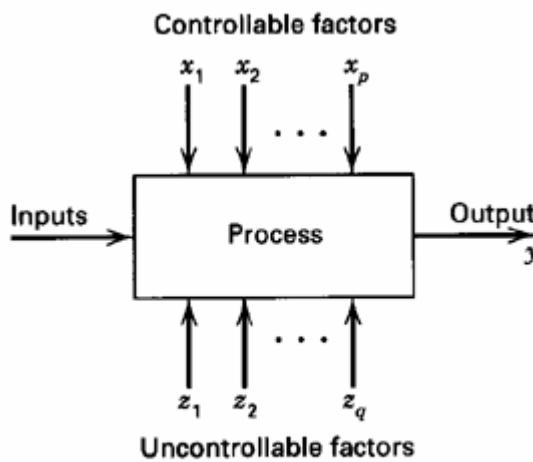


Figure 1-1 General model of a process or system.

Every experiment design has inputs. Back to the cake baking example: we have our ingredients such as flour, sugar, milk, eggs, etc. Regardless of the quality of these ingredients we still want our cake to come out successfully. In every experiment there are inputs and in addition there are factors (such as time of baking, temperature, geometry of the cake pan, etc.), some of which you can control and others that you can't control. The experimenter must think about factors that affect the outcome. We also talk about the output and the yield or the response to your experiment. For the cake, the output might be measured as texture, flavor, height, size, or flavor.

Four Eras in the History of DOE

Here's a quick timeline:

- The agricultural origins, 1918 – 1940s
 - R. A. Fisher & his co-workers
 - Profound impact on agricultural science
 - Factorial designs, ANOVA
- The first industrial era, 1951 – late 1970s
 - Box & Wilson, response surfaces

- Applications in the chemical & process industries
- The second industrial era, late 1970s – 1990
 - Quality improvement initiatives in many companies
 - CQI and TQM were important ideas and became management goals
 - Taguchi and robust parameter design, process robustness
- The modern era, beginning circa 1990, when economic competitiveness and globalization is driving all sectors of the economy to be more competitive.

Notes:

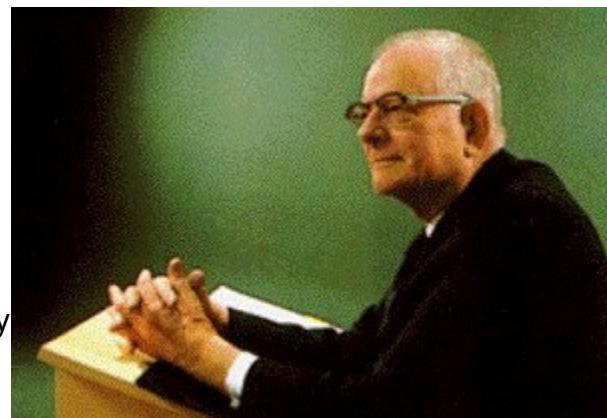
A lot of what we are going to learn in this course goes back to what Sir Ronald Fisher developed in the UK in the first half of the 20th century. He really laid the foundation for statistics and for design of experiments. He and his colleague Frank Yates developed many of the concepts and procedures that we use today. Basic concepts such as orthogonal designs and Latin squares began there in the 20's through the 40's. World War II also had an impact on statistics, inspiring sequential analysis, which arose from World War II as a method to improve the accuracy of long-range artillery guns.

Immediately following World War II the first industrial era marked another resurgence in the use of DOE. It was at this time that Box and Wilson (1951) wrote the key paper in response surface designs thinking of the output as a response function and trying to find the optimum conditions for this function. George Box died early in 2013. And, an interesting fact here - he married Fisher's daughter! He worked in the chemical industry in England in his early career and then came to America and worked at the University of Wisconsin for most of his career.

The Second Industrial Era - or the Quality Revolution

The importance of statistical quality control was taken to Japan in the 1950's by W Edward Deming. This started what Montgomery calls a second Industrial Era, and sometimes the quality revolution. After the second world war Japanese products were of terrible quality.

They were cheaply made and not very good. In the 1960s their quality started improving. The Japanese car industry adopted statistical quality control procedures and conducted experiments which started this new era. Total Quality Management (TQM), Continuous Quality Improvement (CQI) are management techniques that have come out of this statistical quality revolution - statistical quality control and design of experiments.



W. Edwards Deming

Taguchi, a Japanese engineer, discovered and published a lot of the techniques that were later brought to the West, using an independent development of what he referred to as orthogonal arrays. In the West these were referred to as fractional factorial designs. These are both very similar and we will discuss both of these in this course. He came up with the concept of robust parameter design and process robustness.

The Modern Era

Around 1990 Six Sigma, a new way of representing CQI, became popular. Now it is a company and they employ a technique which has been adopted by many of the large manufacturing companies. This is a technique that uses statistics to make decisions based on quality and feedback loops. It incorporates a lot of the previous statistical and management techniques.

Clinical Trials

Montgomery omits in this brief history a major part of design of experimentation that evolved - clinical trials. This evolved in the 1960's when medical advances were previously based on anecdotal data; a doctor would examine six patients and from this wrote a paper and published it. The incredible biases resulting from these kinds of anecdotal studies became known. The outcome was a move toward making the randomized double-blind clinical trial the gold standard for approval of any new product, medical device, or procedure. The scientific application of the statistical procedures became very important.

1.2 - The Basic Principles of DOE

The first three here are perhaps the most important...

Randomization - this is an essential component of any experiment that is going to have validity. If you are doing a comparative experiment where you have two treatments, a treatment and a control for instance, you need to include in your experimental process the assignment of those treatments by some random process. An experiment includes experimental units. You need to have a deliberate process to eliminate potential biases from the conclusions, and random assignment is a critical step.

Replication - is some in sense the heart of all of statistics. To make this point... Remember what the standard error of the mean is? It is the square root of the estimate of the variance of the sample mean, i.e., $\sqrt{\frac{s^2}{n}}$. The width of the confidence interval is determined by this statistic. Our estimates of the mean become less variable as the sample size increases.

Replication is the basic issue behind every method we will use in order to get a handle on how precise our estimates are at the end. We always want to estimate or control the uncertainty in our results. We achieve this estimate through replication. Another way we can achieve short confidence intervals is by reducing the error variance itself. However, when that isn't possible, we can reduce the error in our estimate of the mean by increasing n .

Another way is to reduce the size or the length of the confidence interval is to reduce the error variance - which brings us to blocking.

Blocking - is a technique to include other factors in our experiment which contribute to undesirable variation. Much of the focus in this class will be to creatively use various

blocking techniques to control sources of variation that will reduce error variance. For example, in human studies, the gender of the subjects is often important factor. Age is another factor affecting the response. Age and gender are often considered nuisance factors which contribute to variability and make it difficult to assess systematic effects of a treatment. By using these as blocking factors, you can avoid biases that might occur due to differences between the allocation of subjects to the treatments, and as a way of accounting for some noise in the experiment. We want the unknown error variance at the end of the experiment to be as small as possible. Our goal is usually to find out something about a treatment factor (or a factor of primary interest), but in addition to this we want to include any blocking factors that will explain variation.

Multi-factor Designs - we will spend at least half of this course talking about multi-factor experimental designs: 2^k designs, 3^k designs, response surface designs, etc. The point to all of these multi-factor designs is contrary to the scientific method where everything is held constant except one factor which is varied. The one factor at a time method is a very inefficient way of making scientific advances. It is much better to design an experiment that simultaneously includes combinations of multiple factors that may affect the outcome. Then you learn not only about the primary factors of interest but also about these other factors. These may be blocking factors which deal with nuisance parameters or they may just help you understand the interactions or the relationships between the factors that influence the response.

Confounding - is something that is usually considered bad! Here is an example. Let's say we are doing a medical study with drugs A and B. We put 10 subjects on drug A and 10 on drug B. If we categorize our subjects by gender, how should we allocate our drugs to our subjects? Let's make it easy and say that there are 10 male and 10 female subjects. A balanced way of doing this study would be to put five males on drug A and five males on drug B, five females on drug A and five females on drug B. This is a perfectly balanced experiment such that if there is a difference between male and female at least it will equally influence the results from drug A and the results from drug B.

An alternative scenario might occur if patients were randomly assigned treatments as they came in the door. At the end of the study they might realize that drug A had only been given to the male subjects and drug B was only given to the female subjects. We would call this design totally confounded. This refers to the fact that if you analyze the difference between the average response of the subjects on A and the average response of the subjects on B, this is exactly the same as the average response on males and the average response on females. You would not have any reliable conclusion from this study at all. The difference between the two drugs A and B, might just as well be due to the gender of the subjects, since the two factors are totally confounded.

Confounding is something we typically want to avoid but when we are building complex experiments we sometimes can use confounding to our advantage. We will confound things we are not interested in order to have more efficient experiments for the things we are interested in. This will come up in multiple factor experiments later on. We may be interested in main effects but not interactions so we will confound the interactions in this way in order to reduce the sample size, and thus the cost of the experiment, but still have good information on the main effects.

1.3 - Steps for Planning, Conducting and Analyzing an Experiment

The practical steps needed for planning and conducting an experiment include: recognizing the goal of the experiment, choice of factors, choice of response, choice of the design, analysis and then drawing conclusions. This pretty much covers the steps involved in the scientific method.

1. Recognition and statement of the problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable(s)
4. Choice of design
5. Conducting the experiment
6. Statistical analysis
7. Drawing conclusions, and making recommendations

What this course will deal with primarily is the choice of the design. This focus includes all the related issues about how we handle these factors in conducting our experiments.

Factors

We usually talk about "treatment" factors, which are the factors of primary interest to you. In addition to treatment factors, there are nuisance factors which are not your primary focus, but you have to deal with them. Sometimes these are called blocking factors, mainly because we will try to block on these factors to prevent them from influencing the results.

There are other ways that we can categorize factors:

Experimental vs. Classification Factors

Experimental Factors - these are factors that you can specify (and set the levels) and then assign at random as the treatment to the experimental units. Examples would be temperature, level of an additive fertilizer amount per acre, etc.

Classification Factors - can't be changed or assigned, these come as labels on the experimental units. The age and sex of the participants are classification factors which can't be changed or randomly assigned. But you can select individuals from these groups randomly.

Quantitative vs. Qualitative Factors

Quantitative Factors - you can assign any specified level of a quantitative factor. Examples: percent or pH level of a chemical.

Qualitative Factors - have categories which are different types. Examples might be species of a plant or animal, a brand in the marketing field, gender, - these are not ordered or continuous but are arranged perhaps in sets.

Think About It:

Think about your own field of study and jot down several of the factors that are pertinent in your own research area? Into what categories do these fall?

Get statistical thinking involved early when you are preparing to design an experiment! Getting well into an experiment before you have considered these implications can be disastrous. Think and experiment sequentially. Experimentation is a process where what you know informs the design of the next experiment, and what you learn from it becomes the knowledge base to design the next.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/5>

Links:

[1] [https://bcs.wiley.com/he-bcs/Books?
action=chapter&bcslId=7219&itemId=1118146921&chapterId=79009](https://bcs.wiley.com/he-bcs/Books?action=chapter&bcslId=7219&itemId=1118146921&chapterId=79009)



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 2: Simple Comparative Experiments

Lesson 2: Simple Comparative Experiments

Lesson 2: Introduction

This chapter should be a review for most students who have the required prerequisites. We included it to focus the course and confirm the basics of understanding the assumptions and underpinnings of estimation and hypothesis testing.

Learning objectives & outcomes

Goals for this lesson include the following:

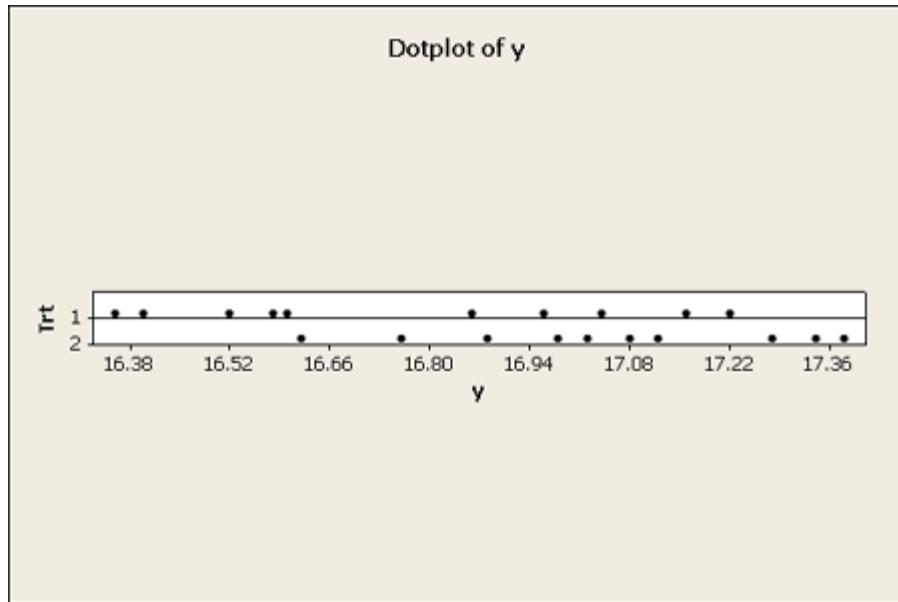
- to review basic statistical concepts
- to review sample size calculation for two sample problems based on the *t*-test
- to review the difference between two independent samples and paired comparison design
- to review the assumptions underlying the *t*-test and how to test for these assumptions

2.1 - Simple Comparative Experiments

Simple comparative experiments are not only preliminary to this course but this takes you back probably into your first course in statistics. We will look at both hypothesis testing and estimation and from these perspectives we will look at sample size determination.

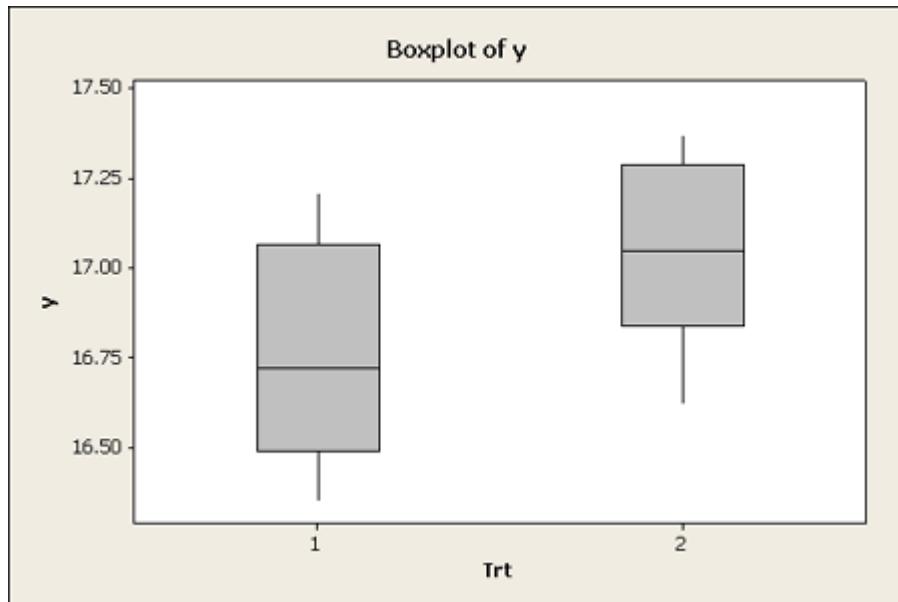
Two Sample Experiment

Here is an example from the text where there are two formulations for making cement mortar. It is hard to get a sense of the data when looking only at a table of numbers. You get a much better understanding of what it is about when looking at a graphical view of the data.



Dot plots work well to get a sense of the distribution. These work especially well for very small sets of data.

Another graphical tool is the boxplot, useful for small or larger data sets. If you look at the box plot you get a quick snapshot of the distribution of the data.



Remember that the box spans the middle 50% of the data (from the 25th to the 75th percentile) and the whiskers extend as far out as the minimum and maximum of the data, to a maximum of 1.5 times the width of the box, or 1.5 times the Interquartile range. So if the data are normal you would expect to see just the box and whisker with no dots outside. Potential outliers will be displayed as single dots beyond the whiskers.

This example is a case where the two groups are different in terms of the median, which is the horizontal line in the box. One cannot be sure simply by visualizing the data if there

is a significant difference between the means of these two groups. However, both the box plots and the dot plot hint at differences.

Testing: The two sample t-test

For the two sample *t*-test both samples are assumed to come from Normal populations with (possibly different) means μ_i and variances σ^2 . When the variances are not equal we will generally try to overcome this by transforming the data. Using a metric where the variation is equal we can use complex ANOVA models, which also assume equal variances. (There is a version of the two sample *t*-test which can handle different variances, but unfortunately this does not extend to more complex ANOVA models.) We want to test the hypothesis that the means μ_i are equal.

Our first look at the data above shows that the means are somewhat different but the variances look to be about the same. We estimate the mean and the sample variance using formulas:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

We divide by $n - 1$ so we can get an unbiased estimate of σ^2 . These are the summary statistics for the two sample problem. If you know the sample size, n , the sample mean, and the sample standard deviation (or the variance), these three quantities for each of the two groups will be sufficient for performing statistical inference. However, it is dangerous to not look at the data and only look at the summary statistics because these summary statistics do not tell you anything about the shape or distribution of the data or about potential outliers, both things you'd want to know about to determine if the assumptions are satisfied.

The two sample *t*-test is basically looking at the difference between the sample means relative to the standard deviation of the difference of the sample means. Engineers would express this as a signal to noise ratio for the difference between the two groups.

If the underlying distributions are normal then the *z*-statistic is the difference between the sample means divided by the true population variance of the sample means. Of course if we do not know the true variances -- we have to estimate them. We therefore use the *t*-distribution and substitute sample quantities for population quantities, which is something we do frequently in statistics. This ratio is an approximate *z*-statistic -- Gosset published the exact distribution under the pseudonym "Student" and the test is often called the "Student *t*" test. If we can assume that the variances are equal, an assumption we will make whenever possible, then we can pool or combine the two sample variances to get the pooled standard deviation shown below.

Our pooled statistic is the pooled standard deviation s_p times the square root of the sum of the inverses of the two sample sizes. The *t*-statistic is a signal-to-noise ratio, a measure of how far apart the means are for determining if they are really different.

Does the data provide evidence that the true means differ? Let's test $H_0: \mu_1 = \mu_2$

We will now calculate the test statistic, which is

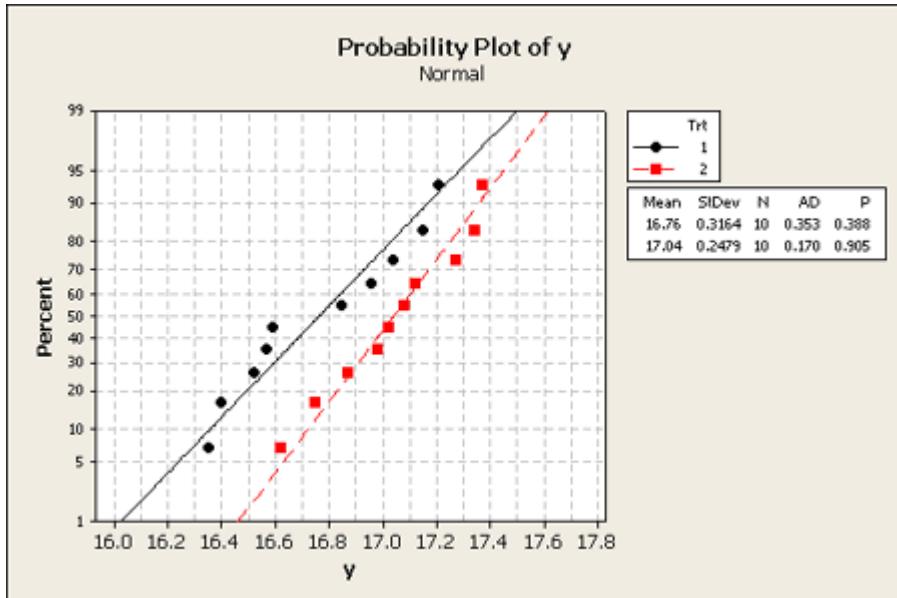
$$t = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

This is always a relative question. Are they different relative to the variation within the groups? Perhaps, they look a bit different. Our t -statistic turns out to be -2.19. If you know the t -distribution, you should then know that this is a borderline value and therefore requires that we examine carefully whether these two samples are really far apart.

We compare the sample t to the distribution with the appropriate $d.f.$. We typically will calculate just the p -value which is the probability of finding the value at least as extreme as the one in our sample. This is under the assumption of the null hypothesis that our means are equal. The p -value in our example is essentially 0.043 as shown in the Minitab output below.

Two-Sample T-Test and CI: y, Trt					
Two-sample T for y					
Trt	N	Mean	StDev	SE Mean	
1	10	16.764	0.316	0.10	
2	10	17.042	0.248	0.078	
 Difference = mu (1) - mu (2)					
Estimate for difference: -0.278					
95% CI for difference: (-0.546, -0.010)					
T-Test of difference = 0 (vs not =): T-Value = -2.19 P-Value = 0.043 DF = 17					

Normal probability plots look reasonable.



Confidence intervals involve finding an interval, in this case the interval is about the difference in means. We want to find upper and lower limits that include the true difference in the means with a specified level of confidence, typically we will use 95%.

In the cases where we have a two-sided hypothesis test which rejects the null hypothesis, then the confidence interval will not contain 0. In our example above we can see in the Minitab output that the 95% confidence interval does not include the value 0, the hypothesized value for the difference, when the null hypothesis assumes the two means are equal.

2.2 - Sample Size Determination

The estimation approach to determining sample size addresses the question: "How accurate do you want your estimate to be?" In this case we are estimating the difference in means. This approach requires us to specify how large a difference we are interested in detecting, say B for the Bound on the margin of error, and then to specify how certain we want to be that we can detect a difference that large. Recall that when we assume equal sample sizes of n , a confidence interval for $\mu_1 - \mu_2$ is given by:

$$\left\{ \bar{Y}_1 - \bar{Y}_2 \pm t(1 - \alpha/2; df) \cdot s \cdot \sqrt{\frac{2}{n}} \right\}$$

Where n is the sample size for each group, and $df = n + n - 2 = 2(n - 1)$ and s is the pooled standard deviation. Therefore, we first specify B and then solve this equation:

$$B = t(1 - \alpha/2; df) \cdot s \cdot \sqrt{\frac{2}{n}}$$

for n . Therefore,

$$n = \left[t(1 - \alpha/2; df) \cdot s \cdot \frac{\sqrt{2}}{B} \right]^2 = \left[\frac{t^2(1 - \alpha/2; df) \cdot s^2 \cdot 2}{B^2} \right]$$

Since in practice, we don't know what s will be, prior to collecting the data, we will need a guesstimate of σ to substitute into this equation. To do this by hand and we use z rather than t since we don't know the df if we don't know the sample size n - the computer will iteratively update the d.f. as it computes the sample size, giving a slightly larger sample size when n is small.

So we need to have an estimate of σ^2 , a desired margin of error bound B , that we want to detect, and a confidence level $1-\alpha$. With this we can determine sample size in this comparative type of experiment. We may or may not have direct control over σ^2 , but by using different experimental designs we do have some control over this and we will address this later in this course. In most cases an estimate of σ^2 is needed in order to determine the sample size.

One special extension of this method is when we have a binomial situation. In this case where we are estimating proportions rather than some quantitative mean level, we know that the worst-case variance, $p(1-p)$, is where p (the true proportion) is equal to 0.5 and then we would have an approximate sample size formula that is simpler, namely $n = 2/B^2$ for $\alpha = 0.05$.

Another Two-Sample Example – Paired Samples

In the paired sample situation, we have a group of subjects where each subject has two measurements taken. For example, blood pressure was measured before and after a treatment was administered for five subjects. These are not independent samples, since for each subject, two measurements are taken, which are typically correlated – hence we call this paired data. If we perform a two sample independent t -test, ignoring the pairing for the moment we lose the benefit of the pairing, and the variability among subjects is part of the error. By using a paired t -test, the analysis is based on the differences (after – before) and thus any variation among subjects is eliminated.

In our Minitab output we show the example with Blood Pressure on five subjects.

2 Sample or Paired Samples? (no sound)



By viewing the output, we see that the different patients' blood pressures seem to vary a lot (standard deviation about 12) but the treatment seems to make a small but consistent difference with each subject. Clearly we have a nuisance factor involved - the subject - which is causing much of this variation. This is a stereotypical situation where because the observations are correlated and paired and we should do a paired t -test.

These results show that by using a paired design and taking into account the pairing of the data we have reduced the variance. Hence our test gives a more powerful conclusion regarding the significance of the difference in means.

The paired *t*-test is our first example of a blocking design. In this context the subject is used as a *block*, and the results from the paired *t*-test are identical to what we will find when we analyze this as a Randomize Complete Block Design from lesson 4.

2.3 - Determining Power

We begin this part by defining the power of a hypothesis test. This also provides another way of determining the sample size. The power is the probability of achieving the desired outcome. What is the desired outcome of a hypothesis test? Usually rejecting the null hypothesis. Therefore, power is the probability of rejecting the null hypothesis when in fact the alternative hypothesis is true.

Decision	H_0	H_A
Reject Null Hypothesis	Type I Error - α	OK
Accept Null Hypothesis	OK	Type II Error - β

Note:

$$P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha: P(\text{Type I Error})$$

$$P(\text{Accept } H_0 \mid H_A \text{ is true}) = \beta: P(\text{Type II Error})$$

Therefore the power of the test is $P(\text{Reject } H_0 \mid H_A \text{ is true}) = 1-\beta$.

Before any experiment is conducted you typically want to know how many observations you will need to run. If you are performing a study to test a hypothesis, for instance in the blood pressure example where we are measuring the efficacy of the blood pressure medication, if the drug is effective there should be a difference in the blood pressure before and after the medication. Therefore we want to reject our null hypothesis, and thus we want the power (i.e. the probability of rejecting the H_0 when it is false) to be as high as possible.

We will describe an approach to determine the power, based on a set of operating characteristic curves traditionally used in determining power for the *t*-test. Power depends on the level of the test, α , the actual true difference in means, and n (the sample size). Figure 2.13 (2.12 in 7th ed) in the text gives the operating characteristic curves where β is calculated for $n^* = 2n - 1$ for an $\alpha = 0.05$ level test. When you design a study you usually plan for equal sample size, since this gives the highest power in your results. We will look at special cases where you might deviate from this but generally this is the case.

To use the Figure in the text, we need to first calculate the difference in means measured in numbers of standard deviation, i.e. $|\mu_1 - \mu_2| / \sigma$. You can think of this as a signal to noise ratio, i.e. how large or strong is the signal, $|\mu_1 - \mu_2|$, in relation to the variation in the measurements, σ . We are not using the symbols in the text, because

the 2 editions define d and δ differently. Different software packages or operating characteristic curves may require either $|\mu_1 - \mu_2| / \sigma$ or $|\mu_1 - \mu_2| / 2\sigma$ to compute sample sizes or estimate power, so you need to be careful in reading the documentation. Minitab avoids this by asking for $|\mu_1 - \mu_2|$ and σ separately, which seems like a very sensible solution.

Again,

Example calculations: Let's consider an example in the two sample situation. We will let $\alpha = .05$, $|\mu_1 - \mu_2| = 8$ (the difference between the two means), and the sigma (assumed true standard deviation) would equal 12, and finally, let the number of observations in each group $n = 5$.

In this case, $|\mu_1 - \mu_2|/\sigma = 8/12 = .66$, and $n^* = 2n - 1 = 9$.

If you look at the Figure you get approximately a β of about 0.9. Therefore, power - or the chance of rejecting the null hypothesis prior to doing the experiment is $1 - \beta$ or $1 - 0.9 = 0.1$ or about ten percent of the time. With such low power we should not even do the experiment!

If we were willing to do a study that would only detect a true difference of, let's say, $|\mu_1 - \mu_2| = 18$ then and n^* would still equal 9, then figure 2-12 the Figure shows that β looks to be about .5 and the power or chance of detecting a difference of 18 is also .5. This is still not very satisfactory since we only have a 50/50 chance of detecting a true difference of 18 even if it exists.

Finally, we calculate the power to detect this difference of 18 if we were to use $n = 10$ observations per group, which gives us $n^* = 19$. For this case $\beta = 0.1$ and thus power = $1 - \beta = 0.9$ or 90%, which is quite satisfactory.

These calculations can also be done in Minitab as shown below. Under the Menu: Stat > Power and Sample Size > 2-sample t, simply input sample sizes, $n = 10$, differences $\delta = 18$, and standard deviation $\sigma = 12$.

Another way to improve power is to use a more efficient procedure - for example if we have paired observations we could use a paired t -test. For instance, if we used the paired t -test, then we would expect to have a much smaller sigma – perhaps somewhere around 2 rather than 12. So, our signal to noise ratio would be larger because the noise component is smaller. We do pay a small price in doing this because our t -test would now have degrees of freedom $n - 1$, instead of $2n - 2$.

The take-home message here is:

If you can reduce variance or noise, then you can achieve an incredible savings in the number of observations you have to collect. Therefore the benefit of a good design is to get a lot more power for the same cost or much decreased cost for the same power.

We now show another approach to calculating power, namely using software tools rather than the graph in Figure 2.12. Let's take a look at how **Minitab** handles this below.

Using Minitab to Calculate Power (no sound)



You can use these dialog boxes to plug in the values that you have assumed and have Minitab calculate the sample size for a specified power, or the power that would result, for a given sample size.

Exercise: Use the assumptions above, and confirm the calculations of power for these values.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/8>



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 3: Experiments with a Single Factor - the Oneway ANOVA - in the Completely Randomized Design (CRD)

Lesson 3: Experiments with a Single Factor - the Oneway ANOVA - in the Completely Randomized Design (CRD)

Lesson 3: Introduction

By the end of this chapter we will understand how to proceed when the ANOVA tells us that the mean responses differ, (i.e., the levels are significantly different), among our treatment levels. We will also briefly discuss the situation that the levels are a random sample from a larger set of possible levels, such as a sample of brands for a product. (Note that this material is in Chapter 3.9 of the 8th edition and Chapter 13.1 of the 7th edition.) We will briefly discuss multiple comparison procedures for qualitative factors, and regression approaches for quantitative factors. These are covered in more detail in the STAT 502 course, and discussed only briefly here.

Learning objectives & outcomes

We focus more on the design and planning aspects of these situations:

- How many observations do we need?
 - to achieve a desired precision when the goal is estimating a parameter, and
 - to achieve a desired level of power when hypothesis testing.
- Which multiple comparison procedure is appropriate for your situation?
- How should we allocate our observations among the k treatment groups? Usually equally, but the Dunnett Test situation has a different optimum allocation.
- The last section describes the F -test as an example of the General Linear Test.

3.1 - Experiments with One Factor and Multiple Levels

Lesson 3 is the beginning of the one-way analysis of variance part of the course, which extends the two sample situation to k samples. In addition to these notes, read Chapter 3 of the text and the on-line supplement. (If you have the 7th edition, also read 13.1.)

We review the issues related to a single factor experiment, which we see in the context of a Completely Randomized Design (CRD). In a single factor experiment with a CRD the levels of the factor are randomly assigned to the experimental units. Alternatively, we can think of randomly assigning the experimental units to the treatments or in some cases, randomly selecting experimental units from each level of the factor.

Example - Cotton Tensile Strength

Let's take a look at an Example, taken from Problem 3.10 of [Montgomery \(308 in the 7th edition\)](#).



This is an investigation into the formulation of synthetic fibers that are used to make cloth. The response is tensile strength, the strength of the fiber. The experimenter wants to determine the best level of the cotton in terms of percent, to achieve the highest tensile strength of the fiber. Therefore, we have a single quantitative factor, the percent of cotton combined with synthetic fabric fibers.



The five treatment levels of percent cotton are evenly spaced from 15% to 35%. We have five replicates, five runs on each of the five cotton weight percentages.

Cotton Weight Percentage	Observations					Total	Average
	1	2	3	4	5		
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						376	15.04

The box plot of the results shows an indication that there is an increase in strength as you increase the cotton and then it seems to drop off rather dramatically after 30%.

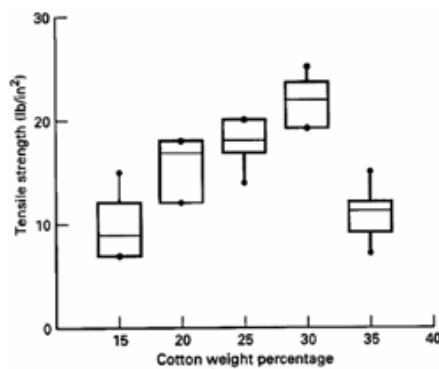


Figure 3-1 Box plots of tensile strength versus cotton weight percentage.

Makes you wonder about all of those 50% cotton shirts that you buy?!

The null hypothesis asks: does the cotton percent make a difference? Now, it seems that it doesn't take statistics to answer this question. All we have to do is look at the side by side box plots of the data and there appears to be a difference – however this difference is not so obvious by looking at the table of raw data. A second question, frequently asked when the factor is quantitative: what is the optimal level of cotton if you only want to consider strength?

There is a point that I probably should emphasize now and repeatedly throughout this course. There is often more than one response measurement that is of interest. You need to think about multiple responses in any given experiment. In this experiment, for some reason, we are interested in only one response, tensile strength, whereas in practice the manufacturer would also consider comfort, ductility, cost, etc.

This single factor experiment can be described as a completely randomized design (CRD). The completely randomized design means there is no structure among the experimental units. There are 25 runs which differ only in the percent cotton, and these will be done in random order. If there were different machines or operators, or other factors such as the order or batches of material, this would

Typesetting Math: 100%

need to be taken into account. We will talk about these kinds of designs later. This is an example of a completely randomized design where there are no other factors that we are interested in other than the treatment factor percentage of cotton.

Analysis of Variance

The Analysis of Variance (ANOVA) is a somewhat misleading name for this procedure. But we call it the analysis of variance because we are partitioning the total variation in the response measurements.

The Model Statement

Each measured response can be written as the overall mean plus the treatment effect plus a random error.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$$i = 1, \dots, a, \quad j = 1, \dots, n_i$$

Generally we will define our treatment effects so that they sum to 0, a constraint on our definition of our parameters, $\sum \tau_i = 0$. This is not the only constraint we could choose, one treatment level could be a reference such as the zero level for cotton and then everything else would be a deviation from that. However, generally we will let the effects sum to 0. The experimental error terms are assumed to be normally distributed, with zero mean and if the experiment has constant variance then there is a single variance parameter σ^2 . All of these assumptions need to be checked. This is called the effects model.

An alternative way to write the model, besides the effects model, where the expected value of our observation, $E(Y_{ij}) = \mu + \tau_i$ or an overall mean plus the treatment effect. This is called the means model and is written as:

$$Y_{ij} = \mu + \epsilon_{ij}$$

$$i = 1, \dots, a, \quad \text{and} \quad j = 1, \dots, n_i.$$

In looking ahead there is also the regression model. Regression models can also be employed but for now we consider the traditional analysis of variance model and focus on the effects of the treatment.

Analysis of variance formulas that you should be familiar with by now are provided in the textbook, (Section 3.3).

The total variation is the sum of the observations minus the overall mean squared, summed over all $a \times n$ observations.

The analysis of variance simply takes this total variation and partitions it into the treatment component and the error component. The treatment component is the difference between the treatment mean and the overall mean. The error component is the difference between the observations and the treatment mean, i.e. the variation not explained by the treatments.

Notice when you square the deviations there are also cross product terms, (see equation 3-5), but these sum to zero when you sum over the set of observations. The analysis of variance is the partition of the total variation into treatment and error components. We want to test the hypothesis that the means are equal versus at least one is different, i.e.

$$H_0: \mu_1 = \dots = \mu_a \text{ versus } H_a: \mu_i \neq \mu_j \text{ for some } i, j$$

Corresponding to the sum of squares (SS) are the degrees of freedom associated with the treatments, $a - 1$, and the degrees of freedom associated with the error, $a \times (n - 1)$, and finally one degree of

Typesetting math: 100%

freedom is due to the overall mean parameter. These add up to the total $N = a \times n$, when the n_i are all equal to n , or $N = \sum n_i$ otherwise.

The mean square treatment (MST) is the sum of squares due to treatment divided by its degrees of freedom.

The mean square error (MSE) is the sum of squares due to error divided by its degrees of freedom.

If the true treatment means are equal to each other, i.e. the μ_i are all equal, then these two quantities should have the same expectation. If they are different then the treatment component, MST will be larger. This is the basis for the F -test.

The basic test statistic for testing the hypothesis that the means are all equal is the F ratio, MST/MSE, with degrees of freedom, $a-1$ and $a \times (n-1)$ or $a-1$ and $N-a$.

We reject H_0 if this quantity is greater than $1-\alpha$ percentile of the F distribution.

Back to the example - Cotton Weight Percent

Here is the Analysis of Variance table from the Minitab output:

One-way ANOVA: Observations versus Cotton Weight %						
Source	DF	SS	MS	F	P	
Cotton Weight %	4	475.76	118.94	14.76	0.000	
Error	20	161.20	8.06			
Total	24	636.96				

S = 2.839	R-Sq = 74.69%	R-Sq(adj) = 69.63%
-----------	---------------	--------------------

Individual 95% CIs For Mean Based on Pooled StDev						
Level	N	Mean	StDev	(-----*-----)	(-----*-----)	(-----*-----)
15	5	9.800	3.347	(-----*-----)		
20	5	15.400	3.130		(-----*-----)	
25	5	17.600	2.074			(-----*-----)
30	5	21.600	2.608			(-----*-----)
35	5	10.800	2.864	(-----*-----)		

10.0	15.0	20.0	25.0
------	------	------	------

Pooled StDev = 2.839

Note a very large F statistic that is, 14.76. The p -value for this F -statistic is $< .0005$ which is taken from an F distribution pictured below with 4 and 20 degrees of freedom.

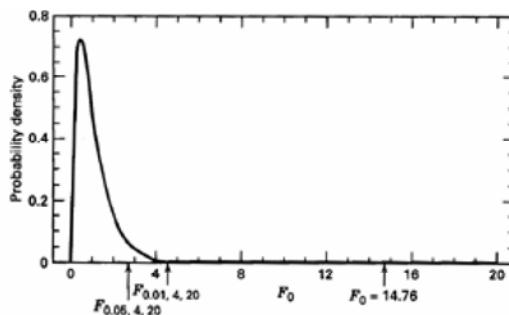


Figure 3-3 The reference distribution ($F_{4,20}$) for the test statistic F_0 in Example 3-1.

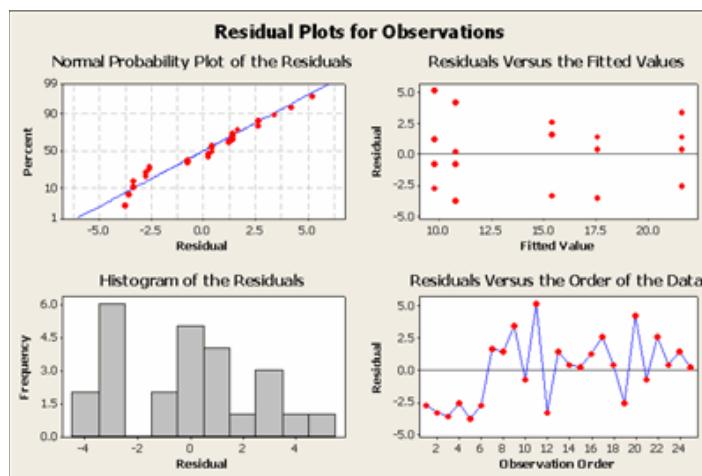
Typesetting math: 100%

We can see that most of the distribution lies between zero and about four. Our statistic, 14.76, is far out in the tail, obvious confirmation about what the data show, that indeed the means are not the same. Hence, we reject the null hypothesis.

Model Assumption Checking

We should check if the data are normal - they should be approximately normal - they should certainly have constant variance among the groups. Independence is harder to check but plotting the residuals in the order in which the operations are done can sometimes detect if there is lack of independence. The question in general is how do we fit the right model to represent the data observed. In this case there's not too much that can go wrong since we only have one factor and it is a completely randomized design. It is hard to argue with this model.

Let's examine the residuals, which are just the observations minus the predicted values, in this case treatment means. Hence, $e_{ij} = y_{ij} - \bar{y}_i$.



These plots don't look exactly normal but at least they don't seem to have any wild outliers. The normal scores plot looks reasonable. The residuals versus the order of the data plot are a plot of the error residuals data in the order in which the observations were taken. This looks a little suspect in that the first six data points all have small negative residuals which are not reflected in the following data points. This looks like it might be a start up problem? These are the kinds of clues that you look for... if you are conducting this experiment you would certainly want to find out what was happening in the beginning.

Post-ANOVA Comparison of Means

So, we found the means are significantly different. Now what? In general, if we had a qualitative factor rather than a quantitative factor we would want to know which means differ from which other ones. We would probably want to do *t*-tests or Tukey maximum range comparisons, or some set of contrasts to examine the differences in means. There are many multiple comparison procedures.

Two methods in particular are Fisher's Least Significant Difference (LSD), and the Bonferroni Method. Both of these are based on the *t*-test. Fisher's LSD says do an *F*-test first and if you reject the null hypothesis, then just do ordinary *t*-tests between all pairs of means. The Bonferroni method is similar, but only requires that you decide in advance how many pairs of means you wish to compare, say *g*, and then perform the *g* *t*-tests with a type I level of α / g . This provides protection for the entire family of *g* tests that the type I error is no more than α . For this setting, with *a* treatments, $g = a(a-1)/2$ when comparing all pairs of treatments.

All of these multiple comparison procedures are simply aimed at interpreting or understanding the overall *F*-test - which means are different? They apply to many situations especially when the factor is

qualitative. However, in this case, since cotton percent is a quantitative factor, doing a test between two arbitrary levels e.g. 15% and 20% level, isn't really what you want to know. What you should focus on is the whole response function as you increase the level of the quantitative factor, cotton percent.

Whenever you have a quantitative factor you should be thinking about modeling that relationship with a regression function.

Review the video that demonstrates the use of polynomial regression to help explain what is going on.

Cotton weight example (no sound)



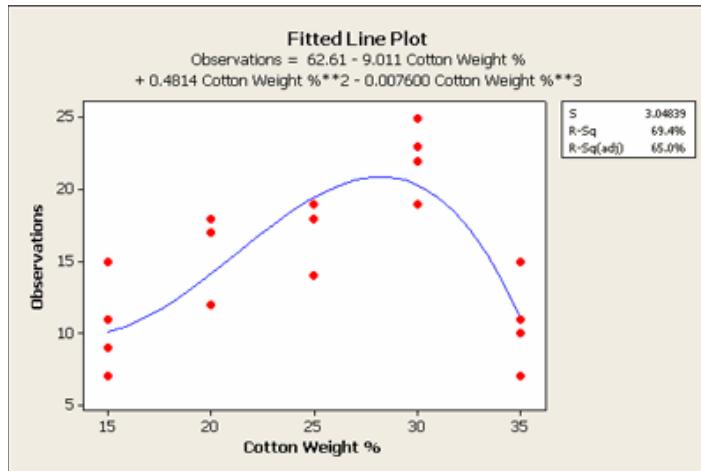
Here is the Minitab output where regression was applied:

Typesetting math: 100%

Polynomial Regression Analysis: Observations versus Cotton Weight %					
The regression equation is					
Observations = 62.61 - 9.011 Cotton Weight % + 0.4814 Cotton Weight %**2 - 0.007600 Cotton Weight %***3					
S = 3.04839 R-Sq = 69.4% R-Sq(adj) = 65.0%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	3	441.814	147.271	15.85	0.000
Error	21	195.146	9.293		
Total	24	636.960			
Sequential Analysis of Variance					
Source	DF	SS	F	P	
Linear	1	33.620	1.28	0.269	
Quadratic	1	343.214	29.03	0.000	
Cubic	1	64.980	6.99	0.015	

Here is a link to the Cotton Weight % dataset ([cotton_weight.MTW](#)). Open this in Minitab so that you can try this yourself.

You can see that the linear term in the regression model is not significant but the quadratic is highly significant. Even the cubic term is significant with p -value = 0.015. In Minitab we can plot this relationship in the fitted line plot as seen below:



This shows the actual fitted equation. Why wasn't the linear term significant? If you just fit a straight line to this data it would be almost flat, not quite but almost. As a result the linear term by itself is not significant. We should still leave it in the polynomial regression model however, because we like to have a hierarchical model when fitting polynomials. What we can learn from this model is that tensile strength of cotton is probably best between the 25 and 30 weight.

This is a more focused conclusion than we get from simply comparing the means of the actual levels in the experiment because the polynomial model reflects the quantitative relationship between the treatment and the response.

We should also check whether the observations have constant variance σ^2 , for all treatments. If they are all equal we can say that they are equal to σ^2 . This is an assumption of the analysis and we need to

Typesetting math: 100%

confirm this assumption. We can either test it with the Bartlett's test, the Levene's test, or simply use the 'eye ball' technique of plotting the residuals versus the fitted values and see if they are roughly equal. The eyeball approach is almost as good as using these tests, since by testing we cannot 'prove' the null hypothesis.

Bartlett's test is very susceptible to non-normality because it is based on the sample variances, which are not robust to outliers. (See Section 3.4 in the text.) We must assume that the data are normally distributed and thus not very long-tailed. When one of the residuals is large and you square it, you get a very large value which explains why the sample variance is not very robust. One or two outliers can cause any particular variance to be very large. Thus simply looking at the data in a box plot is as good as these formal tests. If there is an outlier you can see it. If the distribution has a strange shape you can also see this in a histogram or a box plot. The graphical view is very useful in this regard.

Levene's test is preferred to Bartlett's in my view, because it is more robust. To calculate the Levene's test you take the observations and obtain (not the squared deviations from the mean but) the absolute deviations from the median. Then, you simply do the usual one way ANOVA *F*-test on these absolute deviations from the medians. This is a very clever and simple test that has been around for a long time, created by Levene back in the 1950's. (See 3.4 in the text.) It is much more robust to outliers and non-normality than Bartlett's test.

3.2 - Sample Size Determination

An important aspect of designing an experiment is to know how many observations are needed to make conclusions of sufficient accuracy and with sufficient confidence. We review what we mean by this statement. The sample size needed depends on lots of things; including what type of experiment is being contemplated, how it will be conducted, resources, and desired sensitivity and confidence.

Sensitivity refers to the difference in means that the experimenter wishes to detect, i.e., sensitive enough to detect important differences in the means.

Generally, **increasing** the number of **replications increases** the **sensitivity** and makes it easier to detect small differences in the means. Both power and the margin of error are a function of n and a function of the error variance. Most of this course is about finding techniques to reduce this unexplained residual error variance, and thereby improving the power of hypothesis tests, and reducing the margin of error in estimation.

Hypothesis Testing Approach to Determining Sample Size

Our usual goal is to test the hypothesis that the means are equal, versus the alternative that the means are not equal.

The null hypothesis that the means are all equal implies that the τ_i 's are all equal to 0. Under this framework we want to calculate the power of the *F*-test in the fixed effects case.

Example - Blood Pressure

Consider the situation where we have four treatment groups that will be using four different blood pressure drugs, $a = 4$. We want to be able to detect differences between the mean blood pressure for the subjects after using these drugs.

One possible scenario is that two of the drugs are effective and two are not. e.g. say two of them result in blood pressure at 110 and two of them at 120. In this case the sum of the τ_i^2 for this situation is 100, i.e. $\tau_i = (-5, -5, 5, 5)$ and thus $\sum \tau_i^2 = 100$.

Typesetting math: 100%

Another scenario is the situation where we have one drug at 110, two of them at 115 and one at 120. In this case the sum of the τ_i^2 is 50, i.e. $\tau_i = (-5, 0, 0, 5)$ and thus $\sum \tau_i^2 = 50$.

Considering both of these scenarios, although there is no difference between the minimums and the maximums, the quantities $\sum \tau_i^2$ are very different.

Of the two scenarios, the second is the least favorable configuration (LFC). It is the configuration of means for which you get the least power. The first scenario would be much more favorable. But generally you do not know which situation you are in. The usual approach is to not to try guess exactly what all the values of the τ_i will be but simply to specify δ , which is the maximum difference between the true means, or $\delta = \max(\tau_i) - \min(\tau_i)$.

Going back to our LFC scenario we can calculate this again using $\sum \tau_i^2 = \delta^2/2$, i.e. the maximum difference squared over 2. This is true for the LFC for any number of treatments, since $\sum \tau_i^2 = (\delta/2)^2 \times 2 = \delta^2/2$ since all but the extreme values of τ_i are zero under the LFC.

The Use of Operating Characteristic Curves

The OC curves for the fixed effects model are given in the Appendix V.

The usual way to use these charts is to define the difference in the means, $\delta = \max(\mu_i) - \min(\mu_i)$, that you want to detect, specify the value of σ^2 , and then for the LFC use :

$$\Phi^2 = \frac{n\delta^2}{2a\sigma^2}$$

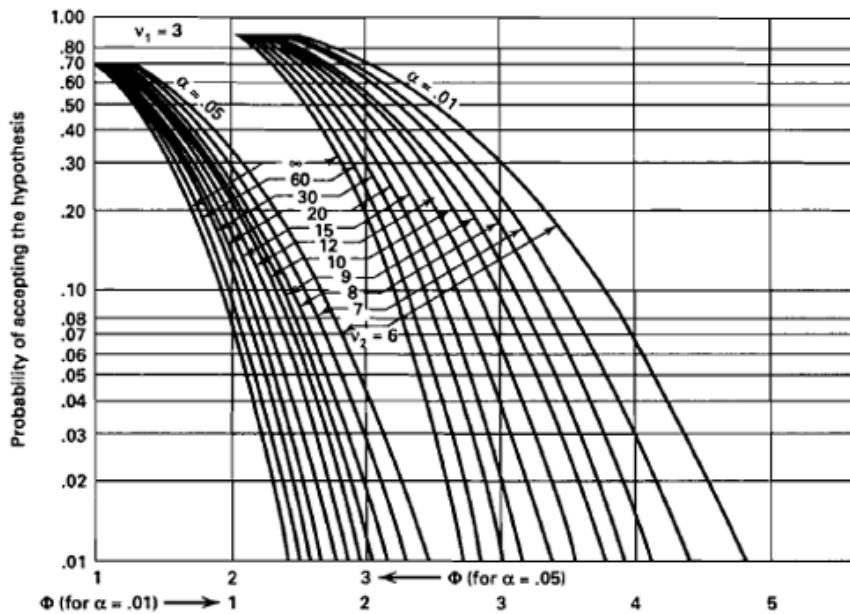
for various values of n . The Appendix V gives β , where $1 - \beta$ is the power for the test where $v_1 = a - 1$ and $v_2 = a(n - 1)$. Thus after setting n , you must calculate v_1 and v_2 to use the table.

Example: We consider an $\alpha = 0.05$ level test for $a = 4$ using $\delta = 10$ and $\sigma^2 = 144$ and we want to find the sample size n to obtain a test with power = 0.9.

Let's guess at what our n is and see how this work. Say we let n be equal to 20, let $\delta = 10$, and $\sigma = 12$ then we can calculate the power using Appendix V. Plugging in these values to find Φ we get $\Phi = 1.3$.

Typesetting math: 100%

V. Operating Characteristic Curves for the Fixed Effects Model Analysis of Variance (continued)



Now go to the chart where v_2 is $80 - 4 = 76$ and $\Phi = 1.3$. This gives us a Type II error of $\beta = 0.45$ and power = $1 - \beta = 0.55$.

It seems that we need a larger sample size.

Well, let's use a sample size of 30. In this case we get $\Phi^2 = 2.604$, so $\Phi = 1.6$.

Now with v_2 a bit more at 116, we have $\beta = 0.30$ and power = 0.70.

So we need a bit more than $n = 30$ per group to achieve a test with power = 0.8.

Review the video below for a 'walk-through' this procedure using Appendix V in the back of the text.

Typesetting math: 100%

Using Power Curve Tables



3.3 - Multiple Comparisons

Scheffé's method for investigating all possible contrasts of the means corresponds exactly to the F -test in the following sense. If the F -test rejects the null hypothesis at level α , then there exists at least one contrast which would be rejected using the Scheffé procedure at level α . Therefore, Scheffé provides α level protection against rejecting the null hypothesis when it is true, regardless of how many contrasts of the means are tested.

Fisher's LSD -- which is the F test, followed by ordinary t -tests among all pairs of means, but only if the F -test rejects the null hypothesis. The F -test provides the overall protection against rejecting H_0 when it is true. The t -tests are each performed at α level and thus likely will reject more than they should, when the F -test rejects. A simple example may explain this statement: assume there are eight treatment groups, and one treatment has a mean higher than the other seven, which all have the same value, and the F -test will rejects H_0 . However, when following up with the pairwise t -tests, the $7 \times 6 / 2 = 21$ pairwise t -tests among the seven means which are all equal, will by chance alone reject at least one pairwise hypothesis, $H_0: \mu_i = \mu_j$ at $\alpha = 0.05$. Despite this drawback Fisher's LSD remains a favorite method since it has overall α level protection, and offers simplicity to understand and interpret.

Bonferroni method for g comparisons – use α / g instead of α for testing each of the g comparisons.

Comparing the Bonferroni Procedure with the Fishers LSD

Fishers's LSD method is an alternative to other pairwise comparison methods (for post ANOVA analysis). This method controls the α -level error rate for each pairwise comparison so it does not control the family error rate. This procedure uses the t statistic for testing $H_0 : \mu_i = \mu_j$ for all i and j pairs.

Typesetting math: 100%

$$t = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSE(\frac{1}{n})}}$$

Alternatively, the Bonferroni method does control the family error rate, by performing the pairwise comparison tests using α/g level of significance, where g is the number of pairwise comparisons. Hence, the Bonferroni confidence intervals for differences of the means are wider than that of Fisher's LSD. In addition, it can be easily shown that the p -value of each pairwise comparison calculated by Bonferroni method is g times the p -value calculated by Fisher's LSD method.

Tukey's Studentized Range considers the differences among all pairs of means divided by the estimated standard deviation of the mean, and compares them with the tabled critical values provided in Appendix VII. Why is it called the studentized range? The denominator uses an estimated standard deviation, hence, the statistic is studentized like the student t -test. The Tukey procedure assumes all n_i are equal say to n .

$$q = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSE(\frac{1}{n})}}$$

Comparing the Tukey Procedure with the Bonferroni Procedure

The Bonferroni procedure is a good all around tool, but for all pairwise comparisons the Tukey studentized range procedure is slightly better as we show here.

The studentized range is the distribution of the difference between the maximum and a minimum over the standard error of the mean. When we calculate a t -test, or when we're using the Bonferroni adjustment where g is the number of comparisons, we are not comparing apples and oranges. In one case (Tukey) the statistic has a denominator with the standard error of a single mean and in the other case (t -test) with the standard error of the difference between means as seen in the equation for t and q above.

Example - Tukey vs. Bonferroni approaches

Here is an example we can work out. Let's say we have 5 means, so $a = 5$, we will let $\alpha = 0.05$, and the total number of observations $N = 35$, so each group has seven observations and $df = 30$.

If we look at the studentized range distribution for 5, 30 degrees of freedom, (the distribution can be found in Appendix VII, p. 630.), we find a critical value of 4.11.

If we took a Bonferroni approach - we would use $g = 5 \times 4 / 2 = 10$ pairwise comparisons since $a = 5$. Thus, again for an $\alpha = 0.05$ test all we need to look at is the t -distribution for $\alpha / 2g = 0.0025$ and $N - a = 30$ df. Looking at the t -table (found in Appendix II, p. 614) we get the value 3.03. However, to compare with the Tukey Studentized Range statistic, we need to multiply the tabled critical value by $\sqrt{2} = 1.414$, therefore $3.03 \times 1.414 = 4.28$, which is slightly larger than the 4.11 obtained for the Tukey table.

The point that we want to make is that the Bonferroni procedure is slightly more conservative than the Tukey result, since the Tukey procedure is exact in this situation whereas Bonferroni only approximate.

The Tukey's procedure is exact for equal samples sizes. However, there is an approximate procedure called the Tukey-Kramer test for unequal n_i .

If you are looking at all pairwise comparisons then Tukey's exact procedure is probably the best procedure to use. The Bonferroni, however, is a good general procedure.

Contrasts of Means

A pairwise comparison is just one example of a contrast of the means. A general contrast can be written as a set of coefficients of the means that sum to zero. This will often involve more than just a pair of treatments. In general we can write a contrast to make any comparison we like. We will also consider sets of orthogonal contrasts.

Example - Gas Mileage

We want to compare the gas mileage on a set of cars: Ford Escape (hybrid), Toyota Camry, Toyota Prius (hybrid), Honda Accord, and the Honda Civic (hybrid). A consumer testing group wants to test each of these cars for gas mileage under certain conditions. They take n prescribed test runs and record the mileage for each vehicle.

Now they first need to define some contrasts among these means. Contrasts are the coefficients which provide a comparison that is meaningful. Then they can test and estimate these contrasts. For the first contrast, C_1 , they could compare the American brand to the foreign brands. We need each contrast to sum to 0, and for convenience only use integers. How about comparing Toyota to Honda (that is C_2), or hybrid compared to non-hybrid (that is C_3).

	Ford Escape	Toyota Camry	Toyota Prius	Honda Accord	Honda Civic
	$Y_1.$	$Y_2.$	$Y_3.$	$Y_4.$	$Y_5.$
C_1	4	-1	-1	-1	-1
C_2	0	-1	-1	1	1
C_3	2	-3	2	-3	2
C_4	0	-1	1	0	0
C_5	0	0	0	-1	1

So the first three contrast coefficients would specify the comparisons described, and the C_4 and C_5 are comparisons within the brands with two models.

After we develop a set of contrasts, we can then test these contrasts or we can estimate them. We can also calculate a confidence intervals around the true contrast of the means by using the estimated contrast \pm the t -distribution times the estimated standard deviation of the contrast. See equation 3-30 in the text.

Concerning Sets of Multiple Contrasts

Scheffé's Method provides α -level protection for all possible contrasts - especially useful when we don't really know how many contrasts we will have in advance. This test is quite conservative, because this test is valid for all possible contrasts of the means. Therefore the Scheffé procedure is equivalent to the F -test, and if the F -test rejects, there will be some contrast that will not contain zero in its confidence interval.

What is an orthogonal contrast?

Two contrasts are orthogonal if the sum of the product of the coefficients of the two contrasts sum to zero. An orthogonal set of contrasts are also orthogonal to the overall mean, since the coefficients sum to zero. See Section 3.5.4 and 3.5.5 of the text.

Type ~~the numbers in the table above~~ and locate which contrasts are orthogonal.

There always exists $a-1$ orthogonal contrasts of a means. When the sample sizes are equal, the sum of squares for these contrasts, when added up, total the sum of squares due to treatment. Any set of orthogonal contrasts partition the variation such that the total variation corresponding to those $a-1$ contrasts equals the total sum of squares among treatments. When the sample sizes are not equal, the definition of orthogonal contrasts involves the sample sizes - this is explained in Section 3.5.5

Dunnett's Procedure

Dunnett's procedure is another multiple comparison procedure specifically designed to compare each treatment to a control. If we have a groups, let the last one be a control group and the first $a-1$ be treatment groups. We want to compare each of these treatment groups to this one control. Therefore, we will have $a-1$ contrasts, or $a-1$ pairwise comparisons. To perform multiple comparisons on these $a-1$ contrasts we use special tables for finding hypothesis test critical values, derived by Dunnett. Section 3.5.8 in the text and compare the test statistics d_i for $i = 1, \dots, a-1$.

Comparing Dunnett's procedure to the Bonferroni procedure

We can compare the Bonferroni approach to the Dunnett procedure. The Dunnett procedure calculates the difference of means for the control versus treatment one, control versus treatment two, etc. to $a-1$. Which provides $a-1$ pairwise comparisons.

So, we now consider an example where we have six groups, $a = 6$, and $t = 5$ and $n = 6$ observations per group. Then, Dunnett's procedure will give the critical point for comparing the difference of means. From the table in the appendix, VIII, we get $\alpha=0.05$ two-sided comparison $d(a-1, f) = 2.66$, where $a-1 = 5$ and $f = df = 30$.

Using the Bonferroni approach, if we look at the t -distribution for $g = 5$ comparisons and a two-sided test with 30 degrees of freedom for error we get 2.75.

Comparing the two, we can see that the Bonferroni approach is a bit more conservative. The Dunnett's is an exact procedure for comparing a control to $a-1$ treatments. Bonferroni is a general tool but not exact. However, there is not much of a difference in this example

Fisher's LSD has the practicality of always using the same measuring stick, the unadjusted t -test. Every one knows that if you do a lot of these tests, that for every 20 tests you do, that one could be wrong by chance. This is another way to handle this uncertainty. All of these methods are protecting you from making too many Type I errors whether you are either doing hypothesis testing or confidence intervals. In your lifetime how many tests are you going to do?

So in a sense you have to ask yourself the question what is the set of tests that I want to protect against making a Type I error. So, in Fisher's LSD procedure each test is standing on its own and is not really a multiple comparisons test. If you are looking for any type of difference and you don't know how many you are going to end up doing, you should probably use Scheffé as to protect you against all of them. But if you know it is all pairwise and that is it, then Tukey's would be best. If you're comparing a bunch of treatments against a control then Dunnett's would be best.

There is a whole family of step-wise procedures which are now available, but we will not consider them here. Each can be shown to be better in certain situations. Another approach to this problem is called False Discovery Rate control. It is used when there are hundreds of hypotheses - a situation that occurs for example in testing gene expression of all genes in an organism, or differences in pixel intensities for pixels in a set of images. The multiple comparisons procedures discussed above all guard against the probability of making one false significant call. But when there are hundreds of tests, we might prefer to make a few false significant calls if it greatly increases our power to detect true difference. False Discovery Rate methods attempt to control the expected percentage of false significant calls among the tests declared significant.

Typesetting math: 100%

3.4 - The Optimum Allocation for the Dunnett Test

The Dunnett test for comparing means is a multiple comparison procedure, but is precisely designed to test t treatments against a control.

We compared the Dunnett test to the Bonferroni - and there was only a slight difference, reflecting the fact that the Bonferroni procedure is an approximation. This is a situation where we have $a = t + 1$ groups; a control group and t treatments.

I like to think of an example where we have a standard therapy, (a control group), and we want to test t new treatments to compare them against the existing acceptable therapy. This is a case where we are not so much interested in comparing each of the treatments against each other, but instead we are interested in finding out whether each of the new treatments are better than the original control treatment.

We have Y_{ij} distributed with mean μ_i , and variance σ^2 , where $i = 1, \dots, t$, and $j = 1, \dots, n_i$ for the t treatment groups and a control group with mean μ_0 with variance σ^2 .

We are assuming equal variance among all treatment groups.

The text describes the Dunnett test in Section 3.5.8.

The question that I want to address here is the design question.

The Dunnett procedure is based on t comparisons for testing H_0 that $\mu_i = \mu_0$, for $i = 1, \dots, t$. This is really t different tests where $t = a - 1$.

The H_a is that the μ_i are not equal to μ_0 .

Or viewing this as an estimation problem, we want to estimate the t differences $\mu_i - \mu_0$.

How Should We Allocate Our Observations?

This is the question we are trying to answer. We have a fixed set of resources and a budget that only allows for only N observations. So, how should we allocate our resources?

Should we assign half to the control group and the rest spread out among the treatments? Or, should we assign an equal number of observations among all treatments and the control? Or what?

We want to answer this question by seeing how we can maximize the power of these tests with the N observations that we have available. We approach this using an estimation approach where we want to estimate the t differences $\mu_i - \mu_0$. Let's estimate the variance of these differences.

What we want to do is minimize the total variance. Remember that the variance of $(\bar{y}_i - \bar{y}_0)$ is $\sigma^2 / n_i + \sigma^2 / n_0$. The total variance is the sum of these t parts.

We need to find n_0 , and n_i that will minimize this total variance. However, this is subject to a constraint, the constraint being that $N = n_0 + (t \times n)$, if the $n_i = n$ for all treatments, an assumption we can reasonably make when all treatments are of equal importance.

Given N observations and a groups, where $a = t + 1$:

Typesetting math: 100%

the model is: $y_{ij} = \mu_i + \varepsilon_{ij}$, where $i = 0, 1, \dots, t$ and $j = 1, \dots, n_i$

sample mean: $\bar{y}_{i\cdot} = \frac{1}{n_i} \sum_j^{n_i} y_{ij}$

and $Var(\bar{y}_{i\cdot}) = \frac{\sigma^2}{n_i}$

Furthermore, $Var(\bar{y}_{i\cdot} - \bar{y}_0) = \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_0}$

Use $\hat{\sigma}^2 = MSE$ and assume $n_i = n$ for $i = 1, \dots, t$.

Then the Total Sample Variance (TSV) = $(TSV) = \sum_{i=1}^t \widehat{var}(\bar{y}_{i\cdot} - \bar{y}_0) = t\left(\frac{\sigma^2}{n} + \frac{\sigma^2}{n_0}\right)$

We want to minimize $t\sigma^2\left(\frac{1}{n} + \frac{1}{n_0}\right)$ where $N = tn + n_0$

This is a LaGrange multiplier problem (calculus): $\min \{TSV + \lambda(N - tn - n_0)\}$:

Solve:

$$1) \frac{\partial(*)}{\partial n} = \frac{-t\sigma^2}{n^2} - \lambda t = 0$$

$$2) \frac{\partial(*)}{\partial n_0} = \frac{-t\sigma^2}{n_0^2} - \lambda = 0$$

From 2) $\lambda = \frac{-t\sigma^2}{n_0^2}$ we can then substitute into 1) as follows:

$$\frac{-t\sigma^2}{n^2} = \lambda t = \frac{-t\sigma^2}{n_0^2} \implies n^2 = \frac{n_0^2}{t} \implies n = \frac{n_0}{\sqrt{t}} \implies n_0 = n\sqrt{t}$$

$$\text{Therefore, from } N = tn + n_0 = tn + \sqrt{tn} = n(t + \sqrt{t}) \implies n = \frac{N}{(t + \sqrt{t})}$$

When this is all worked out we have a nice simple rule to guide our decision about how to allocate our observations:

$$n_0 = n\sqrt{t}$$

Or, the number of observations in the control group should be the square root of the number of treatments times the number of observations in the treatment groups.

If we want to get the exact n based on our resources, let $n = N/(t + \sqrt{t})$ and $n_0 = \sqrt{t} \times n$ and then round to the nearest integers.

Back to our example....

In our example we had $N = 60$ and $t = 4$. Plugging these values into the equation above gives us $n = 10$ and $n_0 = 20$. We should allocate 20 observations in the control and 10 observations in each of the treatments. The purpose is not to compare each of the new drugs to each other but rather to answer whether or not the new drug is better than the control.

Typesetting math: 100%

These calculations demonstrate once again, that the design principles we use in this course are almost always based on trying to minimize the variance and maximizing the power of the experiment. Here is a case where equal allocation is not optimal because you are not interested equally in all comparisons. You are interested in specific comparisons i.e. treatments versus the control, so the control takes on a special importance. In this case we allocate additional observations to the control group for the purpose of minimizing the total variance.

3.5 - One-way Random Effects Models

With quantitative factors, we may want to make inference to levels not measured in the experiment by interpolation or extrapolation on the measurement scale. With categorical factors, we may only be able to use a subset of all possible levels - e.g. brands of popcorn - but we would still like to be able to make inference to other levels. Imagine that we randomly select a of the possible levels of the factor of interest. In this case, we say that the factor is random. As before, the usual single factor ANOVA applies which is

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

However, here both the error term and treatment effects are random variables, that is

$$\varepsilon_{ij} \text{ is } NID(0, \sigma^2) \text{ and } \tau_i \text{ is } NID(0, \sigma_\tau^2)$$

Also, τ_i and ε_{ij} are independent. The variances σ_τ^2 and σ^2 are called **variance components**.

In the fixed effect models we test the equality of the treatment means. However, this is no longer appropriate because treatments are randomly selected and we are interested in the population of treatments rather than any individual one. The appropriate hypothesis test for a random effect is:

$$\begin{aligned} H_0 : \sigma_\tau^2 &= 0 \\ H_1 : \sigma_\tau^2 &> 0 \end{aligned}$$

The standard ANOVA partition of the total sum of squares still works; and leads to the usual ANOVA display. However, as before, the form of the appropriate test statistic depends on the Expected Mean Squares. In this case, the appropriate test statistic would be

$$F_0 = MS_{Treatments} / MS_E$$

which follows an F distribution with $a-1$ and $N-a$ degrees of freedom. Furthermore, we are also interested in estimating the variance components σ_τ^2 and σ^2 . To do so, we use the **analysis of variance method** which consists of equating the expected mean squares to their observed values.

$$\hat{\sigma}^2 = MS_E \text{ and } \hat{\sigma}^2 + n\hat{\sigma}_\tau^2 = MS_{Treatments}$$

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatment} - MS_E}{n}$$

$$\hat{\sigma}^2 = MS_E$$

Potential problem that may arise here is that the estimated treatment variance component may be negative. In such a case, it is proposed to either consider zero in case of a negative estimate or use another method which always results in a positive estimate. A negative estimate for the treatment

Typesetting math: 100%

variance component can also be viewed as evidence that the model is not appropriate, which suggests looking for a better one.

Example 3.11 (13.1 in the 7th ed) discusses a single random factor case about the differences among looms in a textile weaving company. Four looms have been chosen randomly from a population of looms within a weaving shed and four observations were made on each loom. Table 13.1 illustrates the data obtained from the experiment. Here is the Minitab output for this example using **Stat> ANOVA> Balanced ANOVA** command.

Factor	Type	Levels	Values		
Loom	random	4	1 2 3 4		
Analysis of Variance for Y					
Source	DF	SS	MS F P		
Loom	3	89.188	29.729 15.68 0.000		
Error	12	22.750	1.896		
Total	15	111.938			
Source Variance Error Expected Mean Square for Each Term					
component term (using unrestricted model)					
1 Loom	6.958	2	(2) + 4(1)		
2 Error	1.896		(2)		

The interpretation made from the ANOVA table is as before. With the *p*-value equal to 0.000 it is obvious that the looms in the plant are significantly different, or more accurately stated, the variance component among the looms is significantly larger than zero. And confidence intervals can be found for the variance components. The 100(1- α)% confidence interval for σ^2 is

$$\frac{(N - a)MS_E}{\chi_{\alpha/2, N-a}^2} \leq \sigma^2 \leq \frac{(N - a)MS_E}{\chi_{1-\alpha/2, N-a}^2}$$

Confidence intervals for other variance components are provided in the textbook. It should be noted that a closed form expression for the confidence interval on some parameters may not be obtained.

3.6 - The General Linear Test

This is just a general representation of an *F*-test based on a full and a reduced model. We will use this frequently when we look at more complex models.

Let's illustrate the general linear test here for the single factor experiment:

First we write the full model, $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ and then the reduced model, $Y_{ij} = \mu + \varepsilon_{ij}$ where you don't have a τ_i term, you just have an overall mean, μ . This is a pretty degenerate model that just says all the observations are just coming from one group. But the reduced model is equivalent to what we are hypothesizing when we say the μ_i would all be equal, i.e.:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a$$

This is equivalent to our null hypothesis where the τ_i 's are all equal to 0.

Typesetting math: 100%

The reduced model is just another way of stating our hypothesis. But in more complex situations this is not the only reduced model that we can write, there are others we could look at.

The general linear test is stated as an F ratio:

$$F = \frac{(SSE(R) - SSE(F))/(dfR - dfF)}{SSE(F)/dfF}$$

This is a very general test. You can apply any full and reduced model and test whether or not the difference between the full and the reduced model is significant just by looking at the difference in the SSE appropriately. This has an F distribution with $(dfR - dfF)$, dfF degrees of freedom, which correspond to the numerator and the denominator degrees of freedom of this F ratio.

Let's take a look at this general linear test using Minitab...

Example - Cotton Weight

Remember this experiment had treatment levels 15, 20, 25, 30, 35 % cotton weight and the observations were the tensile strength of the material.

The full model allows a different mean for each level of cotton weight %.

We can demonstrate the General Linear Test by viewing the ANOVA table from Minitab:

STAT > ANOVA > Balanced ANOVA



The $SSE(R) = 636.96$ with a $dfR = 24$, and $SSE(F) = 161.20$ with $dfF = 20$. Therefore:

$$F^* = \frac{(636.96 - 161.20)/(24 - 20)}{161.20/20}$$

This demonstrates the equivalence of this test to the F -test. We now use the General Linear Test (GLT) to test for Lack of Fit when fitting a series of polynomial regression models to determine the appropriate degree of polynomial.

We can demonstrate the General Linear Test by comparing the quadratic polynomial model (Reduced model), with the full ANOVA model (Full model). Let $Y_{ij} = \mu + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \varepsilon_{ij}$ be the reduced model, where x_{ij} is the cotton weight percent. Let $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ be the full model.



The viewlet above shows the $SSE(R) = 260.126$ with $dfR = 22$ for the quadratic regression model. The ANOVA shows the full model with $SSE(F) = 161.20$ with $dfF = 20$.

Therefore the GLT is:

Typesetting math: 100%

$$F^* = \frac{(SSE(R) - SSE(F))/(dfR - dfF)}{SSE(F)/dfF}$$

$$= \frac{(260.126 - 161.200)/(22 - 20)}{161.20/20}$$

We reject H_0 : Quadratic Model ~~and claim there is Lack of Fit if $F^* > F_{1-\alpha}(2, 20) = 3.49$~~ .

Therefore, since 6.14 is > 3.49 we ~~reject the null hypothesis of no Lack of Fit from the quadratic equation and fit a cubic polynomial.~~ From the viewlet above we noticed that the cubic term in the equation was indeed significant with ~~p-value = 0.015.~~
~~= 6.14~~

We can apply the General Linear Test again, now testing whether the cubic equation is adequate. The reduced model is:

$$Y_{ij} = \mu + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \varepsilon_{ij}$$

and the full model is the same as before, the full ANOVA model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

The General Linear Test is now a test for Lack of Fit from the cubic model:

$$F^* = \frac{(SSE(R) - SSE(F))/(dfR - dfF)}{SSE(F)/dfF}$$

$$= \frac{(195.146 - 161.200)/(21 - 20)}{161.20/20}$$

$$= \frac{33.95/1}{8.06}$$

$$= 4.21$$

We reject if $F^* > F_{0.95}(1, 20) = 4.35$.

Therefore we do not reject H_a : Lack of Fit and conclude the data are consistent with the cubic regression model, and higher order terms are not necessary.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/12>

Links:

- [1] https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson03/cotton_weight.MTW
- [2] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson03/L03_cotton_weight_viewlet_swf.html', 'l03_cotton_weight', 704, 652\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson03/L03_cotton_weight_viewlet_swf.html', 'l03_cotton_weight', 704, 652);)

Typesetting math: 100%



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 4: Blocking

Lesson 4: Blocking

Introduction

Blocking factors and nuisance factors provide the mechanism for explaining and controlling variation among the experimental units from sources that are not of interest to you and therefore are part of the error or noise aspect of the analysis. Block designs help maintain internal validity, by reducing the possibility that the observed effects are due to a confounding factor, while maintaining external validity by allowing the investigator to use less stringent restrictions on the sampling population.

The single design we looked at so far is the completely randomized design (CRD) where we only have a single factor. In the CRD setting we simply randomly assign the treatments to the available experimental units in our experiment.

When we have a single blocking factor available for our experiment we will try to utilize a randomized complete block design (RCBD). We also consider extensions when more than a single blocking factor exists which takes us to Latin Squares and their generalizations. When we can utilize these ideal designs, which have nice simple structure, the analysis is still very simple, and the designs are quite efficient in terms of power and reducing the error variation.

Learning outcomes & objectives

By the end of this lesson, students are supposed to know

- Concept of Blocking in Design of Experiment
- Dealing with missing data cases in Randomized Complete Block Design
- Application of Latin Square Designs in presence of two nuisance factors
- Application of Graeco-Latin Square Design in presence of three blocking factor sources of variation
- Crossover Designs and their special clinical applications
- Balanced Incomplete Block Designs (BIBD)

References

In this lesson specific references to material in the textbook come from Chapter 4, including:

The Hardness Testing Example, Section 4.1

Vascular Graft Example, Example 4.1

The Latin Square Design, Section 4.2

4.1 - Blocking Scenarios

To compare the results from the RCBD, we take a look at the table below. What we did here was use the one-way analysis of variance instead of the two-way to illustrate what might have occurred if we had not blocked, if we had ignored the variation due to the different specimens. **Blocking** is a technique for dealing with nuisance factors.

[Learn More About Blocking](#)

A **nuisance** factor is a factor that has some effect on the response, but is of no interest to the experimenter; however, the variability it transmits to the response needs to be minimized or explained. We will talk about treatment factors, which we are interested in, and blocking factors, which we are not interested in. We will try to account for these nuisance factors in our model and analysis.

Typical nuisance factors include *batches* of raw material if you are in a production situation, different *operators*, nurses or subjects in studies, the *pieces* of test equipment, when studying a process, and *time* (shifts, days, etc.) where the time of the day or the shift can be a factor that influences the response.

Many industrial and human subjects experiments involve blocking, or when they do not, probably should in order to reduce the unexplained variation.

Where does the term *block* come from? The original use of the term block for removing a source of variation comes from agriculture. Given that you have a plot of land and you want to do an experiment on crops, for instance perhaps testing different varieties or different levels of fertilizer, you would take a section of land and divide it into plots and assigned your treatments at random to these plots. If the section of land contains a large number of plots, they will tend to be very variable - heterogeneous.

A block is characterized by a set of homogeneous plots or a set of similar experimental units. In agriculture a typical block is a set of contiguous plots of land under the assumption that fertility, moisture, weather, will all be similar, and thus the plots are homogeneous.

Failure to block is a common flaw in designing an experiment. Can you think of the consequences?

If the nuisance variable is **known** and **controllable**, we use **blocking** and control it by including a blocking factor in our experiment.

If you have a nuisance factor that is **known** but **uncontrollable**, sometimes we can use **analysis of covariance** (see Chapter 15) to measure and remove the effect of the nuisance factor from the analysis. In that case we adjust statistically to account for a covariate, whereas in blocking, we design the experiment with a block factor as an essential component of the design. Which do you think is preferable?

Many times there are nuisance factors that are **unknown** and **uncontrollable** (sometimes called a “lurking” variable). We use **randomization** to balance out their impact. We always randomize so that every experimental unit has an equal chance of being assigned to a given treatment. Randomization is our insurance against a systematic bias due to a nuisance factor.

Sometimes several sources of variation are **combined** to define the block, so the block becomes an aggregate variable. Consider a scenario where we want to test various subjects with different treatments.

Age classes and gender

In studies involving human subjects, we often use gender and age classes as the blocking factors. We could simply divide our subjects into age classes, however this does not consider gender. Therefore we partition our subjects by gender and from there into age classes. Thus we have a block of subjects that is defined by the combination of factors, gender and age class.

Institution (size, location, type, etc)

Often in medical studies, the blocking factor used is the type of institution. This provides a very useful blocking factor, hopefully removing institutionally related factors such as size of the institution, types of populations served, hospitals versus clinics, etc., that would influence the overall results of the experiment.

The Hardness Testing Example

Loading [MathJax]/extensions/MathMenu.js

In this example we wish to determine whether 4 different tips (the treatment factor) produce different (mean) hardness readings on a Rockwell hardness tester. The treatment factor is the design of the tip for the machine that determines the hardness of metal. The tip is one component of the testing machine.

To conduct this experiment we assign the tips to an **experimental unit**; that is, to a test specimen (called a coupon), which is a piece of metal on which the tip is tested.

If the structure were a completely randomized experiment (CRD) that we discussed in lesson 3, we would assign the tips to a random piece of metal for each test. In this case, the test specimens would be considered a source of **nuisance variability**. If we conduct this as a blocked experiment, we would assign all four tips to the same test specimen, randomly assigned to be tested on a different location on the specimen. Since each treatment occurs once in each block, the number of test specimens is the number of replicates.

Back to the hardness testing example, the experimenter may very well want to test the tips across specimens of various hardness levels. This shows the importance of blocking. To conduct this experiment as a RCBD, we assign all 4 tips to each specimen.

In this experiment, each specimen is called a “**block**”; thus, we have designed a more homogenous set of experimental units on which to test the tips.

Variability **between** blocks can be large, since we will remove this source of variability, whereas variability **within** a block should be relatively small. In general, a **block** is a specific level of the nuisance factor.

Another way to think about this is that a complete replicate of the basic experiment is conducted in each block. In this case, a block represents an experimental-wide **restriction on randomization**. However, experimental runs **within** a block are **randomized**.

Suppose that we use $b = 4$ blocks as shown in the table below:

Table 4-1 Randomized Complete Block Design for the Hardness Testing Experiment

Test Coupon (Block)			
1	2	3	4
Tip 3	Tip 3	Tip 2	Tip 1
Tip 1	Tip 4	Tip 1	Tip 4
Tip 4	Tip 2	Tip 3	Tip 2
Tip 2	Tip 1	Tip 4	Tip 3

Notice the **two-way structure** of the experiment. Here we have four blocks and within each of these blocks is a random assignment of the tips within each block.

We are primarily interested in testing the equality of treatment means, but now we have the ability to remove the variability associated with the nuisance factor (the blocks) through the grouping of the experimental units prior to having assigned the treatments.

The ANOVA for Randomized Complete Block Design (RCBD)

In the RCBD we have one run of each treatment in each block. In some disciplines each block is called an experiment (because a copy of the entire experiment is in the block) but in statistics we call the block to be a replicate. This is a matter of scientific jargon, the design and analysis of the study is an RCBD in both cases.

Suppose that there are a treatments (factor levels) and b blocks.

Leading Math/Stat Extensions/Mathematics

A **statistical model** (effects model) for the RCBD is:

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

This is just an extension of the model we had in the one-way case. We have for each observation Y_{ij} an additive model with an overall mean, plus an effect due to treatment, plus an effect due to block, plus error.

The relevant (fixed effects) hypothesis for the treatment effect is:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a \quad \text{where} \quad \mu_i = (1/b) \sum_{j=1}^b (\mu + \tau_i + \beta_j) = \mu + \tau_i$$

$$\text{if } \sum_{j=1}^b \beta_j = 0$$

We make the assumption that the errors are independent and normally distributed with constant variance σ^2 .

The ANOVA is just a partitioning of the variation:

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 &= \sum_{i=1}^a \sum_{j=1}^b [(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &\quad + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})]^2 \\ &= b \sum_{i=1}^a a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^b b (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned}$$

$$SS_T = SS_{Treatments} + SS_{Blocks} + SS_E$$

The algebra of the sum of squares falls out in this way. We can partition the effects into three parts: sum of squares due to treatments, sum of squares due to the blocks and the sum of squares due to error.

The degrees of freedom for the sums of squares in:

$$SS_T = SS_{Treatments} + SS_{Blocks} + SS_E$$

are as follows for a treatments and b blocks:

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

The partitioning of the variation of the sum of squares and the corresponding partitioning of the degrees of freedom provides the basis for our orthogonal analysis of variance.

Loading [MathJax]/extensions/MathMenu.js

ANOVA Display for the RCBD

Table 4-2 Analysis of Variance for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{Treatments}$	$a - 1$	$\frac{SS_{Treatments}}{a - 1}$	$\frac{MS_{Treatments}}{MS_E}$
Blocks	SS_{Blocks}	$b - 1$	$\frac{SS_{Blocks}}{b - 1}$	
Error	SS_E	$(a - 1)(b - 1)$	$\frac{SS_E}{(a - 1)(b - 1)}$	
Total	SS_T	$N - 1$		

In Table 4.2 we have the sum of squares due to treatment, the sum of squares due to blocks, and the sum of squares due to error. The degrees of freedom add up to a total of $N-1$, where $N = ab$. We obtain the Mean Square values by dividing the sum of squares by the degrees of freedom.

Then, under the null hypothesis of no treatment effect the ratio of the mean square for treatments to the error mean square is an F statistic that is used to test the hypothesis of equal treatment means.

The text provides manual computing formulas; however, we will use Minitab to analyze the RCBD.

Back to the Tip Hardness example:

Remember, the hardness of specimens (coupons) is tested with 4 different tips.

Note: tips are the treatment factor levels, and the coupons are the block levels, composed of homogeneous specimens.

Here is the data for this experiment ([tip_hardness.txt](#) [1]):

Obs	Tip	Hardness	Coupon
1	1	9.3	1
2	1	9.4	2
3	1	9.6	3
4	1	10.0	4
5	2	9.4	1
6	2	9.3	2
7	2	9.8	3
8	2	9.9	4
9	3	9.2	1
10	3	9.4	2
11	3	9.5	3
12	3	9.7	4
13	4	9.7	1
14	4	9.6	2
15	4	10.0	3
16	4	10.2	4

Here is the output from Minitab. We can see four levels of the Tip and four levels for Coupon:

The Analysis of Variance table shows three degrees of freedom for Tip three for Coupon, and the error degrees of freedom is nine. The ratio of mean squares of treatment over error gives us an *F* ratio that is equal to 14.44 which is highly significant since it is greater than the .001 percentile of the *F* distribution with three and nine degrees of freedom.

ANOVA: Hardness versus Tip, Coupon					
Factor	Type	Levels	Values		
Tip	fixed	4	1, 2, 3, 4		
Coupon	fixed	4	1, 2, 3, 4		
Analysis of Variance for Hardness					
Source	DF	SS	MS	F	P
Tip	3	0.38500	0.12833	14.44	0.001
Coupon	3	0.82500	0.27500	30.94	0.000
Error	9	0.08000	0.00889		
Total	15	1.29000			
 S = 0.0942809 R-Sq = 93.80% R-Sq(adj) = 89.66%					

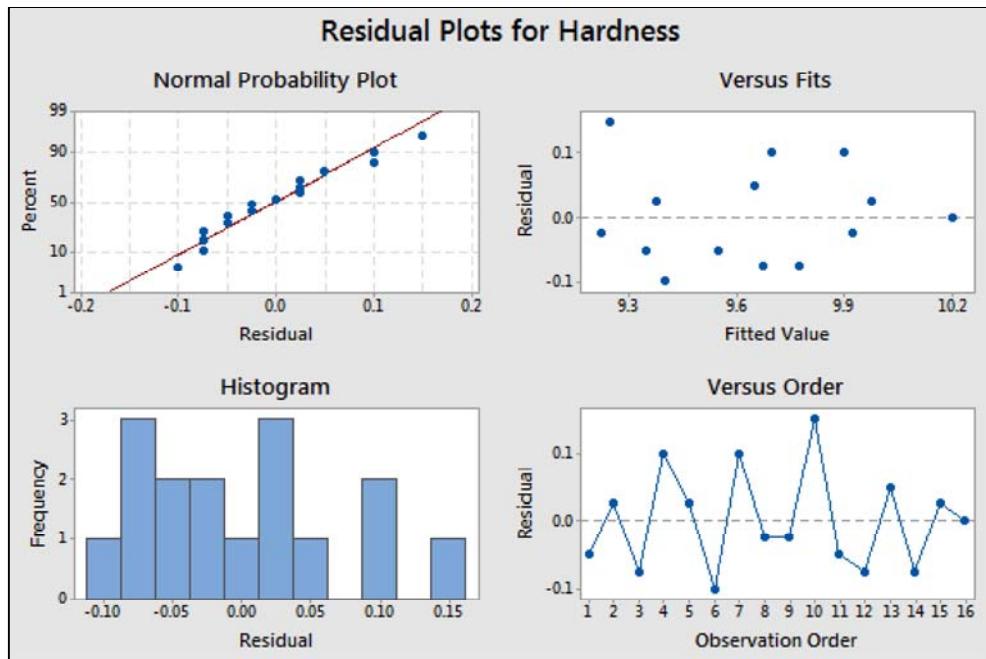
Our 2-way analysis also provides a test for the block factor, Coupon. The ANOVA shows that this factor is also significant with an *F*-test = 30.94. So, there is a large amount of variation in hardness between the pieces of metal. This is why we used specimen (or coupon) as our blocking factor. We expected in advance that it would account for a large amount of variation. By including block in the model and in the analysis, we removed this large portion of the variation, such that the residual error is quite small. By including a block factor in the model, the error variance is reduced, and the test on treatments is more powerful.

The test on the block factor is typically not of interest except to confirm that you used a good blocking factor. The results are summarized by the table of means given below.

Means

Tip	N	Hardness	Coupon	N	Hardness
1	4	9.5750	1	4	9.4000
2	4	9.6000	2	4	9.4250
3	4	9.4500	3	4	9.7250
4	4	9.8750	4	4	9.9500

Here is the residual analysis from the two-way structure.



Comparing the CRD to the RCBD

To compare the results from the RCBD, we take a look at the table below. What we did here was use the one-way analysis of variance instead of the two-way to illustrate what might have occurred if we had not blocked, if we had ignored the variation due to the different specimens.

ANOVA: Hardness versus Tip					
Factor	Type	Levels	Values		
Tip	fixed	4	1, 2, 3, 4		
Analysis of Variance for Hardness					
Source	DF	SS	MS	F	P
Tip	3	0.38500	0.12833	1.70	0.220
Error	12	0.90500	0.07542		
Total	15	1.29000			
 $S = 0.274621$ $R-Sq = 29.84\%$ $R-Sq(adj) = 12.31\%$					

This isn't quite fair because we did in fact block, but putting the data into one-way analysis we see the same variation due to tip, which is 3.85. So we are explaining the same amount of variation due to the tip. That has not changed. But now we have 12 degrees of freedom for error because we have not blocked and the sum of squares for error is much larger than it was before, thus our F -test is 1.7. If we hadn't blocked the experiment our error would be much larger and in fact we would not even show a significant difference among these tips. This provides a good illustration of the benefit of blocking to reduce error. Notice that the standard deviation, $S = \sqrt{MSE}$, would be about three times larger if we had not blocked.

Loading [MathJax]/extensions/MathMenu.js

Other Aspects of the RCBD

See Text, Section 4.1.3.

The RCBD utilizes an **additive model** – one in which there is no interaction between treatments and blocks. *The error term in a randomized complete block model reflects how the treatment effect varies from one block to another.*

Both the treatments and blocks can be looked at as random effects rather than fixed effects, if the levels were selected at random from a population of possible treatments or blocks. We consider this case later, but it does not change the test for a treatment effect.

What are the **consequences** of **not blocking** if we should have? Generally the unexplained error in the model will be larger, and therefore the test of the treatment effect less powerful.

How to determine the sample size in the RCBD? The **OC curve** approach can be used to determine the number of blocks to run. See Section 4.1.3. In a RCBD, b , the number of blocks represents the number of replications. The power calculations that we looked at before would be the same, except that we use b rather than n , and we use the estimate of error, σ^2 , that reflects the improved precision based on having used blocks in our experiment. So, the major benefit or power comes not from the number of replications but from the error variance which is much smaller because you removed the effects due to block.

4.2 - RCBD and RCBD's with Missing Data

Vascular Graft Example 4.1.

This example investigates a procedure to create artificial arteries using a resin. The resin is pressed or extruded through an aperture that forms the resin into a tube.

To conduct this experiment as a RCBD, we need to assign all 4 pressures at random to each of the 6 batches of resin. Each batch of resin is called a “**block**”, since a batch is a more homogenous set of experimental units on which to test the extrusion pressures. Below is a table which provides percentages of those products that met the specifications.

Table 4-3 Randomized Complete Block Design for the Vascular Graft Experiment

Extrusion Pressure (PSI)	Batch of Resin (Block)						Treatment Total
	1	2	3	4	5	6	
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.5	90.6	94.7	87.0	95.8	550.1
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
Block Totals	350.8	359.0	364.0	362.2	341.3	377.8	$y_u = 2155.1$

Note: Since percent response data does not generally meet the assumption of constant variance, we might consider a variance stabilizing transformation, i.e., the arcsine square root of the proportion. However, since the range of the percent data is quite limited, it goes from the high 70s through the 90s, this data seems fairly homogeneous.

Figure 4.2 in the text gives the output from the statistical software package Design Expert:

Loading [MathJax]/extensions/MathMenu.js

Response: Yield
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	192.25	5	38.45		
Model	178.17	3	59.39	8.11	0.0019
A	178.17	3	59.39	8.11	0.0019
Residual	109.89	15	7.33		
Cor Total	480.31	23			
Std. Dev.	2.71			R-Squared	0.6185
Mean	89.80			Adj R-Squared	0.5422
C.V.	3.01			Pred R-Squared	0.0234
PRESS	281.31			Adeq Precision	9.759

Notice that Design Expert does not perform the hypothesis test on the block factor. Should we test the block factor?

Below is the Minitab output which treats both batch and treatment the same and tests the hypothesis of no effect.

ANOVA: Yield versus Batch, Pressure

Factor Type Levels Values
 Batch random 6 1, 2, 3, 4, 5, 6
 Pressure fixed 4 8500, 8700, 8900, 9100

Analysis of Variance for Yield

Source	DF	SS	MS	F	P
Batch	5	192.252	38.450	5.25	0.006
Pressure	3	178.171	59.390	8.11	0.002
Error	15	109.886	7.326		
Total	23	480.310			

S = 2.70661 R-Sq = 77.12% R-Sq(adj) = 64.92%

This example shows the output from the ANOVA command in Minitab (Menu >> Stat >> ANOVA >> Balanced ANOVA). It does hypothesis tests for both batch and pressure, and they are both significant. Otherwise, the results from both programs are very similar.

Again, should we test the block factor? Generally, the answer is no, but in some instances this might be helpful. We use the RCBD design because we hope to remove from error the variation due to the block. If the block factor is not significant, then the block variation, or mean square due to the block treatments is no greater than the mean square due to the error. In other words, if the block F ratio is close to 1 (or generally not greater than 2), you have wasted effort in doing the experiment as a block design, and used in this case 5 degrees of freedom that could be part of error degrees of freedom, hence the design could actually be less efficient!

Therefore, one can test the block simply to confirm that the block factor is effective and explains variation that would otherwise be part of your experimental error. However, you generally cannot make any stronger conclusions from the test on a block factor, because you may not have randomly selected the blocks from any population, nor randomly assigned the levels.

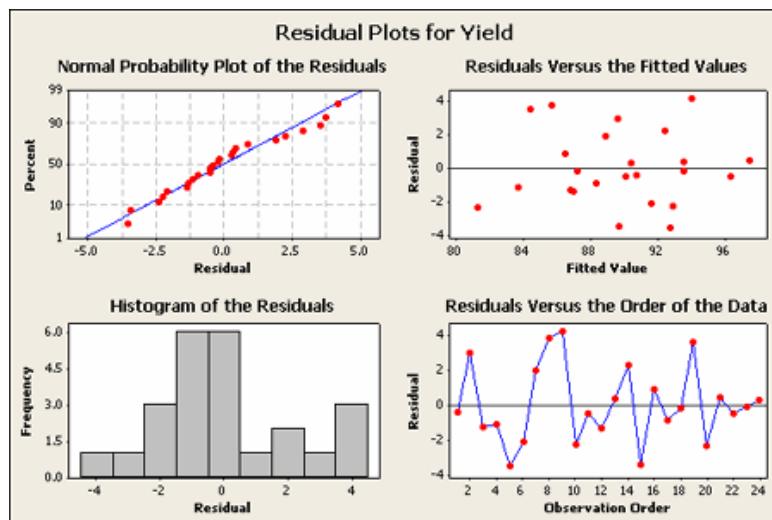
Why did I first say no?

Loading [MathJax]/extensions/MathMenu.js

There are two cases we should consider separately, when blocks are: 1) a classification factor and 2) an experimental factor. In the case where blocks are a batch, it is a classification factor, but it might also be subjects or plots of land which are also classification factors. For a RCBD you can apply your experiment to convenient subjects. In the general case of classification factors, you should sample from the population in order to make inferences about that population. These observed batches are not necessarily a sample from any population. If you want to make inferences about a factor then there should be an appropriate randomization, i.e. random selection, so that you can make inferences about the population. In the case of experimental factors, such as oven temperature for a process, all you want is a representative set of temperatures such that the treatment is given under homogeneous conditions. The point is that we set the temperature once in each block; we don't reset it for each observation. So, there is no replication of the block factor. We do our randomization of treatments within a block. In this case there is an asymmetry between treatment and block factors. In summary, you are only including the block factor to reduce the error variation due to this nuisance factor, not to test the effect of this factor.

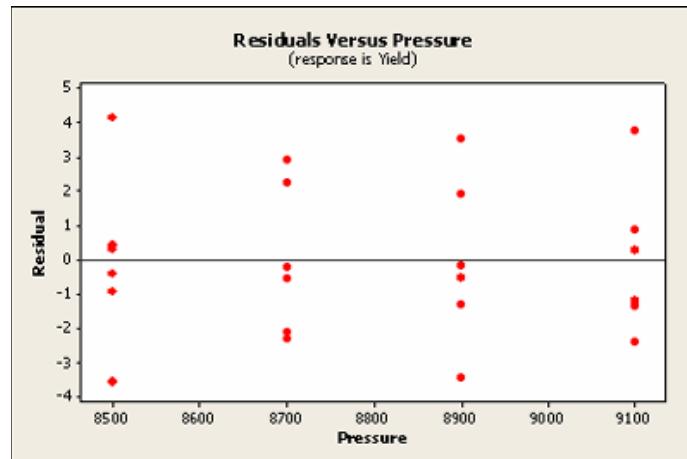
ANOVA: Yield versus Batch, Pressure

The residual analysis for the Vascular Graft example is shown:



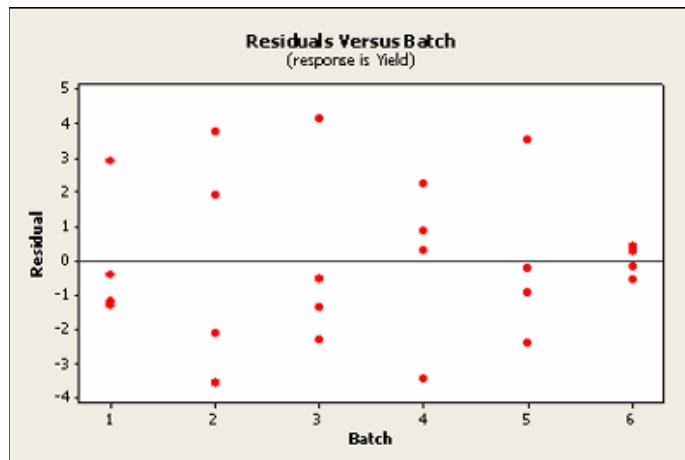
The pattern does not strike me as indicating an unequal variance.

Another way to look at these residuals is to plot the residuals against the two factors. Notice that pressure is the treatment factor and batch is the block factor. Here we'll check for homogeneous variance. Against treatment these look quite homogeneous.



Loading [MathJax]/extensions/MathMenu.js

Plotted against block the sixth does raise ones eyebrow a bit. It seems to be very close to zero.



Basic residual plots indicate that **normality**, **constant variance** assumptions are satisfied. Therefore, there seems to be no obvious problems with **randomization**. These plots provide more information about the constant variance assumption, and can reveal possible outliers. The plot of residuals versus order sometimes indicates a problem with the independence assumption.

Missing Data

In the example dataset above, what if the data point 94.7 (second treatment, fourth block) was missing? What data point can I substitute for the missing point?

If this point is missing we can substitute x , calculate the sum of squares residuals, and solve for x which minimizes the error and gives us a point based on all the other data and the two-way model. We sometimes call this an imputed point, where you use a least squares approach to estimate this missing data point.

After calculating x , you could substitute the estimated data point and repeat your analysis. Now you have an artificial point with known residual zero. So you can analyze the resulting data, but now should reduce your error degrees of freedom by one. In any event, these are all approximate methods, i.e., using the best fitting or imputed point.

Before high-speed computing, data imputation was often done because the ANOVA computations are more readily done using a balanced design. There are times where imputation is still helpful but in the case of a two-way or multiway ANOVA we generally will use the General Linear Model (GLM) and use the full and reduced model approach to do the appropriate test. This is often called the General Linear Test (GLT). Note that text book has mentioned this test as the General Regression Significance Test in Section 4.1.4.

Let's take a look at this in Minitab now...



The sum of squares you want to use to test your hypothesis will be based on the adjusted treatment sum of squares, $R(\tau_i | \mu, \beta_j)$ using the notation in Section 4.1.4 of the text, for testing:

$$H_0 : \tau_i = 0$$

The numerator of the F -test, for the hypothesis you want to test should be based on the adjusted SS's that is last in the sequence or is obtained from the adjusted sums of squares. That will be very close to

Loading [MathJax]/extensions/MathMenu.js

what you would get using the approximate method we mentioned earlier. The general linear test is the most powerful test for this type of situation with unbalanced data.

The General Linear Test can be used to test for significance of multiple parameters of the model at the same time. Generally, significance of all those parameters which are in the Full model but are not included in the Reduced model are tested, simultaneously. The F test statistic is defined as

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

Where F stands for "Full" and R stands for "Reduced." Numerator and denominator degrees of freedom for the F statistic are $df_R - df_F$ and df_F , respectively.

Here are the results for the GLM with all the data intact. There are 23 degrees of freedom total here so this is based on the full set of 24 observations.

General Linear Model: Yield versus Batch, Pressure																																									
Factor Type Levels Values																																									
Batch fixed 6 1, 2, 3, 4, 5, 6																																									
Pressure fixed 4 8500, 8700, 8900, 9100																																									
Analysis of Variance for Yield, using Adjusted SS for Tests																																									
<table> <thead> <tr><th>Source</th><th>DF</th><th>Seq SS</th><th>Adj SS</th><th>Adj MS</th><th>F</th><th>P</th></tr> </thead> <tbody> <tr><td>Batch</td><td>5</td><td>192.252</td><td>192.252</td><td>38.450</td><td>5.25</td><td>0.006</td></tr> <tr><td>Pressure</td><td>3</td><td>178.171</td><td>178.171</td><td>59.390</td><td>8.11</td><td>0.002</td></tr> <tr><td>Error</td><td>15</td><td>109.886</td><td>109.886</td><td>7.326</td><td></td><td></td></tr> <tr><td>Total</td><td>23</td><td>480.310</td><td></td><td></td><td></td><td></td></tr> </tbody> </table>							Source	DF	Seq SS	Adj SS	Adj MS	F	P	Batch	5	192.252	192.252	38.450	5.25	0.006	Pressure	3	178.171	178.171	59.390	8.11	0.002	Error	15	109.886	109.886	7.326			Total	23	480.310				
Source	DF	Seq SS	Adj SS	Adj MS	F	P																																			
Batch	5	192.252	192.252	38.450	5.25	0.006																																			
Pressure	3	178.171	178.171	59.390	8.11	0.002																																			
Error	15	109.886	109.886	7.326																																					
Total	23	480.310																																							
$S = 2.70661$ R-Sq = 77.12% R-Sq(adj) = 64.92%																																									
Least Squares Means for Yield																																									
<table> <thead> <tr><th>Pressure</th><th>Mean</th><th>SE Mean</th></tr> </thead> <tbody> <tr><td>8500</td><td>92.82</td><td>1.105</td></tr> <tr><td>8700</td><td>91.68</td><td>1.105</td></tr> <tr><td>8900</td><td>88.92</td><td>1.105</td></tr> <tr><td>9100</td><td>85.77</td><td>1.105</td></tr> </tbody> </table>							Pressure	Mean	SE Mean	8500	92.82	1.105	8700	91.68	1.105	8900	88.92	1.105	9100	85.77	1.105																				
Pressure	Mean	SE Mean																																							
8500	92.82	1.105																																							
8700	91.68	1.105																																							
8900	88.92	1.105																																							
9100	85.77	1.105																																							
Main Effects Plot (fitted means) for Yield																																									

When the data are complete this analysis from GLM is correct and equivalent to the results from the two-way command in Minitab. When you have missing data, the raw marginal means are wrong. What if the missing data point were from a very high measuring block? It would reduce the overall effect of that treatment, and the estimated treatment mean would be biased.

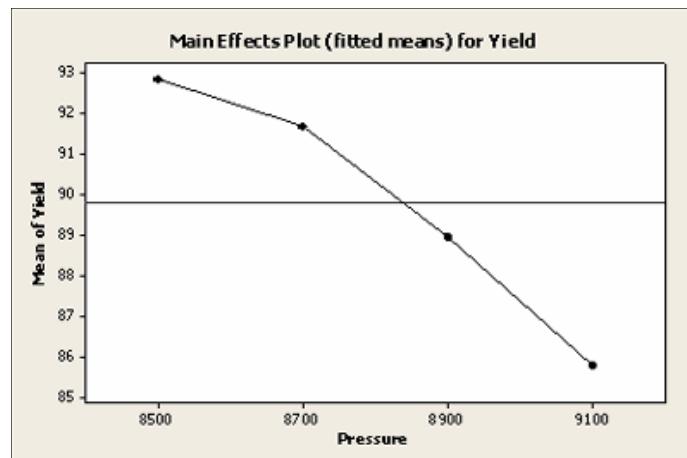
Above you have the least squares means that correspond exactly to the simple means from the earlier analysis.

We now illustrate the GLM analysis based on the missing data situation - one observation missing (Batch 4, pressure 2 data point removed). The least squares means as you can see (below) are slightly different, for pressure 8700. What you also want to notice is the standard error of these means, i.e., the S.E., for the second treatment is slightly larger. The fact that you are missing a point is reflected in the estimate of error. You do not have as many data points on that particular treatment.

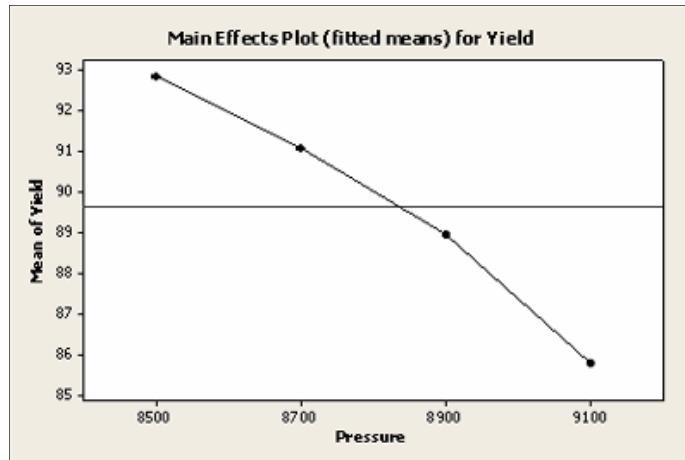
Results for: Ex4-1miss.MTW					
General Linear Model: Yield versus Batch, Pressure					
Factor Type Levels Values					
Batch fixed 6 1, 2, 3, 4, 5, 6					
Pressure fixed 4 8500, 8700, 8900, 9100					
Analysis of Variance for Yield, using Adjusted SS for Tests					
Source	DF	Seq SS	Adj SS	Adj MS	F P
Batch	5	190.119	189.522	37.904	5.22 0.007
Pressure	3	163.398	163.398	54.466	7.50 0.003
Error	14	101.696	101.696	7.264	
Total	22	455.213			
$S = 2.69518 \text{ R-Sq} = 77.66\% \text{ R-Sq(adj)} = 64.89\%$					
Least Squares Means for Yield					
Pressure	Mean	SE Mean			
8500	92.82	1.100			
8700	91.08	1.238			
8900	88.92	1.100			
9100	85.77	1.100			
Main Effects Plot (fitted means) for Yield					

The overall results are similar. We have only lost one point and our hypothesis test is still significant, with a p -value of 0.003 rather than 0.002.

Here is a plot of the least squares means for Yield with all of the observations included.



Here is a plot of the least squares means for Yield with the missing data, not very different.



Again, for any unbalanced data situation we will use the GLM. For most of our examples GLM will be a useful tool for analyzing and getting the analysis of variance summary table. Even if you are unsure whether your data are orthogonal, one way to check if you simply made a mistake in entering your data is by checking whether the sequential sums of squares agree with the adjusted sums of squares.

4.3 - The Latin Square Design

Text reference, Section 4.2.

Latin Square Designs are probably not used as much as they should be - they are very efficient designs. Latin square designs allow for two blocking factors. In other words, these designs are used to simultaneously control (or eliminate) **two sources of nuisance variability**. For instance, if you had a plot of land the fertility of this land might change in both directions, North -- South and East -- West due to soil or moisture gradients. So, both rows and columns can be used as blocking factors. However, you can use Latin squares in lots of other settings. As we shall see, Latin squares can be used as much as the RCBD in industrial experimentation as well as other experiments.

Whenever, you have more than one blocking factor a Latin square design will allow you to remove the variation for these two sources from the error variation. So, consider we had a plot of land, we might have blocked it in columns and rows, i.e. each row is a level of the row factor, and each column is a level of the column factor. We can remove the variation from our measured response in both directions if we consider both rows and columns as factors in our design.

The Latin Square Design gets its name from the fact that we can write it as a square with Latin letters to correspond to the treatments. The treatment factor levels are the Latin letters in the Latin square design. The number of rows and columns has to correspond to the number of treatment levels. So, if we have four treatments then we would need to have four rows and four columns in order to create a Latin square. This gives us a design where we have each of the treatments and in each row and in each column.

This is just one of many 4×4 squares that you could create. In fact, you can make any size square you want, for any number of treatments - it just needs to have the following property associated with it - that each treatment occurs only once in each row and once in each column.

Consider another example in an industrial setting: the rows are the batch of raw material, the columns are the operator of the equipment, and the treatments (A, B, C and D) are an industrial process or protocol for producing a particular product.

What is the model? We let:

$$y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + e_{ijk}$$

Loading [MathJax]/extensions/MathMenu.js

$i = 1, \dots, t$
 $j = 1, \dots, t$
 $[k = 1, \dots, t]$ where - $k = d(i, j)$ and the total number of observations

$N = t^2$ (the number of rows times the number of columns) and t is the number of treatments.

Note that a Latin Square is an incomplete design, which means that it does not include observations for all possible combinations of i, j and k . This is why we use notation $k = d(i, j)$. Once we know the row and column of the design, then the treatment is specified. In other words, if we know i and j , then k is specified by the Latin Square design.

This property has an impact on how we calculate means and sums of squares, and for this reason we can not use the balanced ANOVA command in Minitab even though it looks perfectly balanced. We will see later that although it has the property of orthogonality, you still cannot use the balanced ANOVA command in Minitab because it is not complete.

An assumption that we make when using a Latin square design is that the three factors (treatments, and two nuisance factors) **do not interact**. If this assumption is violated, the Latin Square design error term will be inflated.

The randomization procedure for assigning treatments that you would like to use when you actually apply a Latin Square, is somewhat restricted to preserve the structure of the Latin Square. The ideal randomization would be to select a square from the set of all possible Latin squares of the specified size. However, a more practical randomization scheme would be to select a standardized Latin square at random (these are tabulated) and then:

1. randomly permute the columns,
2. randomly permute the rows, and then
3. assign the treatments to the Latin letters in a random fashion.

Consider a factory setting where you are producing a product with 4 operators and 4 machines. We call the columns the operators and the rows the machines. Then you can randomly assign the specific operators to a row and the specific machines to a column. The treatment is one of four protocols for producing the product and our interest is in the average time needed to produce each product. If both the machine and the operator have an effect on the time to produce, then by using a Latin Square Design this variation due to machine or operators will be effectively removed from the analysis.

The following table gives the degrees of freedom for the terms in the model.

AOV	df	df for the example
Rows	$t-1$	3
Cols	$t-1$	3
Treatments	$t-1$	3
Error	$(t-1)(t-2)$	6
Total	$(t^2 - 1)$	15

A Latin Square design is actually easy to analyze. Because of the restricted layout, one observation per treatment in each row and column, the model is orthogonal.

If the row, ρ_i , and column, β_j , effects are random with expectations zero, the expected value of Y_{ijk} is $\mu + \tau_k$. In other words, the treatment effects and treatment means are orthogonal to the row and column effects. We can also write the sums of squares, as seen in Table 4.10 in the text.

We can test for row and column effects, but our focus of interest in a Latin square design is on the treatments just as in RCBD the row and column factors are included to reduce the error variation but

are not typically of interest. And, depending on how we've conducted the experiment they often haven't been randomized in a way that allows us to make any reliable inference from those tests.

Note: if you have missing data then you need to use the general linear model and test the effect of treatment after fitting the model that would account for the row and column effects.

In general, the General Linear Model tests the hypothesis that:

$$H_0: \tau_i = 0 \quad \text{vs.} \quad H_a: \tau_i \neq 0$$

To test this hypothesis we will look at the F -ratio which is written as:

$$F = \frac{MS(\tau_k | \mu, \rho_i, \beta_j)}{MSE(\mu, \rho_i, \beta_j, \tau_k)} \sim F((t - 1), (t - 1)(t - 2))$$

To get this in Minitab you would use GLM and fit the three terms: rows, columns and treatments. The F statistic is based on the adjusted MS for treatment.

The Rocket Propellant Problem – A Latin Square Design (Table 4.9 in 8th ed and Table 4-8 in 7th ed)

Table 4-8 Latin Square Design for the Rocket Propellant Problem

Batches of Raw Material	Operators				
	1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

Table 4-13 (4-12 in 7th ed) shows some other Latin Squares from $t = 3$ to $t = 7$ and states the number of different arrangements available.

Statistical Analysis of the Latin Square Design

The statistical (effects) model is:

$$Y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + \varepsilon_{ijk} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases}$$

but $k = d(i, j)$ shows the dependence of k in the cell i, j on the design layout, and $p = t$ the number of treatment levels.

The statistical analysis (ANOVA) is much like the analysis for the RCBD.

See the ANOVA table, Table 4.10 (Table 4-9 in 7th ed)

The analysis for the rocket propellant example is presented in Example 4.3.

4.4 - Replicated Latin Squares

Latin Squares are very efficient by including two blocking factors, however the *d.f.* for error are often too small. In these situations, we consider replicating a Latin Square. Let's go back to the factory scenario again as an example and look at $n = 3$ repetitions of a 4×4 Latin square.

Loading [MathJax]/extensions/MathMenu.js

We labeled the row factor the machines, the column factor the operators and the Latin letters denoted the *protocol* used by the operators which was the treatment factor. We will replicate this Latin Square experiment $n = 3$ times. Now we have total observations equal to $N = t^2n$.

You could use the same squares over again in each replicate, but we prefer to randomize these separately for each replicate. It might look like this:

		operators			
		1	2	3	4
machines	1	A	B	C	D
	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C

		operators			
		1	2	3	4
machines	1	D	A	B	C
	2	A	B	C	D
	3	B	C	D	A
	4	C	D	A	B

		operators			
		1	2	3	4
machines	1	C	D	A	B
	2	D	A	B	C
	3	A	B	C	D
	4	B	C	D	A

Ok, with this scenario in mind, let's consider three cases that are relevant and each case requires a different model to analyze. The cases are determined by whether or not the blocking factors are the same or different across the replicated squares. The treatments are going to be the same but the question is whether the levels of the blocking factors remain the same.

Case 1

Here we will have the same row and column levels. For instance, we might do this experiment all in the same factory using the same machines and the same operators for these machines. The first replicate would occur during the first week, the second replicate would occur during the second week, etc. Week one would be replication one, week two would be replication two and week three would be replication three.

We would write the model for this case as:

$$Y_{hijk} = \mu + \delta_h + \rho_i + \beta_j + \tau_k + e_{hijk}$$

where:

$$\begin{aligned} h &= 1, \dots, n \\ i &= 1, \dots, t \\ j &= 1, \dots, t \\ k &= d_h(i, j) \text{ - the Latin letters} \end{aligned}$$

This is a simple extension of the basic model that we had looked at earlier. We have added one more term to our model. The row and column and treatment all have the same parameters, the same effects that we had in the single Latin square. In a Latin square the error is a combination of any interactions that might exist and experimental error. Remember, we can't estimate interactions in a Latin square.

Let's take a look at the analysis of variance table.

AOV	df	df for Case 1	SS
rep=week	$n - 1$	2	See text p. 143
row=machine	$t - 1$	3	
column=operator	$t - 1$	3	

treatment=protocol	$t - 1$	3	
error	$(t - 1)[n(t + 1) - 3]$	36	
Total	$nt^2 - 1$	47	

Case 2

In this case, one of our blocking factors, either row or column, is going to be the same across replicates whereas the other will take on new values in each replicate. Back to the factory example e.g., we would have a situation where the machines are going to be different (you can say they are nested in each of the repetitions) but the operators will stay the same (crossed with replicates). In this scenario, perhaps, this factory has three locations and we want to include machines from each of these three different factories. To keep the experiment standardized, we will move our operators with us as we go from one factory location to the next. This might be laid out like this:

		operators					
		1	2	3	4		
machines	1	A	B	C	D		
	2	B	C	D	A		
	3	C	D	A	B		
	4	D	A	B	C		
		operators					
machines	5	D	A	B	C		
	6	A	B	C	D		
	7	B	C	D	A		
	8	C	D	A	B		
		operators					
machines	9	C	D	A	B		
	10	D	A	B	C		
	11	A	B	C	D		
	12	B	C	D	A		

Rep 1
Factory 1Rep 2
Factory 2Rep 3
Factory 3

There is a subtle difference here between this experiment in a Case 2 and the experiment in Case 1 - but it does affect how we analyze the data. Here the model is written as:

$$Y_{hijk} = \mu + \delta_h + \rho_{i(h)} + \beta_j + \tau_k + e_{hijk}$$

where:

$$\begin{aligned} h &= 1, \dots, n \\ i &= 1, \dots, t \\ j &= 1, \dots, t \\ k &= d_h(i, j) \text{ - the Latin letters} \end{aligned}$$

and the 12 machines are distinguished by nesting the i index within the h replicates.

This affects our ANOVA table. Compare this to the previous case:

AOV	df	df for Case 2	SS
rep = factory	$n - 1$	2	See text p. 144.
row (rep) = machine (factory)	$n(t - 1)$	9	
column = operator	$t - 1$	3	
treatment = protocol	$t - 1$	3	
error		30	

	$(t - 1)(nt - 2)$		
Total	$nt^2 - 1$	47	

Note that Case 2 may also be flipped where you might have the same machines, but different operators.

Case 3

In this case we have different levels of both the row and the column factors. Again, in our factory scenario we would have different machines **and** different operators in the three replicates. In other words, both of these factors would be nested within the replicates of the experiment.

	1	2	3	4		5	6	7	8		9	10	11	12
1	A	B	C	D	5	D	A	B	C	9	C	D	A	B
2	B	C	D	A	6	A	B	C	D	10	D	A	B	C
3	C	D	A	B	7	B	C	D	A	11	A	B	C	D
4	D	A	B	C	8	C	D	A	B	12	B	C	D	A

Rep 1
Factory 1 Rep 2
Factory 2 Rep 3
Factory 3

We would write this model as:

$$Y_{hijk} = \mu + \delta_h + \rho_{i(h)} + \beta_{j(h)} + \tau_k + e_{hijk}$$

where:

$$\begin{aligned} h &= 1, \dots, n \\ i &= 1, \dots, t \\ j &= 1, \dots, t \\ k &= d_h(i, j) \quad - \text{the Latin letters} \end{aligned}$$

Here we have used nested terms for both of the block factors representing the fact that the levels of these factors are not the same in each of the replicates.

The analysis of variance table would include:

AOV	df	df for Case 3	SS
rep = factory	$n - 1$	2	See text p. 144.
row (rep) = machine (factory)	$n(t - 1)$	9	
column (rep) = operator (factory)	$n(t - 1)$	9	
treatment protocol	$t - 1$	3	
error	$(t - 1)[n(t - 1) - 1]$	24	
Total	$nt^2 - 1$	47	

Loading [MathJax]/extensions/MathMenu.js

Which case is best? There really isn't a best here... the choice of case depends on how you need to conduct the experiment. If you are simply replicating the experiment with the same row and column levels, you are in Case 1. If you are changing one or the other of the row or column factors, using different machines or operators, then you are in Case 2. If both of the block factors have levels that differ across the replicates, then you are in Case 3. The third case, where the replicates are different factories, can also provide a comparison of the factories. The fact that you are replicating Latin Squares does allow you to estimate some interactions that you can't estimate from a single Latin Square. If we added a treatment by factory interaction term, for instance, this would be a meaningful term in the model, and would inform the researcher whether the same protocol is best (or not) for all the factories.

The degrees of freedom for error grows very rapidly when you replicate Latin squares. But usually if you are using a Latin Square then you are probably not worried too much about this error. The error is more dependent on the specific conditions that exist for performing the experiment. For instance, if the protocol is complicated and training the operators so they can conduct all four becomes an issue of resources then this might be a reason why you would bring these operators to three different factories. It depends on the conditions under which the experiment is going to be conducted.

Situations where you should use a Latin Square are where you have a **single** treatment factor and you have two blocking or nuisance factors to consider, which can have the same number of levels as the treatment factor.

4.5 - What do you do if you have more than 2 blocking factors?

When might this occur? Let's consider the factory example again. In this factory you have four machines and four operators to conduct your experiment. You want to complete the experimental trials in a week. Use the animation below to see how this example of a typical treatment schedule pans out...

As the treatments were assigned you should have noticed that the treatments have become confounded with the days. Days of the week are not all the same, Monday is not always the best day of the week!

Just like any other factor not included in the design you hope it is not important or you would have included it into the experiment in the first place.

What we now realize is that two blocking factors is not enough! We should also include the day of the week in our experiment. It looks like day of the week could affect the treatments and introduce bias into the treatment effects, since not all treatments occur on Monday. We want a design with 3 blocking factors; machine, operator, and day of the week.

One way to do this would be to conduct the entire experiment on one day and replicate it four times. But this would require $4 \times 16 = 64$ observations not just 16. Or, we could use what is called a Graeco-Latin Square...

Graeco-Latin Squares

We write the Latin square first then each of the Greek letters occurs alongside each of the Latin letters. A Graeco-Latin square is a set of two orthogonal Latin squares where each of the Greek and Latin letters is a Latin square and the Latin square is orthogonal to the Greek square. Use the animation below to explore a Graeco-Latin square:

The Greek letters each occur one time with each of the Latin letters. A Graeco-Latin square is orthogonal between rows, columns, Latin letters and Greek letters. It is completely orthogonal.

How do we use this design?

We let the row be the machines, the column be the operator, (just as before) and the Greek letter the day, ~~Laying out the extension of this month~~ the order in which it was produced). Therefore the Greek letter could serve

the multiple purposes as the day effect or the order effect. The Latin letters are assigned to the treatments as before.

We want to account for all three of the blocking factor sources of variation, and remove each of these sources of error from the experiment. Therefore we must include them in the model.

Here is the model for this design:

$$Y_{ijkl} = \mu + \rho_i + \beta_j + \tau_k + \gamma_l + e_{ijkl}$$

So, we have three blocking factors and one treatment factor.

and i, j, k and l all go from $1, \dots, t$, where i and j are the row and column indices, respectively, and k and l are defined by the design, that is, k and l are specified by the Latin and Greek letters, respectively.

This is a highly efficient design with $N = t^2$ observations.

You could go even farther and have more than two orthogonal latin squares together. These are referred to a Hyper-Graeco-Latin squares!

Fisher, R.A. *The Design of Experiments*, 8th edition, 1966, p.82-84 [4], gives examples of hyper-Graeco-Latin squares for $t = 4, 5, 8$ and 9 .

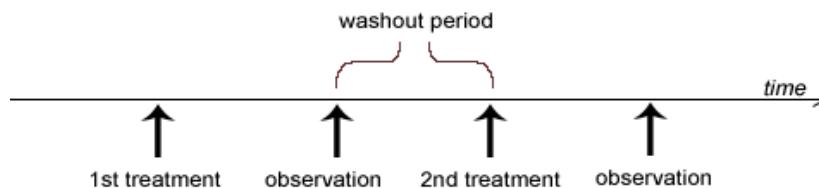
(Note: it is impossible to have a 6×6 Graeco-Latin square! So in designing your experiment with a Graeco-Latin Square - don't have 6 treatments! Add another, or drop one!)

4.6 - Crossover Designs

Crossover designs use the same experimental unit for multiple treatments. The common use of this design is where you have subjects (human or animal) on which you want to test a set of drugs -- this is a common situation in clinical trials for examining drugs.

The simplest case is where you only have 2 treatments and you want to give each subject both treatments. Here as with all crossover designs we have to worry about carryover effects.

Here is a timeline of this type of design.



We give the treatment, then we later observe the effects of the treatment. This is followed by a period of time, often called a washout period, to allow any effects to go away or dissipate. This is followed by a second treatment, followed by an equal period of time, then the second observation.

If we only have two treatments, we will want to balance the experiment so that half the subjects get treatment A first, and the other half get treatment B first. For example, if we had 10 subjects we might have half of them get treatment A and the other half get treatment B in the first period. After we assign the first treatment, A or B, and make our observation, we then assign our second treatment.

This situation can be represented as a set of 5, 2×2 Latin squares.

Loading [MathJax]/extensions/MathMenu.js

		subjects									
		1 2		3 4		5 6		7 8		9 10	
order	1	A	B	A	B	B	A	A	B	A	B
	2	B	A	B	A	A	B	B	A	B	A

We have not randomized these, although you would want to do that, and we do show the third square different from the rest. The row effect is the order of treatment, whether A is done first or second or whether B is done first or second. And the columns are the subjects. So, if we have 10 subjects we could label all 10 of the subjects as we have above, or we could label the subjects 1 and 2 nested in a square. This is similar to the situation where we have replicated Latin squares - in this case five reps of 2×2 Latin squares, just as was shown previously in Case 2.

This crossover design has the following AOV table set up:

AOV	df	df for this example
rep = square	$n - 1$	4
column = subject(sq)	$n(t - 1)$	5
row = order	$t - 1$	1
treatment = A vs. B	$t - 1$	1
error	$(t - 1)[n(t - 1) - 1]$	8
Total	$nt^2 - 1$	19

We have five squares and within each square we have two subjects. So we have 4 degrees of freedom among the five squares. We have 5 degrees of freedom representing the difference between the two subjects in each square. If we combine these two, $4 + 5 = 9$, which represents the degrees of freedom among the 10 subjects. This representation of the variation is just the partitioning of this variation. The same thing applies in the earlier cases we looked at.

With just two treatments there are only two ways that we can order them. Let's look at a crossover design where $t = 3$. If $t = 3$ then there are more than two ways that we can represent the order. The basic building block for the crossover design is the Latin Square.

Here is a 3×3 Latin Square. To achieve replicates, this design could be replicated several times.

Subject					
			1	2	3
Order	1	A	B	C	
	2	B	C	A	
	3	C	A	B	

In this Latin Square we have each treatment occurring in each period. Even though Latin Square guarantees that treatment A occurs once in the first, second and third period, we don't have all sequences represented. It is important to have all sequences represented when doing clinical trials with drugs.

Crossover Design Balanced for Carryover Effects.

The following crossover design, is based on two orthogonal Latin squares.

Loading [MathJax]/extensions/MathMenu.js

Order	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 33.33%;">A</td><td style="width: 33.33%;">B</td><td style="width: 33.33%;">C</td></tr> <tr><td>B</td><td>C</td><td>A</td></tr> <tr><td>C</td><td>A</td><td>B</td></tr> </table>	A	B	C	B	C	A	C	A	B	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 33.33%;">A</td><td style="width: 33.33%;">B</td><td style="width: 33.33%;">C</td></tr> <tr><td>C</td><td>A</td><td>B</td></tr> <tr><td>B</td><td>C</td><td>A</td></tr> </table>	A	B	C	C	A	B	B	C	A
A	B	C																		
B	C	A																		
C	A	B																		
A	B	C																		
C	A	B																		
B	C	A																		

Together, you can see that going down the columns every pairwise sequence occurs twice, AB, BC, CA, AC, BA, CB going down the columns. The combination of these two Latin squares gives us this additional level of balance in the design, than if we had simply taken the standard Latin square and duplicated it.

To do a crossover design, each subject receives each treatment at one time in some order. So, one of its benefits is that you can use each subject as its own control, either as a paired experiment or as a randomized block experiment, the subject serves as a block factor. For each subject we will have each of the treatments applied. The number of periods is the same as the number of treatments. It is just a question about what order you give the treatments. The smallest crossover design which allows you to have each treatment occurring in each period would be a single Latin square.

A 3×3 Latin square would allow us to have each treatment occur in each time period. We can also think about period as the order in which the drugs are administered. One sense of balance is simply to be sure that each treatment occurs at least one time in each period. If we add subjects in sets of complete Latin squares then we retain the orthogonality that we have with a single square.

Order	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 33.33%;">A</td><td style="width: 33.33%;">B</td><td style="width: 33.33%;">C</td></tr> <tr><td>B</td><td>C</td><td>A</td></tr> <tr><td>C</td><td>A</td><td>B</td></tr> </table>	A	B	C	B	C	A	C	A	B	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 33.33%;">A</td><td style="width: 33.33%;">B</td><td style="width: 33.33%;">C</td></tr> <tr><td>C</td><td>A</td><td>B</td></tr> <tr><td>B</td><td>C</td><td>A</td></tr> </table>	A	B	C	C	A	B	B	C	A
A	B	C																		
B	C	A																		
C	A	B																		
A	B	C																		
C	A	B																		
B	C	A																		

In designs with two orthogonal Latin Squares we have all ordered pairs of treatments occurring twice and only twice throughout the design. Take a look at the animation below to get a sense of how this occurs:

All ordered pairs occur an equal number of times in this design. It is balanced in terms of residual effects, or carryover effects.

For an odd number of treatments, e.g. 3, 5, 7, etc., it requires two orthogonal Latin squares in order to achieve this level of balance. For even number of treatments, 4, 6, etc., you can accomplish this with a single square. This form of balance is denoted balanced for carryover (or residual) effects.

Here is an actual data example for a design balanced for carryover effects. In this example the subjects are cows and the treatments are the diets provided for the cows. Using the two Latin squares we have three diets A, B, and C that are given to 6 different cows during three different time periods of six weeks each, after which the weight of the milk production was measured. In between the treatments a wash out period was implemented.

cows			cows					
period			1	2	3	4	5	6
1	A	B	C	A	B	C	A	B
2	B	C	A	C	A	B	B	C
3	C	A	B	B	C	A	C	A

Loading [MathJax]/extensions/MathMenu.js

How do we analyze this? If we didn't have our concern for the residual effects then the model for this experiment would be:

$$Y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + e_{ijk}$$

where:

ρ_i = period

β_j = cows

τ_k = treatment

$i = 1, \dots, 3$ (the number of treatments)

$j = 1, \dots, 6$ (the number of cows)

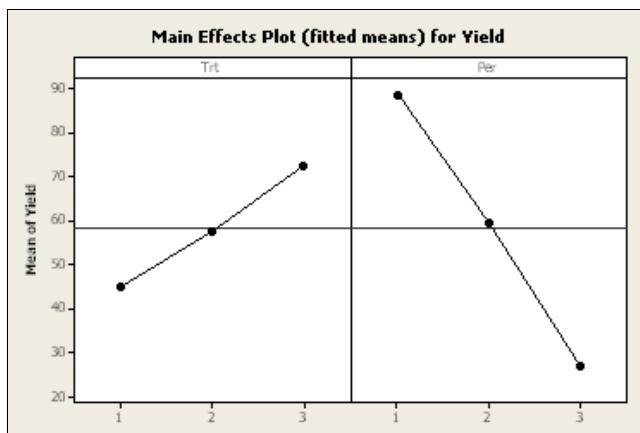
$k = 1, \dots, 3$ (the number of treatments)

Let's take a look at how this is implemented in Minitab using GLM. Use the viewlet below to walk through an initial analysis of the data ([cow_diets.MTW](#) [5]) for this experiment with cow diets. Reference: W.G. Cochran and G.M. Cox, 1957, Experimental Designs, 2nd edition, p. 135.

These demonstrations [Inspect] are based on Minitab Version 16 or earlier. The GLM command menus in Minitab Version 17 have changed.



Why do we use GLM? We do not have observations in all combinations of rows, columns and treatments, since the design is based on the Latin square.



Here is a plot of the least square means for treatment and period. We can see in the table below that the other blocking factor, cow, is also highly significant.

Loading [MathJax]/extensions/MathMenu.js

General Linear Model: Yield versus Per, Cow, Trt						
Factor	Type	Levels	Values			
Per	fixed	3	1, 2, 3			
Cow	fixed	6	1, 2, 3, 4, 5, 6			
Trt	fixed	3	1, 2, 3			

Analysis of Variance for Yield, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Per	2	11480.1	11480.1	5740.1	55.70	0.000
Cow	5	5781.1	5781.1	1156.2	11.22	0.002
Trt	2	2276.8	2276.8	1138.4	11.05	0.005
Error	8	824.4	824.4	103.1		
Total	17	20362.4				

S = 10.1516	R-Sq = 95.95%	R-Sq(adj) = 91.40%
-------------	---------------	--------------------

So, let's go one step farther...

Is this an example of the Case 2 or the Case 3 of the multiple Latin Squares that we had looked at earlier?

This is a Case 2 where the column factor, the cows are nested within the square, but the row factor, *period*, is the same across squares.

Notice the sum of squares for cows is 5781.1. Let's change the model slightly using the general linear model in Minitab again. Follow along with the animation.



[7]

Now I want to move from Case 2 to Case 3. Is the *period* effect in the first square the same as the *period* effect in the second square? If it only means order and all the cows start lactating at the same time it might mean the same. But if some of the cows are done in the spring and others are done in the fall or summer, then the *period* effect has more meaning than simply the order. Although this represents order it may also involve other effects you need to be aware of this. A Case 3 approach involves estimating separate *period* effects within each square.



[8]

My guess is that they all started the experiment at the same time - in this case the first model would have been appropriate.

How Do We Analyze Carryover Effect?

OK, we are looking at the main treatment effects. With our first cow, during the first period, we give it a treatment or diet and we measure the yield. Obviously you don't have any carryover effects here because it is the first period. However, what if the treatment they were first given was a really bad treatment? In fact in this experiment the diet A consisted of only roughage, so, the cow's health might in fact deteriorate as a result of this treatment. This could carry over into the next period. This carry over would hurt the second treatment if the washout period isn't long enough. The measurement at this point is a direct reflection of treatment B but may also have some influence from the previous treatment, treatment A.

If you look at how we have coded data here, we have another column called residual treatment. For the first six observations we have just assigned this a value of 0 because there is no residual treatment. But for the first observation in the second row, we have labeled this with a value of one indicating that this

Loading [MathJax]/extensions/MathMenu.js

was the treatment prior to the current treatment (treatment A). In this way the data is coded such that this column indicates the treatment given in the prior period for that cow.

Now we have another factor that we can put in our model. Let's take a look at how this looks in Minitab:



We have learned everything we need to learn. We have the appropriate analysis of variance here. By fitting in order, when residual treatment (i.e., ResTrt) was fit last we get:

$$\begin{aligned} \text{SS(treatment | period, cow)} &= 2276.8 \\ \text{SS(ResTrt | period, cow, treatment)} &= 616.2 \end{aligned}$$

When we flip the order of our treatment and residual treatment, we get the sums of squares due to fitting residual treatment after adjusting for period and cow:

$$\begin{aligned} \text{SS(ResTrt | period, cow)} &= 38.4 \\ \text{SS(treatment | period, cow, ResTrt)} &= 2854.6 \end{aligned}$$

Which of these are we interested in? If we wanted to test for residual treatment effects how would we do that? What would we use to test for treatment effects if we wanted to remove any carryover effects?

4.7 - Incomplete Block Designs

In using incomplete block designs we will use the notation $t = \#$ of treatments. We define the block size as k . And, as you will see, in incomplete block designs k will be less than t . You cannot assign all of the treatments in each block. In short,

$$\begin{aligned} t &= \# \text{ of treatments}, \\ k &= \text{block size}, \\ b &= \# \text{ of blocks}, \\ r_i &= \# \text{ of replicates for treatment } i, \text{ in the entire design}. \end{aligned}$$

Remember that an equal number of replications is the best way to be sure that you have minimum variance if you're looking at all possible pairwise comparisons. If $r_i = r$ for all treatments, the total number of observations in the experiment is N where:

$$N = t(r) = b(k)$$

The incidence matrix which defines the design of the experiment, gives the number of observations say n_{ij} for the i th treatment in the j th block. This is what it might look like here:

n_{ij}	blocks
treatments	1 2 ... b
1	
2	
.	
.	
t	

Loading [MathJax]/extensions/MathMenu.js

Here we have treatments 1, 2, up to t and the blocks 1, 2, up to b . For a complete block design we would have each treatment occurring one time within each block, so all entries in this matrix would be 1's. For an incomplete block design the incidence matrix would be 0's and 1's simply indicating whether or not that treatment occurs in that block.

Example 1

The example that we will look at is Table 4.22 (4.21 in 7th ed). Here is the incidence matrix for this example:

	1	2	3	4
1	1	1	0	1
2	0	1	1	1
3	1	1	1	0
4	1	0	1	1

Here we have $t = 4$, $b = 4$, (four rows and four columns) and $k = 3$ (so at each block we can only put three of the four treatments leaving one treatment out of each block). So, in this case, the row sums (r_i) and the column sums, k , are all equal to 3.

	1	2	3	4	
1	1	1	0	1	3
2	0	1	1	1	3
3	1	1	1	0	3
4	1	0	1	1	3
	3	3	3	3	

In general, we are faced with a situation where the number of treatments is specified, and the block size, or number of experimental units per block (k) is given. This is usually a constraint given from the experimental situation. And then, the researcher must decide how many blocks are needed to run and how many replicates that provides in order to achieve the precision or the power that you want for the test.

Example 2

Here is another example of an incidence matrix for allocating treatments and replicates in an incomplete block design. Let's take an example where $k = 2$, still $t = 4$, and $b = 4$. That gives us a case $r = 2$. In This case we could design our incidence matrix so that it might look like this:

	1	2	3	4	
1	1	0	1	0	2
2	0	1	0	1	2
3	1	0	1	0	2
4	0	1	0	1	2
	2	2	2	2	

This example has two observations per block so $k = 2$ in each case and for all treatments $r = 2$.

Balanced Incomplete Block Design (BIBD)

A BIBD is an incomplete block design where all pairs of treatments occur together within a block an equal number of times (λ). In general, we will specify $\lambda_{ii'}$ as the number of times treatment i occurs with i' , in a block.

Let's look at the previous cases. How many times does treatment one and two occur together in this first example design?

		blocks				
		1	2	3	4	
treatments	1	1	1	0	1	3
	2	0	1	1	1	3
	3	1	1	1	0	3
	4	1	0	1	1	3
		3	3	3	3	

It occurs together in block 2 and then again in block 4 (highlighted in light blue). So, $\lambda_{12} = 2$. If we look at treatment one and three, this occurs together in block one and in block two therefore $\lambda_{13} = 2$. In this design you can look at all possible pairs. Let's look at 1 and 4 - they occur together twice, 2 and 3 occur together twice, 2 and 4 twice, and 3 and 4 occur together twice. For this design $\lambda_{ii'} = 2$ for all ii' treatment pairs defining the concept of balance in this incomplete block design.

If the number of times treatments occur together within a block is equal across the design for all pairs of treatments then we call this a balanced incomplete block design (BIBD).

Now look at the incidence matrix for the second example.

	1	2	3	4	
1	1	0	1	0	2
2	0	1	0	1	2
3	1	0	1	0	2
4	0	1	0	1	2
	2	2	2	2	

We can see that:

- λ_{12} occurs together 0 times.
- λ_{13} occurs together 2 times.
- λ_{14} occurs together 0 times.
- λ_{23} occurs together 0 times.
- λ_{24} occurs together 2 times.
- λ_{34} occurs together to 0 times.

Here we have two pairs occurring together 2 times and the other four pairs occurring together 0 times. Therefore, this is not a balanced incomplete block design (BIBD).

What else is there about BIBD?

We can define λ in terms of our design parameters when we have equal block size k , and equal replication $r_i = r$. For a given set of t , k , and r we define λ as:

$$\lambda = r(k-1) / t-1$$

So, for the first example that we looked at earlier - let's plug in the values and calculate λ :

$$\lambda = 3(3-1) / (4-1) = 2$$

Here is the key: when λ is equal to an integer number it tells us that a balanced incomplete block design exists. Let's look at the second example and use the formula and plug in the values for this second example. So, for $t = 4$, $k = 2$, $r = 2$ and $b = 4$, we have:

$$\lambda = 2(2-1) / (4-1) = 0.666$$

Since λ is not an integer there does not exist a balanced incomplete block design for this experiment. We would either need more replicates or a larger block size. Seeing as how the block size in this case is fixed, we can achieve a balanced complete block design by adding more replicates so that λ equals at least 1. It needs to be a whole number in order for the design to be balanced.

We will talk about partially balanced designs later. But in thinking about this case we note that a balanced design doesn't exist so what would be the best partially balanced design? That would be a question that you would ask if you could only afford four blocks and the block size is two. Given this situation, is the design in Example 2 the best design we can construct? The best partially balanced design is where $\lambda_{ii'}$ should be the nearest integers to the λ that we calculated. In our case each $\lambda_{ii'}$ should be either 0 or 1, the integers nearest 0.667. This example is not as close to balanced as it could be. In fact, it is not even a connected design where you can go from any treatment to any other treatment within a block. More about this later...

Loading [MathJax]/extensions/MathMenu.js

How do you Construct a BIBD?

In some situations, it is easy to construct the best IBD, however, for other cases it can be quite difficult and we will look them up in a reference.

Let's say that we want six blocks, we still want 4 treatments and our block size is still 2. Calculate $\lambda = r(k - 1) / (t - 1) = 1$. We want to create all possible pairs of treatments because lambda is equal to one. We do this by looking at all possible combinations of four treatments taking two at a time. We could set up the incidence matrix for the design or we could represent it like this - entries in the table are treatment labels: {1, 2, 3, 4}.

	1	2	3	4	5	6
1	1	1	1	2	2	3
2	2	3	4	3	4	4

However, this method of constructing a BIBD using all possible combinations, does not always work as we now demonstrate. If the number of combinations is too large then you need to find a subset - - not always easy to do. However, sometimes you can use Latin Squares to construct a BIBD. As an example, let's take any 3 columns from a 4×4 Latin Square design. This subset of columns from the whole Latin Square creates a BIBD. However, not every subset of a Latin Square is a BIBD.

Let's look at an example. In this example we have $t = 7$, $b = 7$, and $k = 3$. This means that $r = 3 = (bk) / t$. Here is the 7×7 Latin square :

A	B	C	D	E	F	G
B	C	D	E	F	G	A
C	D	E	F	G	A	B
D	E	F	G	A	B	C
E	F	G	A	B	C	D
F	G	A	B	C	D	E
G	A	B	C	D	E	F

We want to select ($k = 3$) three columns out of this design where each treatment occurs **once** with every other treatment because $\lambda = 3(3 - 1) / (7 - 1) = 1$.

We could select the first three columns - let's see if this will work. Click the animation below to see whether using the first three columns would give us combinations of treatments where treatment pairs are not repeated.

Since the first three columns contain some pairs more than once, let's try columns 1, 2, and now we need a third...how about the fourth column. If you look at all possible combinations in each row, each treatment pair occurs only one time.

What if we could afford a block size of 4 instead of 3? Here $t = 7$, $b = 7$, $k = 4$, then $r = 4$. We calculate $\lambda = r(k - 1) / (t - 1) = 2$ so a BIBD does exist. For this design with a block size of 4 we can select 4 columns (or rows) from a Latin square. Let's look at columns again... can you select the correct 4?

Now consider the case with 8 treatments. The number of possible combinations of 8 treatments taking 4 at a time is 70. Thus with 70 sets of 4 from which you have to choose 14 blocks - - wow, this is a big job!

Loading [MathJax]/extensions/MathMenu.js

At this point, we should simply look at an appropriate reference. Here is a handout - a catalog that will help you with this selection process [10] - taken from Cochran & Cox, *Experimental Design*, p. 469-482.

Analysis of BIBD's

When we have missing data, it affects the average of the remaining treatments in a row, i.e., when complete data does not exist for each row - this affects the means. When we have complete data the block effect and the column effects both drop out of the analysis since they are orthogonal. With missing data or IBDs that are not orthogonal, even BIBD where orthogonality does not exist, the analysis requires us to use GLM which codes the data like we did previously. The GLM fits first the block and then the treatment.

The sequential sums of squares (Seq SS) for block is not the same as the Adj SS.

We have the following:

Seq SS

$$\begin{aligned} \text{SS}(\beta | \mu) &= 55.0 \\ \text{SS}(\tau | \mu, \beta) &= 22.50 \end{aligned}$$

Adj SS

$$\begin{aligned} \text{SS}(\beta | \mu, \tau) &= 66.08 \\ \text{SS}(\tau | \mu, \beta) &= 22.75 \end{aligned}$$

Switch them around...now first fit treatments and then the blocks.

Seq SS

$$\begin{aligned} \text{SS}(\tau | \mu) &= 11.67 \\ \text{SS}(\beta | \mu, \tau) &= 66.08 \end{aligned}$$

Adj SS

$$\begin{aligned} \text{SS}(\tau | \mu, \beta) &= 22.75 \\ \text{SS}(\beta | \mu, \tau_i) &= 66.08 \end{aligned}$$

The 'least squares means' come from the fitted model. Regardless of the pattern of missing data or the design we can conceptually think of our design represented by the model:

$$Y_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$

$$i = 1, \dots, b, \quad j = 1, \dots, t$$

You can obtain the 'least squares means' from the estimated parameters from the least squares fit of the model.

Optional Section

See the discussion in the text for *Recovery of Interblock Information*, p. 154. This refers to a procedure which allows us to extract additional information from a BIBD when the blocks are a random effect.

Optionally you can read this section. We illustrate the analysis by the use of the software, PROC Mixed in SAS (L03_sas_Ex_4_5.sas [11]):

```

data: input blk trt Y; cards;
1 1 73
1 3 73
1 4 75
2 1 74
2 2 75
2 3 75
3 2 67
3 3 68
3 4 72
4 1 71
4 2 72
4 4 75
;;
/*This data is from Example 4-5 in Montgomery, Design and Analysis of experiments, 6th edition, */
/* Wiley, 2005, pages 147-154. This demonstrates the recovery of interblock information when */
/* the blocks are considered random. */
proc glm; class trt blk;
model Y = blk trt;
lsmeans trt/ e stder pdiff;

proc mixed; class trt blk;
model Y = trt;
random blk;
lsmeans trt/ e pdiff;
/* The next 4 estimate statements calculate the treatment effects from the solution*/
estimate "trt effect 1" trt +.75 -.25 -.25 -.25/e;
estimate "trt effect 2" trt -.25 +.75 -.25 -.25/e;
estimate "trt effect 3" trt -.25 -.25 +.75 -.25/e;
estimate "trt effect 4" trt -.25 -.25 -.25 +.75/e;
/* The next 3 contrast statements show one set of orthogonal contrasts*/
contrast "trt1 vs trt2-4" trt 3 -1 -1 -1;
contrast "trt2 vs trt3-4" trt 0 2 -1 -1 ;
contrast "trt3 vs trt4" trt 0 0 1 -1 ;
run;

```



[12]

SAS output file ([L03_sas_Ex_4_5.lst](#) [13])

Note that the least squares means for treatments, when using PROC Mixed, correspond to the combined intra- and inter-block estimates of the treatment effects.

Random Effect Factor

So far we have discussed experimental designs with fixed factors, that is, the levels of the factors are fixed and constrained to some specific values. However, this is often not the case. In some cases, the levels of the factors are selected at random from a larger population. In this case, the inference made on the significance of the factor can be extended to the whole population but the factor effects are treated as contributions to variance.

Random effect models are the topic of Chapter 13 in the text book and we will go through them in Lesson 12. Minitab's General Linear Command handles random factors appropriately as long as you are careful to select which factors are fixed and which are random.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/18>

Links:

- [1] https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/data/tip_hardness.txt
- [2] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_block_tip_viewlet_swf.html', 'l04_block_tip', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_block_tip_viewlet_swf.html', 'l04_block_tip', 718, 668);)
- [3] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_missing_data_viewlet_swf.html', 'l04_missing_data', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_missing_data_viewlet_swf.html', 'l04_missing_data', 718, 668);)
- [4] https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/graeo_latin_fisher.pdf
- [5] https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/cow_diets.MTW
- [6] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('https://onlinecourses.science.psu.edu/stat503/print/book/export/html/18'\)](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('https://onlinecourses.science.psu.edu/stat503/print/book/export/html/18'))

('/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_cow_GLM_step01_viewlet_swf.html',
'l04_cow_glm_step01', 718, 668);
[7] https://onlinecourses.science.psu.edu/stat503/javascript:popup_window
('/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_cow_GLM_step02_viewlet_swf.html',
'l04_cow_glm_step02', 718, 668);
[8] https://onlinecourses.science.psu.edu/stat503/javascript:popup_window
('/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_cow_GLM_step03_viewlet_swf.html',
'l04_cow_glm_step03', 718, 668);
[9] https://onlinecourses.science.psu.edu/stat503/javascript:popup_window
('/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_cow_GLM_step04_viewlet_swf.html',
'l04_cow_glm_step04', 718, 668);
[10]
https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/Cochran_Cox.pdf
[11]
https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L03_sas_Ex_4_5.sas
[12] https://onlinecourses.science.psu.edu/stat503/javascript:popup_window
('/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L04_interblock_viewlet_swf.html', 'l04_interblock', 821,
653);
[13]
https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson04/L03_sas_Ex_4_5.lst

Loading [MathJax]/extensions/MathMenu.js



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 5: Introduction to Factorial Designs

Lesson 5: Introduction to Factorial Designs

Introduction

Factorial designs are the basis for another important principle besides blocking - examining several factors simultaneously. We will start by looking at just two factors and then generalize to more than two factors. Investigating multiple factors in the same design automatically gives us replication for each of the factors.

Learning objectives & outcomes

Goals for this lesson includes the following

- Introductory understanding of Factorial Designs as among the most common experimental designs
- Two factor Factorial Design and its extension to the General Factorial Designs
- Sample size determination in Factorial Designs

5.1 - Factorial Designs with Two Treatment Factors

For now we will just consider two treatment factors of interest. It looks almost the same as the randomized block design model only now we are including an interaction term:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

where $i = 1, \dots, a$, $j = 1, \dots, b$, and $k = 1, \dots, n$. Thus we have two factors in a factorial structure with n observations per cell. As usual, we assume the $e_{ijk} \sim N(0, \sigma^2)$, i.e. independently and identically distributed with the normal distribution. Although it looks like a multiplication, the interaction term need not imply multiplicative interaction.

The Effects Model vs. the Means Model

The cell means model is written:

Here the cell means are: $\mu_{11}, \dots, \mu_{1b}, \dots, \mu_{a1} \dots \mu_{ab}$. Therefore we have $a \times b$ cell means, μ_{ij} . We will define our marginal means as the simple average over our cell means as shown below:

$$\bar{\mu}_{i\cdot} = \frac{1}{b} \sum_j \mu_{ij}, \quad \bar{\mu}_{\cdot j} = \frac{1}{a} \sum_i \mu_{ij}$$

From the cell means structure we can talk about marginal means and row and column means. But first we want to look at the effects model and define more carefully what the interactions are. We can write the cell means in terms of the full effects model:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

It follows that the interaction terms $(\alpha\beta)_{ij}$ are defined as the difference between our cell means and the additive portion of the model:

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$$

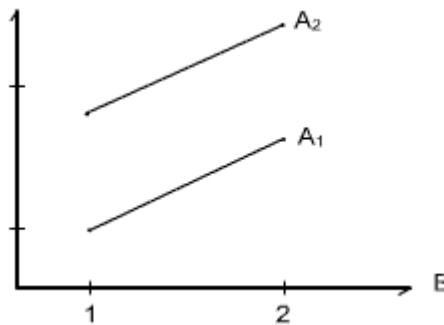
If the true model structure is additive then the interaction terms $(\alpha\beta)_{ij}$ are equal to zero. Then we can say that the true cell means, $\mu_{ij} = (\mu + \alpha_i + \beta_j)$, have additive structure.

Example 1

Let's illustrate this by considering the true means μ_{ij} :

		B		$\bar{\mu}_{i\cdot}$	α_i
		1	2		
A	1	5	11	8	-2
	2	9	15	12	2
		$\bar{\mu}_{\cdot j}$	7	13	10
		β_j	-3	3	

Note that both a and b are 2, thus our marginal row means are 8 and 12, and our marginal column means are 7 and 13. Next, let's calculate the α and the β effects; since the overall mean is 10, our α effects are -2 and 2 (which sum to 0), and our β effects are -3 and 3 (which also sum to 0). If you plot the cell means you get two lines that are parallel.



The difference between the two means at the first β factor level is $9 - 5 = 4$. The difference between the means for the second β factor level is $15 - 11 = 4$. We can say that the effect of α at the first level of β is the same as the effect of α at the second level of β . Therefore we say that there is no interaction and as we will see the interaction terms are equal to 0.

This example simply illustrates that the cell means in this case have additive structure. A problem with data that we actually look at is that you do not know in advance whether the effects are additive or not. Because of random error, the interaction terms are seldom exactly zero. You may be involved in a situation that is either additive or non-additive, and the first task is to decide between them.

Now consider the non-additive case. We illustrate this with Example 2 which follows.

Example 2

This example was constructed so that the marginal means and the overall means are the same as in Example 1. However, it does not have additive structure.

μ_{ij}			$\bar{\mu}_{i\cdot}$	α_i	
		3	13	8	-2
		11	13	12	2
$\bar{\mu}_{j\cdot}$	7	13	10		
β_j	-3	3			

Using the definition of interaction:

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$$

which gives us $(\alpha\beta)_{ij}$ interaction terms that are -2, 2, 2, -2. Again, by the definition of our interaction effects these $(\alpha\beta)_{ij}$ terms should sum to zero in both directions.

$$(\alpha\beta)_{ij}$$

-2	2
2	-2

We generally call the α_i terms the treatment effects for treatment factor A and the β_j terms for treatment factor B, and the $(\alpha\beta)_{ij}$ terms the interaction effects.

The model we have written gives us a way to represent in a mathematical form a two factor design, whether we use the means model or the effects model, i.e.,

$$Y_{ijk} = \mu_{ij} + e_{ijk}$$

or

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

There is really no benefit to the effects model when there is interaction, except that it gives us a mechanism for partitioning the variation due to the two treatments and their interactions. Both models have the same number of distinct parameters. However, when there is no interaction then we can remove the interaction terms from the model and use the reduced additive model.

Now, we'll take a look at the strategy for deciding whether our model fits, whether the assumptions are satisfied and then decide whether we can go forward with an interaction model or an additive model. This is the first decision. When you can eliminate the interactions because they are not significantly different from zero, then you can use the simpler additive model. This should be the goal whenever possible because then you have fewer parameters to estimate, and a simpler structure to represent the underlying scientific process.

Before we get to the analysis, however, we want to introduce another definition of effects - rather than defining the α_i effects as deviation from the mean, we can look at the difference between the high and the low levels of factor A. These are two different definitions of effects that will be introduced and discussed in this chapter and the next, the α_i effects and the difference between the high and low levels, which we will generally denote as the A effect.

Factorial Designs with 2 Treatment Factors, cont'd

For a completely randomized design, which is what we discussed for the one-way ANOVA, we need to have $n \times a \times b = N$ total experimental units available. We randomly assign n of those experimental units to each of the $a \times b$ treatment combinations. For the moment we will only consider the model with fixed effects and constant experimental random error.

The model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

$$\begin{aligned} i &= 1, \dots, a \\ j &= 1, \dots, b \\ k &= 1, \dots, n \end{aligned}$$

Read the text section 5.3.2 for the definitions of the means and the sum of squares.

Testing Hypotheses

We can test the hypotheses that the marginal means are all equal, or in terms of the definition of our effects that the α 's are all equal to zero, and the hypothesis that the β 's are all equal to zero. And, we can test the hypothesis that the interaction effects are all equal to zero. The alternative hypotheses are that at least one of those effects is not equal to zero.

How we do this, in what order, and how do we interpret these tests?

One of the purposes of a factorial design is to be efficient about estimating and testing factors A and B in a single experiment. Often we are primarily interested in the main effects. Sometimes, we are also interested in knowing whether the factors interact. In either case, the first test we should do is the test on the interaction effects.

The Test of $H_0: (\alpha\beta)_{ij} = 0$

If there is interaction and it is significant, i.e. the p -value is less than your chosen cut off, then what do we do? If the interaction term is significant that tells us that the effect of A is different at each level of B. Or you can say it the other way, the effect of B differs at each level of A. Therefore, when we have significant interaction, it is not very sensible to even be talking about the main effect of A and B, because these change depending on the level of the other factor. If the interaction is significant then we want to estimate and focus our attention on the cell means. If the interaction is not significant, then we can test the main effects and focus on the main effect means.

The estimates of the interaction and main effects are given in the text in section 5.3.4.

Note that the estimates of the marginal means for A are the marginal means:

$$\bar{y}_{i..} = \frac{1}{bn} \sum_j \sum_k y_{ijk}, \text{ with } var(\bar{y}_{i..}) = \frac{\sigma^2}{bn}$$

A similar formula holds for factor B, with

$$var(\bar{y}_{.j.}) = \frac{\sigma^2}{an}$$

Just the form of these variances tells us something about the efficiency of the two factor design. A benefit of a two factor design is that the marginal means have $n \times b$ number of replicates for factor A and $n \times a$ for factor B . The factorial structure, when you do not have interactions, gives us the efficiency benefit of having additional replication, the number of observations per cell times the number of levels of the other factor. This benefit arises from factorial experiments rather than single factor experiments with n observations per cell. An alternative design choice could have been to do two one-way experiments, one with a treatments and the other with b treatments, with n observations per cell. However, these two experiments would not have provided the same level of precision, nor the ability to test for interactions.

Another practical question: If the interaction test is not significant what should we do?

Do we get remove the interaction term in the model? You might consider dropping that term from the model. If n is very small and your df for error are small, then this may be a critical issue. There is a 'rule of thumb' that I sometimes use in these cases. If the p-value for the interaction test is greater than 0.25 then you can drop the interaction term. This is not an exact cut off but a general rule. Remember, if you drop the interaction term, then a variation accounted for by SS_{ab} would become part of the error and increasing the SSE, however your error df would also become larger in some cases enough to increase the power of the tests for the main effects. Statistical theory shows that in general dropping the interaction term increases your false rejection rate for subsequent tests. Hence we usually do not drop nonsignificant terms when there are adequate sample sizes.

However, if we are doing an independent experiment with the same factors we might not include interaction in the model for that experiment.

What if $n = 1$, and we have only 1 observation per cell? If $n = 1$ then we have 0 df for SS_{error} and we cannot estimate the error variance with MSE. What should we do in order to test our hypothesis? We obviously cannot perform the test for interaction because we have no error term.

If you are willing to assume, and if it is true that there is no interaction, then you can use the interaction as your F -test denominator for testing the main effects. It is a fairly safe and conservative thing to do. If it is not true then the MS_{ab} will tend to be larger than it should be, so the F -test is conservative. You're not likely to reject a main effect if it is not true. You won't make a Type I error, however you could more likely make a Type II error.

Extension to a 3 Factor Model

The factorial model with three factors can be written as:

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

where $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, l = 1, \dots, n$.

We extend the model in the same way. Our analysis of variance has three main effects, three two-way interactions, a three-way interaction and error. If this were conducted as a

Completely Randomized Design experiment, each of the $a \times b \times c$ treatment combinations would be randomly assigned to n of the experimental units.

Sample Size Determination [Section 5.3.5]

We first consider the two factor case where $N = a \times b \times n$, (n = the number of replicates per cell). The non-centrality parameter for calculating sample size for the A factor is:

$$\Phi^2 = (nb \times D^2) / (2a \times \sigma^2)$$

where D is the difference between the maximum of $\bar{\mu}_{i\cdot}$ and the minimum of $\bar{\mu}_{i\cdot}$, and where b is the number of observations in each level of factor A .

Actually at the beginning of our design process we should decide how many observations we should take, if we want to find a difference of D , between the maximum and the minimum of the true means for the factor A . There is a similar equation for factor B .

$$\Phi^2 = (na \times D^2) / (2b \times \sigma^2)$$

where na is the number of observations in each level of factor B .

In the two factor case this is just an extension of what we did in the one factor case. But now we have the marginal means benefiting from a number of observations per cell and the number of levels of the other factor. In this case we have n observations per cell, and we have b cells. So, we have nb observations.

5.2 - Another Factorial Design Example - Cloth Dyes

Consider the cloth dyes data from Problem 5.19 in the text:

For each combination of time, temperature and operator, there are three observations. Now we have a case where there are three factors and three observations per cell. Let's run this model in Minitab.

General Linear Model: Score versus Temperature, Cycle Time, Operator

Factor	Type	Levels	Values
Temperature	fixed	2	300, 350
Cycle Time	fixed	3	40, 50, 60
Operator	fixed	3	1, 2, 3

Analysis of Variance for Score, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Temperature	1	50.074	50.074	50.074	15.28	0.000
Cycle Time	2	436.000	436.000	218.000	66.51	0.000
Operator	2	261.333	261.333	130.667	39.86	0.000
Temperature*Cycle Time	2	78.815	78.815	39.407	12.02	0.000
Temperature*Operator	2	11.259	11.259	5.630	1.72	0.194
Cycle Time*Operator	4	355.667	355.667	88.917	27.13	0.000
Temperature*Cycle Time*Operator	4	46.185	46.185	11.546	3.52	0.016
Error	36	118.000	118.000	3.278		
Total	53	1357.333				

S = 1.81046 R-Sq = 91.31% R-Sq(adj) = 87.20%

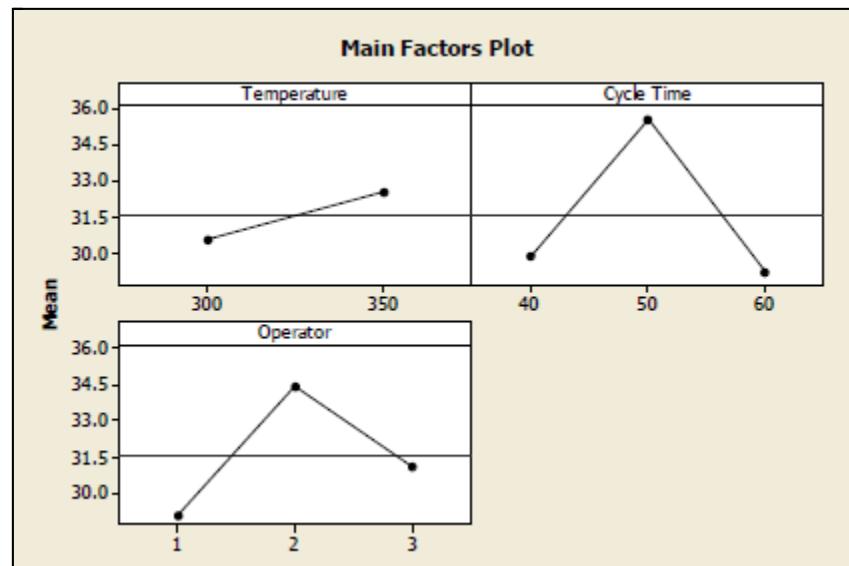
Unusual Observations for Score

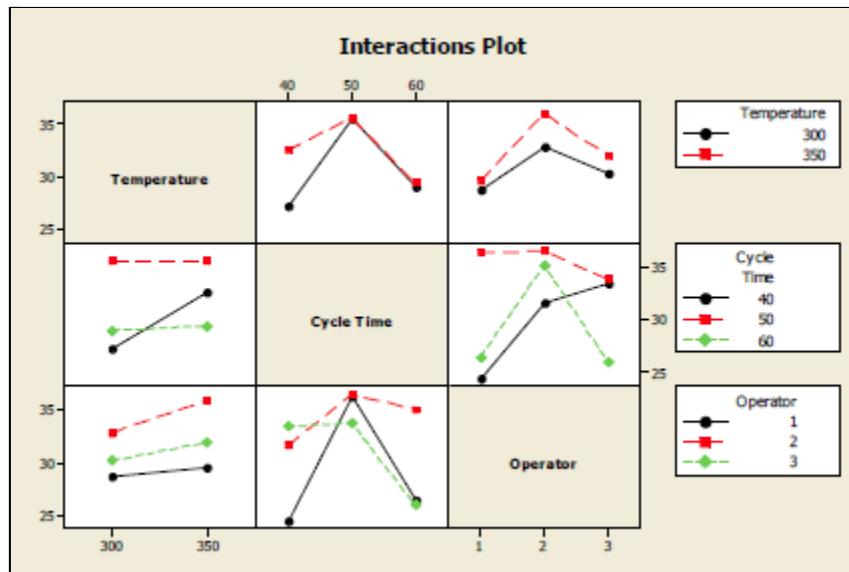
Obs	Score	Fit	SE Fit	Residual	St Resid
5	34.0000	37.0000	1.0453	-3.0000	-2.03 R
46	28.0000	25.0000	1.0453	3.0000	2.03 R

R denotes an observation with a large standardized residual.

The ANOVA table shows us that the main effects due to cycle time, operator, and temperature are all significant. The two-way interactions for cycle time by operator and cycle time by temperature are significant. But the operator by temperature is not significant but the dreaded three-way interaction is significant. What does it mean when a three-way interaction is significant?

Let's take a look at the factor plots:





These interaction plots show us the three sets of two-way cell means, each of the three are plotted in two different ways. This is a useful plot to try to understand what is going on. These are all the two-way plots.

Typically a three-way interaction would be plotted as two panels... showing how the two-way interactions differ across the levels of the third factor. Minitab does not do that for you automatically.

Let's think about how this experiment was done. There are three observations for each combination of factors. Are they actually separate experimental units or are they simply three measurements on the same experimental unit? If they are simply three measurements on the same piece of cloth that was all done in the same batch, for instance, then they are not really independent. If this is the case, then another way to look at this data would be to average those replications. In this case there is only 1 observation for each treatment, so that there would be no d.f. for error. However, the way the problem is presented in the text, they appear to have been treated independently and thus are true replicates, leading to 36 d.f. for error.

You could also think about the operator not as a factor that you're interested in but more as a block factor, i.e. a source of variation that we want to remove from the study. What we're really interested in is the effect of temperature and time on the process of dyeing the cloth. In this case we could think about using the operator as a block effect. Running the analysis again, now we get the same plot but look at the ANOVA table: now the interactions related to operator have been pooled as a part of the error. So the residual error term now has $2 + 4 + 4 + 36 = 46$ df for error. Note also that if you do use the operator as a treatment factor, it probably should be considered random. In this case, you would probably want to consider the 2 and 3-way interactions involving operator to be random effects. Experiments in which some factors are fixed and others are random are called mixed effects experiments. The analysis of mixed effects experiments is discussed in Chapter 13.

General Linear Model: Score versus Temperature, Cycle Time, Operator						
Factor Type Levels Values						
Temperature fixed 2 300, 350						
Cycle Time fixed 3 40, 50, 60						
Operator fixed 3 1, 2, 3						
Analysis of Variance for Score, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Temperature	1	50.07	50.07	50.07	4.34	0.043
Cycle Time	2	436.00	436.00	218.00	18.88	0.000
Operator	2	261.33	261.33	130.67	11.32	0.000
Temperature*Cycle Time	2	78.81	78.81	39.41	3.41	0.041
Error	46	531.11	531.11	11.55		
Total	53	1357.33				
S = 3.39792 R-Sq = 60.87% R-Sq(adj) = 54.92%						
Unusual Observations for Score						
Obs	Score	Fit	SE Fit	Residual	St Resid	
28	23.0000	30.1111	1.3079	-7.1111	-2.27	R
48	39.0000	32.1111	1.3079	6.8889	2.20	R
R denotes an observation with a large standardized residual.						

What this points out is the importance of distinguishing what is a block factor, and which are the treatment factors when you have a multifactor experimental design. This should be apparent from how the experiment was conducted, but if the data are already collected when you are introduced to the problem, you need to inquire carefully to understand how the experiment was actually conducted to know what model to use in the analysis.

Let's take a look two examples using this same dataset using **Minitab v.16**. First we will analyze the quantitative factors involved, *Cycle Time* and *Temperature* and as though they were qualitative - simply nominal factors.



[1]

Next, using Operator as a block we will now use **Minitab v.16** to treat the quantitative factors as qualitative factors and apply these in a regression analysis.



[2]

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/26>

Links:

- [1] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson05/L05_clothdyes_viewlet_swf.html','l05_clothdyes', 724, 708\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson05/L05_clothdyes_viewlet_swf.html','l05_clothdyes', 724, 708);)
- [2] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson05/L05_clothdyes_02_viewlet_swf.html','l05_clothdyes_02', 724, 708\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson05/L05_clothdyes_02_viewlet_swf.html','l05_clothdyes_02', 724, 708);)



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 6: The 2^k Factorial Design

Lesson 6: The 2^k Factorial Design

Introduction

The 2^k designs are a major set of building blocks for many experimental designs. These designs are usually referred to as screening designs. The 2^k refers to designs with k factors where each factor has just two levels. These designs are created to explore a large number of factors, with each factor having the minimal number of levels, just two. By screening we are referring to the process of screening a large number of factors that might be important in your experiment, with the goal of selecting those important for the response that you're measuring. We will see that k can get quite large. So far we have been looking at experiments that have one, two or three factors, maybe a blocking factor and one or two treatment factors, but when using screening designs k can be as large as 8, 10 or 12. For those of you familiar with chemical or laboratory processes, it would not be hard to come up with a long list of factors that would affect your experiment. In this context we need to decide which factors are important.

In these designs we will refer to the levels as high and low, +1 and -1, to denote the high and the low level of each factor. In most cases the levels are quantitative, although they don't have to be. Sometimes they are qualitative, such as gender, or two types of variety, brand or process. In these cases the +1 and -1 are simply used as labels.

Learning objectives & outcomes

Upon completion of this lesson, you should be able to do the following:

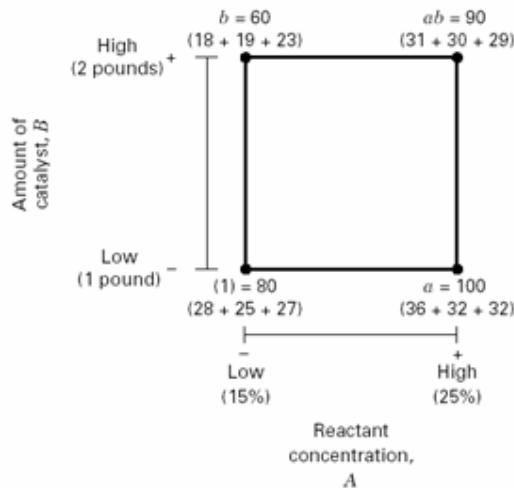
- The idea of 2-level Factorial Designs as one of the most important screening designs
- Defining a “contrast” which is an important concept and how to derive Effects and Sum of Squares using the Contrasts
- Process of analyzing Unreplicated or Single replicated factorial designs, and
- How to use Transformation as a tool in dealing with inadequacy of either variance homogeneity or normality of the data as major hypothetical assumptions.

6.1 - The Simplest Case

The simplest case is 2^k where $k = 2$. We will define a new notation which is known as Yates notation. We will refer to our factors using the letters A, B, C, D, etc. as arbitrary labels of the factors. In the chemical process case A is the concentration of the reactant and B is the amount of catalyst, both of which are quantitative. The yield of the process is our response variable.

Since there are two levels of each of two factors, 2^k equals four. Therefore, there are four treatment combinations and the data are given below:

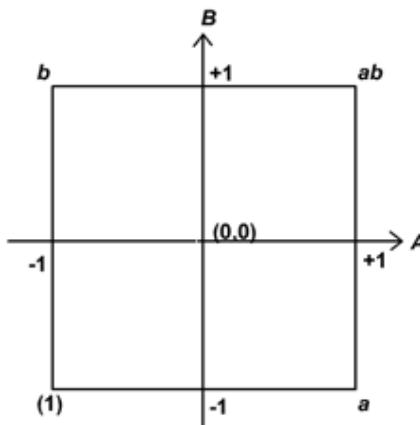
Loading [MathJax]/extensions/MathZoom.js



You can see that we have 3 observations at each of $4 = 2^k$ combinations for $k = 2$. So we have $n = 3$ replicates.

A	B	Yates Notation
-	-	(1)
+	-	a
-	+	b
+	+	ab

The table above gives the data with the factors coded for each of the four combinations and below is a plot of the region of experimentation in two dimensions for this case.



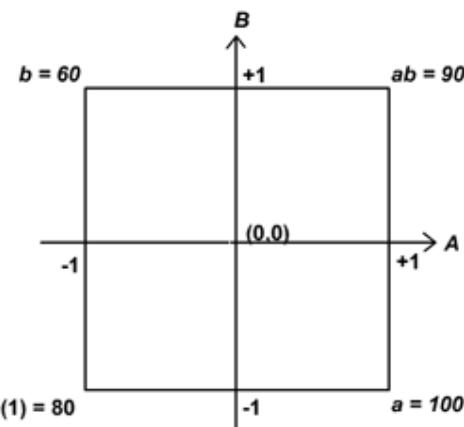
The Yates notation used for denoting the factor combinations is as follows:

We use "(1)" to denote that both factors are at the low level, "a" for when A is at its high level and B is at its low level, "b" for when B is at its high level and A is at its low level, and "ab" when both A and B factors are at their high level.

The use of this Yates notation indicates the high level of any factor simply by using the small letter of that level factor. This notation actually is used for two purposes. One is to denote the

Loading [MathJax]/extensions/MathZoom.js

total sum of the observations at that level. In the case below $b = 60$ is the sum of the three observations at the level b .



This shortcut notation, using the small letters, shows which level for each of our k factors we are at just by its presence or absence.

We will also connect this to our previous notation for the two factor treatment design:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + e_{ijk}$$

What is the primary goal of these screening experiments?

The goal is to decide which of these factors is important. After determining which factors are important, then we will typically plan for a secondary experiment where the goal is to decide what level of the factors gives us the optimal response. Thus the screening 2^k experiment is the first stage, generally, of an experimental sequence. In the second stage one is looking for a response surface or an experiment to find the optimal level of the important factors.

Estimation of Factors Effects (in the Yates tradition)

The definition of an effect in the 2^k context is the difference in the means between the high and the low level of a factor. From this notation A is the difference between the averages of the observations at the high level of A minus the average of the observations at the low level of A .

Therefore, $A = \bar{y}_{A^+} - \bar{y}_{A^-}$, in the example above:

$$A = 190/6 - 140/6 = 50/6 = 8.33$$

Similarly, $B = \bar{y}_{B^+} - \bar{y}_{B^-}$, is the similar only looking in the other direction. In our example:

$$B = 150/6 - 180/6 = 25 - 30 = -5$$

and finally, $AB = \frac{ab+(1)}{2n} - \frac{a+b}{2n}$

$$AB = [(90 + 80)/6 - (100 + 60)/6] = 10/6 = 1.67$$

Therefore in the Yates notation, we define an effect as the difference in the means between the high and the low levels of a factor whereas in previous models we defined an effect as the

Loading [MathJax]/extensions/MathZoom.js

coefficients of the model, which are the differences between the marginal mean and the overall mean. To restate this, in terms of A, the A effect is the difference between the means at the high levels of A versus the low levels of A, whereas the coefficient, a_i , in the model is the difference between the marginal mean and the overall mean. So the Yates "effect" is twice the size of the estimated coefficient a_i in the model, which is also usually called the effect of the factor A.

The confusion is all in the notation used in the definition.

Let's look at another example in order to reinforce your understanding of the notation for these types of designs. Here is an example in three dimensions, with factors A, B and C. Below is a figure of the factors and levels as well as the table representing this experimental space.

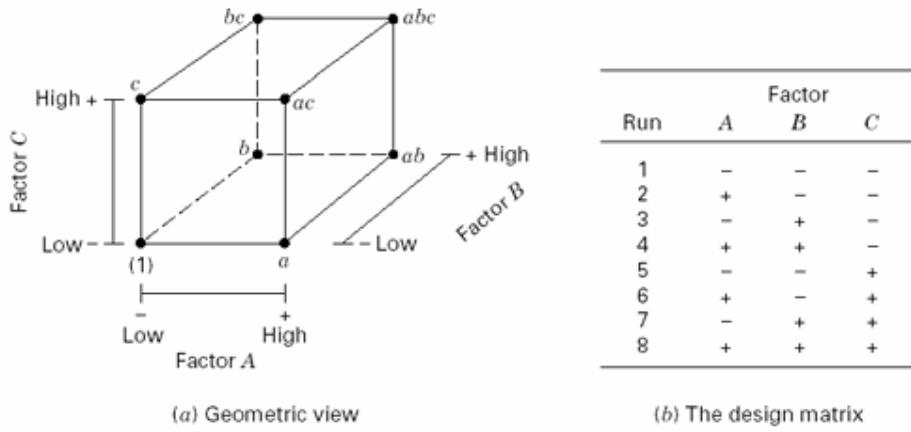


Figure 6-4 The 2^3 factorial design.

In table 6.4 you can see the eight points coded by the factor levels +1 and -1. This example has two replicates so $n = 2$. Notice that the Yates notation is included as the total of the two replicates.

One nice feature of the Yates notation is that every column has an equal number of pluses and minuses so these columns are contrasts of the observations. For instance, take a look at the A column. This column has four pluses and four minuses, therefore, the A effect is a contrast defined on page 216.

This is the principle that gives us all sorts of useful characterizations in these 2^k designs.

In the example above the A, B and C each are defined by a contrast of the data observation totals. Therefore you can define the contrast AB as the product of the A and B contrasts, the contrast AC by the product of the A and C contrasts, and so forth.

Therefore all the two-way and three-way interaction effects are defined by these contrasts. The product of any two gives you the other contrast in that matrix. (See Table 6.3 in the text.)

From these contrasts we can define the effect of A, B, and C, using these coefficients. The general form of an effect for k factors is:

$$\text{Effect} = (1/2^{(k-1)}n) [\text{contrast of the totals}]$$

The sum of the products of the contrast coefficients times the totals will give us the estimate of the effects. See equations (6-11), (6-12), and (6-13).

Loading [MathJax]/extensions/MathZoom.js

We can also write the variance of the effect using the general form used previously. This would be:

$$\begin{aligned} \text{Variance(Effect)} &= [1/(2^{k-1}n)^2]V(\text{contrast}), \text{ or} \\ &= [1/(2^{k-1}n)^2]2^k n\sigma^2 \\ &= \sigma^2 / 2^{k-2}n \end{aligned}$$

Also, we can write the sum of squares for the effects which looks like:

$$\text{SS(effect)} = (\text{contrast})^2 / 2^k n$$

To summarize what we have learned in this lesson thus far, we can write a contrast of the totals which defines an effect, we can estimate the variance for this effect and we can write the sum of squares for an effect. We can do this very simply using Yates notation which historically has been the value of using this notation.

6.2 - Estimated Effects and the Sum of Squares from the Contrasts

How can we apply what we learned in the preceding section?

In general for 2^k factorials the effect of each factor and interaction is:

$$\text{Effect} = (1/2^{k-1}n) [\text{contrast of the totals}]$$

We also defined the variance as follows:

$$\text{Variance(Effect)} = \sigma^2 / 2^{k-2}n$$

The true but unknown residual variance σ^2 , which is also called the within cell variance, can be estimated by the MSE.

If we want to test an effect, for instance, say A = 0, then we can construct a t -test which is the effect over the square root of the estimated variance of the effect as follows:

$$t^* = \frac{\text{Effect}}{\sqrt{\frac{\text{MSE}}{n2^{k-2}}}} \sim t(2^k(n-1))$$

where \sim means that it has a t distribution with $2^k(n-1)$ degrees of freedom.

Finally, here is the equation for the sum of squares due to an effect to complete the story here:

$$\text{SS(Effect)} = (\text{contrast of totals})^2 / 2^k n$$

Where does all of this come from? Each effect in a 2^k model has one degree of freedom. In the simplest case we have two main effects and an interaction. They each have 1 degree of freedom. So the t statistic is the ratio of the effect over its estimated standard error (standard

Loading [MathJax]/extensions/MathZoom.js

deviation of the effect). You will recall that if you have a t statistic with v degrees of freedom and square it, you get an F distribution with one and v degrees of freedom.

$$t^2(v) = F(1, v)$$

We can use this fact to confirm the formulas just developed. We see that the

$$(t^*(v))^2 = \frac{(Effect)^2}{MSE/n2^{k-2}}$$

and from the definition of an F -test, when the numerator has 1 degree of freedom:

$$F(1, v) = \frac{SS(Effect)/1}{MSE} = \frac{(contrast)^2}{2^k n(MSE)}$$

But from the definition of an Effect, we can write $(Effect)^2 = (contrast)^2 / (n2^{k-1})^2$ and thus $F(1, v) = (t^*(v))^2$ which you can show by some algebra or by calculating an example.

Hint: Multiply $F(1, v)$ by $(2^{k-1}n)^2 / (2^{k-1}n)^2$ and simplify.

Once you have these contrasts, you can easily calculate the effect, you can calculate the estimated variance of the effect and the sum of squares due to the effect as well.

Creating a Factorial Design in Minitab

Let's use Minitab to help us create a factorial design and then add data so that we can analyze it. Click on the 'Inspect' button to walk through this process using **Minitab v.16**. The data come from Figure 6.1.



In Minitab we use the software under Stat > Design of Experiments to create our full factorial design. We will come back to this command another time to look at fractional factorial and other types of factorial designs.

In the example that was shown above we did not randomize the runs but kept them in standard order for the purpose of the seeing more clearly the order of the runs. In practice you would want to randomize the order of run when you are designing the experiment.

Once we have created a factorial design within the Minitab worksheet we then need to add the response data so that the design can be analyzed. These response data, Yield, are the individual observations not the totals. So, we again go to the Stat >> DOE >> Factorial menu where we will analyze the data set from the factorial design.

We began with the full model with all the terms included, both the main effects and all of the interactions. From here we were able to determine which effects were significant and should remain in the model and which effects were not significant and can be removed to form a simpler reduced model.

Loading [MathJax]/extensions/MathZoom.js

A Second Example - The Plasma Etch Experiment

Similar to the previous example, in this second industrial process example we have three factors, A equals Gap, B = Flow, C = Power and our response y = Etch Rate. (The data are from Table 6-4 in the text.) Once again in Minitab we will create a similar layout for a full factorial design for three factors with two replicates which gives us 16 observations. Next, we add the response data, Etch Rate, to this worksheet and analyze this data set. These are the results we get:

Factorial Fit: EtchRate versus A, B, C						
Estimated Effects and Coefficients for EtchRate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		776.06	11.87	65.41	0.000	
A	-101.62	-50.81	11.87	-4.28	0.003	
B	7.37	3.69	11.87	0.31	0.764	
C	306.13	153.06	11.87	12.90	0.000	
A*B	-24.88	-12.44	11.87	-1.05	0.325	
A*C	-153.63	-76.81	11.87	-6.47	0.000	
B*C	-2.12	-1.06	11.87	-0.09	0.931	
A*B*C	5.62	2.81	11.87	0.24	0.819	
 S = 47.4612 R-Sq = 96.61% R-Sq(adj) = 93.64%						
 Analysis of Variance for EtchRate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	416378	416378	138793	61.62	0.000
2-Way Interactions	3	96896	96896	32299	14.34	0.001
3-Way Interactions	1	127	127	127	0.06	0.819
Residual Error	8	18020	18020	2253		
Pure Error	8	18021	18021	2253		
Total	15	531421				

The analysis of variance shows the individual effects and the coefficients, (which are half of the effects), along with the corresponding t -tests. Now we can see from these results that the A effect and C effect are highly significant. The B effect is not significant. In looking at the interactions, AB , is not significant, BC is not significant, and the ABC are not significant. However the other interaction, AC is significant.

This is a nice example to illustrate the purpose of a screening design. You want to test a number of factors to see which ones are important. So what have we learned here? Two of these factors are clearly important, A and C . But B appears not to be important either as a main effect or within any interaction. It simply looks like random noise. B was the rate of gas flow across the edging process and it does not seem to be an important factor in this process, at least for the levels of the factor used in the experiment.

The analysis of variance summary table results show us that the main effects overall are significant. That is because two of them, A and C , are highly significant. The two-way interactions overall are significant. That is because one of them is significant. So, just looking at this summary information wouldn't tell us what to do except that we could drop the 3-way interaction.

Now we can go back to Minitab and use the Analyze command under Design of Experiments and we can remove all the effects that were seemingly not important such as any term having to do with B in the model. In running this new reduced model we get:

Factorial Fit: EtchRate versus A, C						
Estimated Effects and Coefficients for EtchRate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		776.06	10.42	74.46	0.000	
A		-101.62	-50.81	10.42	-4.88	0.000
C		306.13	153.06	10.42	14.69	0.000
A*C		-153.63	-76.81	10.42	-7.37	0.000
 S = 41.6911 R-Sq = 96.08% R-Sq(adj) = 95.09%						
Analysis of Variance for EtchRate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	2	416161	416161	208080	119.71	0.000
2-Way Interactions	1	94403	94403	94403	54.31	0.000
Residual Error	12	20858	20858	1738		
Pure Error	12	20858	20858	1738		
Total	15	531421				

For this model, all three terms are significant.

6.3 - Unreplicated 2^k Factorial Designs

These are 2^k factorial designs with **one observation** at each corner of the "cube". An unreplicated 2^k factorial design is also sometimes called a "single replicate" of the 2^k experiment.

You would find these types of designs used where k is very large or the process for instance is very expensive or takes a long time to run. In these cases, for the purpose of saving time or money, we want to run a screening experiment with as few observations as possible. When we introduced this topic we wouldn't have dreamed of running an experiment with only one observation. As a matter of fact, the general rule of thumb is that you would have at least two replicates. This would be a minimum in order to get an estimate of variation - but when we are in a tight situation, we might not be able to afford this due to time or expense. We will look at an example with one observation per cell, no replications, and what we can do in this case.

Where are we going with this? We have first discussed factorial designs with replications, then factorial designs with one replication, now factorial designs with one observation per cell and no replications, which will lead us eventually to fractional factorial designs. This is where we are headed, a steady progression to designs with more and more factors, but fewer observations and less direct replication.

Unreplicated 2^k Factorial Designs

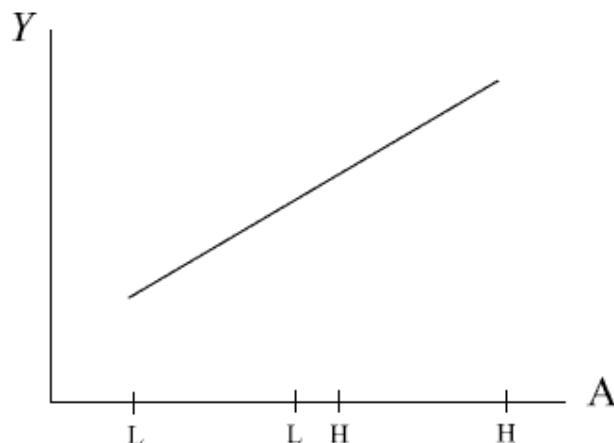
Let's look at the situation where we have one observation per cell. We need to think about where the variation occurs within this design. These designs are very widely used. However,

there are risks...if there is only one observation at each corner, there is a high chance of an unusual response observation spoiling the results. What about an outlier? There would be no way to check if this was the case and thus it could distort the results fairly significantly. You have to remind yourself that these are not the definitive experiments but simply just screening experiments to determine which factors are important.

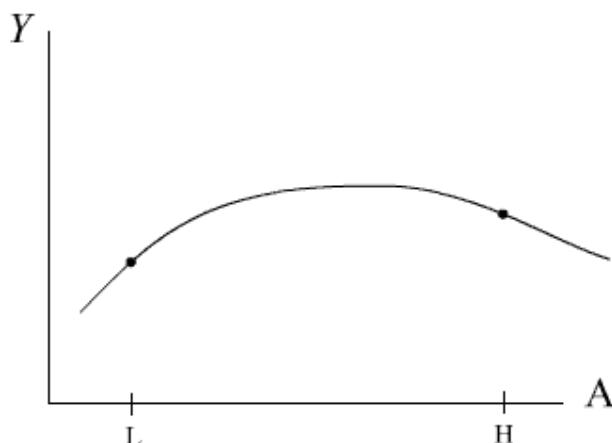
In these experiments one really cannot model the "noise" or variability very well. These experiments cannot really test whether or not the assumptions are being met - again this is another shortcoming, or the price of the efficiency of these experiment designs.

Spacing of Factor Levels in the Unreplicated 2^k Factorial Designs

When choosing the levels of your factors, we only have two options - low and high. You can pick your two levels low and high close together or you can pick them far apart. As most of you know from regression the further apart your two points are the less variance there is in the estimate of the slope. The variance of the slope of a regression line is inversely related to the distance between the extreme points. You can reduce this variance by choosing your high and low levels far apart.



However, consider the case where the true underlying relationship is curved, i.e., more like this:



... and you picked your low and high level as illustrated above, then you would have missed capturing the true relationship. Your conclusion would probably be that there is no effect of that

factor. You need to have some understanding of what your factor is to make a good judgment about where the levels should be. In the end, you want to make sure that you choose levels in the region of that factor where you are actually interested and are somewhat aware of a functional relationship between the factor and the response. This is a matter of knowing something about the context for your experiment.

How do we analyze our experiment when we have this type of situation? We must realize that the lack of replication causes potential problems in statistical testing:

- Replication provides an estimate of "pure error" (a better phrase is an internal estimate of error), and
- With no replication, fitting the full model results in zero degrees of freedom for error.

Potential solutions to this problem might be:

- Pooling high-order interactions to estimate error, (something we have done already in randomized block design),
- Normal probability plotting of effects (Cuthbert and Daniels, 1959), and/or
- Dropping entire factors from the model and other methods.

Example of an Unreplicated 2^k Design

The following 2^4 factorial (Example 6-2 in the text) was used to investigate the effects of four factors on the filtration rate of a resin for a chemical process plant. The factors are A = temperature, B = pressure, C = mole ratio (concentration of chemical formaldehyde), D = stirring rate. This experiment was performed in a pilot plant.

Here is the dataset for this Resin Plant experiment. You will notice that all of these factors are quantitative.

Table 6-10 Pilot Plant Filtration Rate Experiment

Run Number	Factor				Run Label	Filtration Rate (gal/h)
	A	B	C	D		
1	-	-	-	-	(1)	45
2	+	-	-	-	a	71
3	-	+	-	-	b	48
4	+	+	-	-	ab	65
5	-	-	+	-	c	68
6	+	-	+	-	ac	60
7	-	+	+	-	bc	80
8	+	+	+	-	abc	65
9	-	-	-	+	d	43
10	+	-	-	+	ad	100
11	-	+	-	+	bd	45
12	+	+	-	+	abd	104
13	-	-	+	+	cd	75
14	+	-	+	+	acd	86
15	-	+	+	+	bcd	70
16	+	+	+	+	abcd	96

Notice also the use of the Yates notation here that labels the treatment combinations where the high level for each factor is involved. If only A is high then that combination is labeled with the small letter a . In total, there are 16 combinations represented.

Here is a visual representation of this - it would be impossible to show this in a 4 dimensional cube but here are two cubes which attempt to do the same thing.

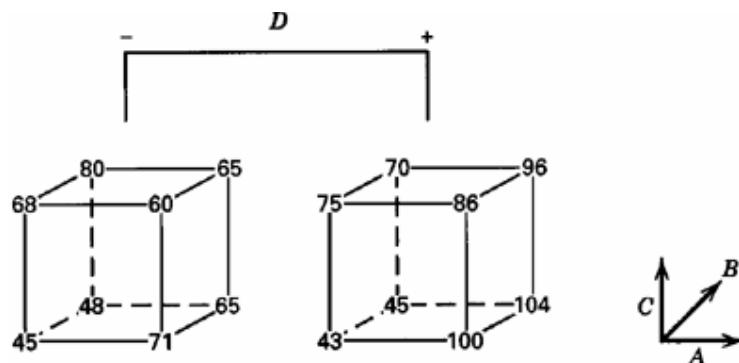


Figure 6-10 Data from the pilot plant filtration rate experiment for Example 6-2.

...

Sequential Procedure for Strategically Finding a Model

Let's use the dataset ([Ex6-2.MTW](#) [2]) and work at finding a model for this data with Minitab...



Even with just one observation per cell, by carefully looking at the results we can come to some understanding as to which factors are important. We do have to take into account that these actual p -values are not something that you would consider very reliable because you are fitting this sequence of models, i.e., fishing for the best model. We have optimized with several decisions that invalidates the actual p -value of the true probability that this could have occurred by chance.

This is one approach to assume that some interactions are not important and use this to test lower order terms of the model and finally come up with a model that is more focused. Based on this for this example that we have just looked at, we can conclude that following factors are important, A , C , D , (of the main effects) and AC and AD of the two-way interactions.

Now I suggest you try this procedure and then go back and check to see what the final model looks like. Here is what we get when we drop factor B and all the interactions that we decided were not important:

Factorial Fit: Rate versus A, C, D						
Estimated Effects and Coefficients for Rate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		70.063	1.104	63.44	0.000	
A		21.625	10.812	1.104	9.79	0.000
C		9.875	4.938	1.104	4.47	0.001
D		14.625	7.312	1.104	6.62	0.000
A*C		-18.125	-9.062	1.104	-8.21	0.000
A*D		16.625	8.313	1.104	7.53	0.000

S = 4.41730 R-Sq = 96.60% R-Sq(adj) = 94.89%

Analysis of Variance for Rate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	3116.19	3116.19	1038.73	53.23	0.000
2-Way Interactions	2	2419.62	2419.62	1209.81	62.00	0.000
Residual Error	10	195.12	195.12	19.51		
Lack of Fit	2	15.62	15.62	7.81	0.35	0.716
Pure Error	8	179.50	179.50	22.44		
Total	15	5730.94				

The important factors didn't change much here. However, we have slightly higher degrees of freedom for error. But now what the design looks like, by having dropped B totally, is that we now have a 2^3 design with 2 replicates per cell. We have moved from a four factor with one observation per cell, to a three factor with two observations per cell.

So, we have looked at two strategies here. The first is to take a higher order interactions out of the model and use them as the estimate of error. Next, what we did at the end of the process is drop that factor entirely. If a particular factor in the screening experiment turns out to be not important either as a main effect or as part of any interaction we can remove it. This is the second strategy, and for instance in this example we took out factor *B* completely from the analysis.

Graphical Approaches to Finding a Model

Let's look at some more procedures - this time graphical approaches for us to look at our data in order to find the best model. This technique is really cool. Get a cup of coffee and click:



[4]

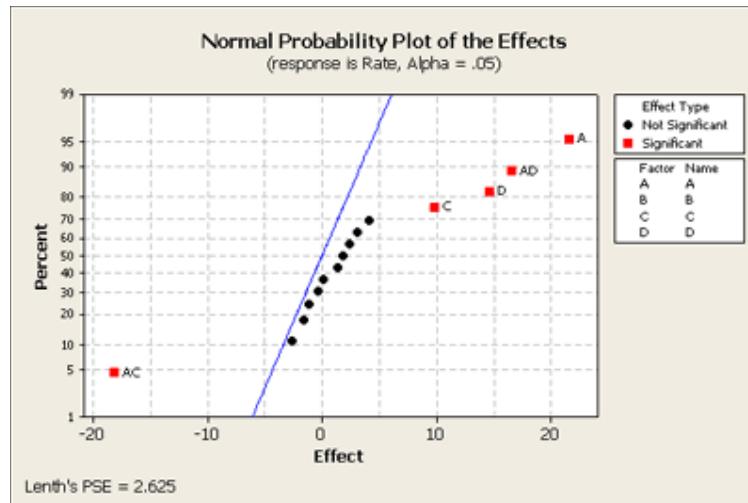
Normal Probability Plot for the Effects

Having included all the terms back into a full model we have shown how to produce a normal plot. Remember that all of these effects are 1 degree of freedom contrasts of the original data, each one of these is a linear combination of the original observations, which are normally distributed with constant variance. Then these 15 linear combinations or contrasts are also normally distributed with some variance. If we assume that none of these effects are significant, the null hypothesis for all of the terms in the model, then we simply have 15 normal random variables, and we will do a normal random variable plot for these. That is what we will ask

Minitab to plot for us. We get a normal probability plot, not of the residuals, not of the original

Load Minitab extensions Minitab 2010

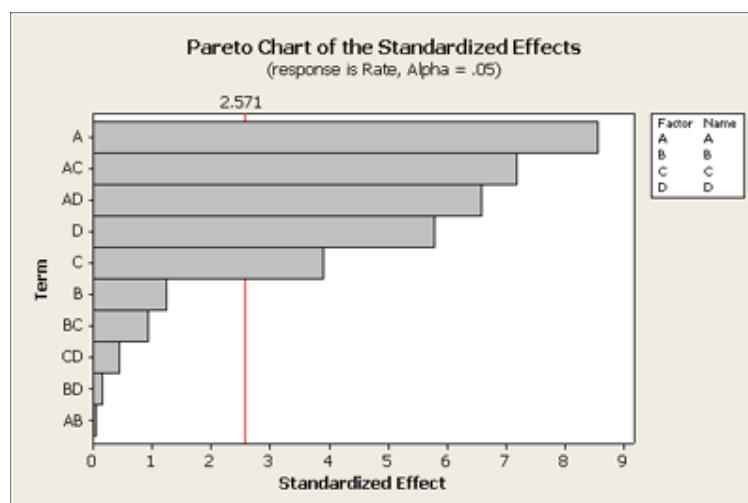
observations but of the effects. We have plotted these effects against what we would expect if they were normally distributed.



In the middle - the points in black, they are pretty much in a straight line - they are following a normal distribution. In other words, their expectation or percentile is proportionate to the size of the effect. The ones in red are like outliers and stand away from the ones in the middle and indicate that they are not just random noise but there must be an actual affect. Without making any assumptions about any of these terms this plot is an overall test of the hypothesis based on simply assuming all of the effects are normal. This is a very helpful - a good quick and dirty first screen - or assessment of what is going on in the data, and this corresponds exactly with what we found in our earlier screening procedures.

The Pareto Plot

Let's look at another plot - the Pareto plot. This is simply a plot that can quickly show you what is important. It looks at the size of the effects and plots the effect size on a horizontal axis ranked from largest to smallest effect.



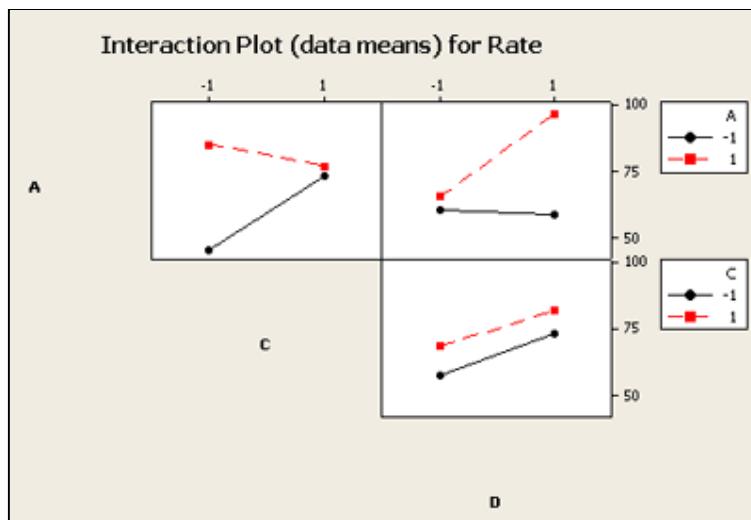
Having dropped some of the terms out of the model, for instance the three and four way interactions, Minitab plots the remaining effects, but now it is the standardized effect. Basically it

Loading [MathJax]/extensions/MathZoom.js

is plotting the t -value, the effect over its standard deviation and then plotting it in ranked order. It also displays the t critical point as a red line at alpha = 0.05.

Effects and Interaction Plots

Another Minitab command that we can take a look at is the subcommand called Factorial Plots. Here we can create plots for main effects telling Minitab which factors you want to plot. As well you can plot two-way interactions. Here is a plot of the interactions (which are more interesting to interpret), for the example we've been looking at:

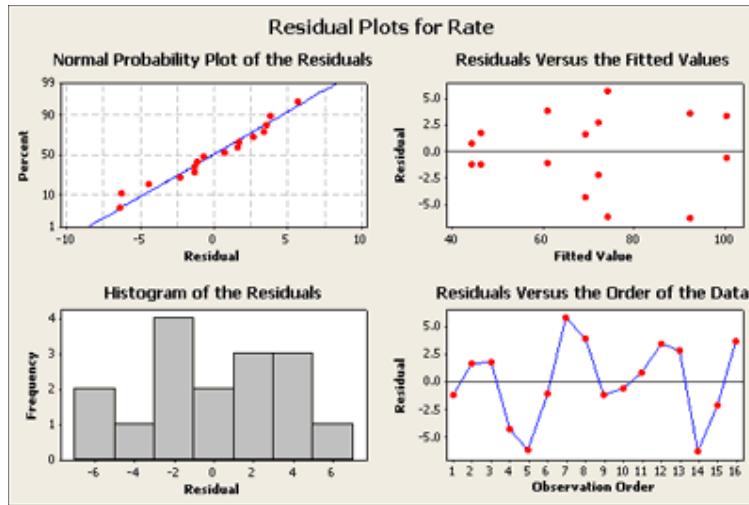


You can see that the C and D interaction plot the lines are almost parallel and therefore do not indicate interaction effects that are significant. However the other two combinations, A and C and A and D, indicate that significant interaction exists. If you just looked at the main effects plot you would likely miss the interactions that are obvious here.

Checking Residuals Using Minitab's Four in One Plot

We have reduced the model to include only those terms that we found were important. Now we want to check the residuals in order to make sure that our assumptions are not out of line with any conclusions that we are making. We can ask Minitab to produce a Four in One residuals plot which, for this example, looks like this:

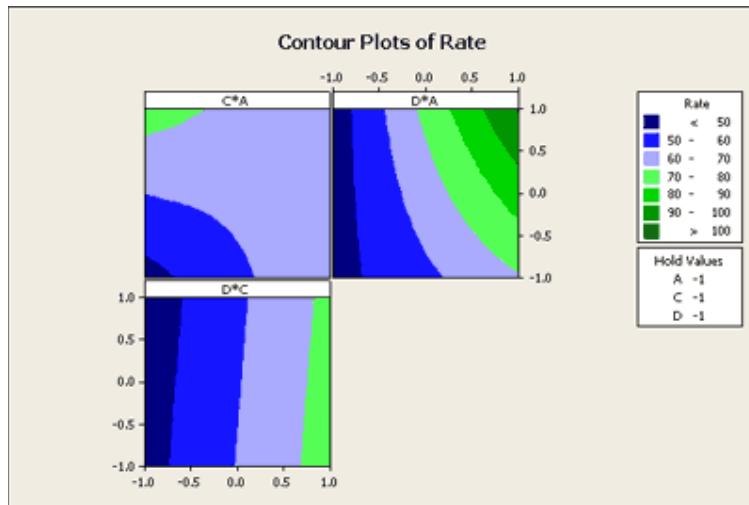
Loading [MathJax]/extensions/MathZoom.js



In visually checking the residuals we can see that we have nothing to complain about. There does not seem to be any great deviation in the normal probability plot of the residuals. There's nothing here that is very alarming and it seems acceptable. In looking at the residuals versus the fitted values plot in the upper right of this four in one plot - except for the lower values on the left where there are smaller residuals and you might be somewhat concerned here, the rest do not set off any alarms - but we will come back to this later.

Contour and Surface Plots

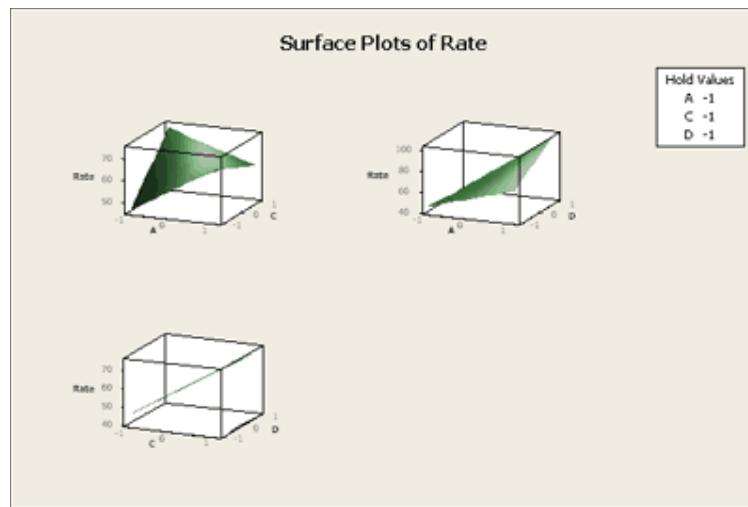
We may also want contour plots of all pairs of our numeric factors. These can be very helpful to understand and present the relationship between several factors on the response. The contour plots below for our example show the color coded average response over the region of interest. The effect of these changes in colors is to show the twist in the plane.



In the D*C plot area you can see that there is no curvature in the colored areas, hence no evidence of interaction. However, if you look at C*A display you can see that if C is low you get a dramatic change. If C is high it makes very little difference. In other words, the response due to A depends on the level of C. This is what the interaction means and it shows up nicely in this contour plot.

Loading [MathJax]/extensions/MathZoom.js

Finally, we can also ask Minitab to give us a surface plot. We will set this up the same way in Minitab and this time Minitab will show the plot in three dimensions, two variables at a time.

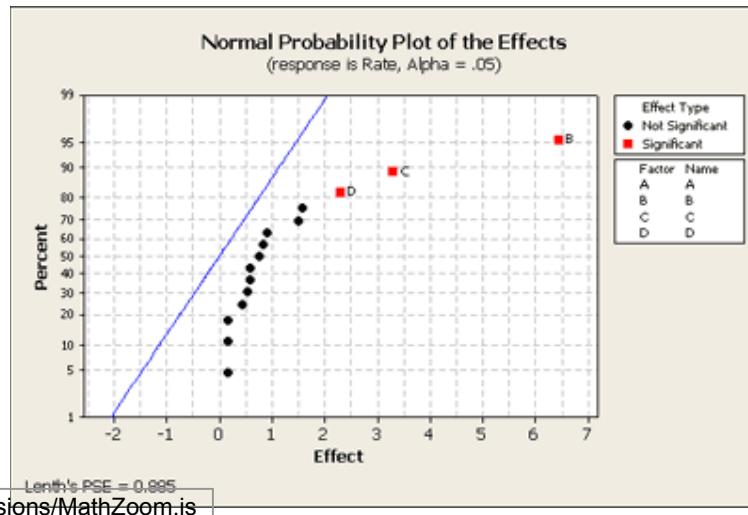


The surface plot shows us the same interaction effect in three dimensions in the twisted plane. This might be a bit easier to interpret. In addition you can ask Minitab to provide you with 3-D graphical tools that will allow you to grab these boxes and twist them around so that you can look at these boxes in space from different perspectives. Pretty cool! Give it a try. These procedures are all 'illustrated in the "Inspect" Flash movie at the beginning of this section.

Another Example - The Drilling Example 6.3

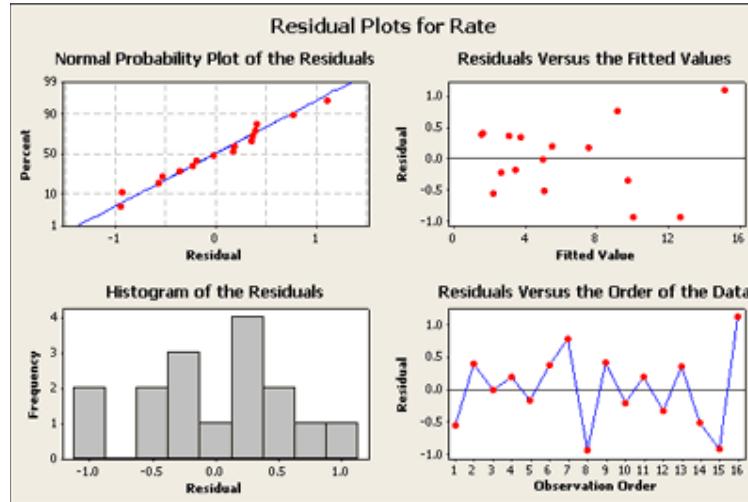
This is another fairly similar example to the one we just looked at. This drilling example (Example 6-3) is a 2^4 design - again, the same design that we looked at before. It is originally from C. Daniel, 1976. It has four factors, A = Drill load, B = Flow of a lubricant, C = Speed of drill, D = Type of mud, Y is the Response - the advance rate of the drill, (how fast can you drill an oil or gas well?).

We've used Minitab to create the factorial design and added the data from the experiment into the Minitab worksheet. First, we will produce a normal probability plot of the effects for this data with all terms included in a full model.

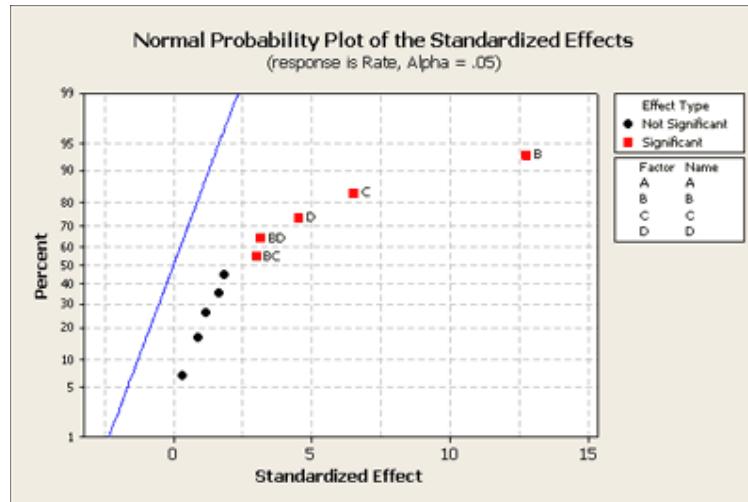


Here's what it looks like. It shows a strange pattern! No negative and all positive effects. All of the black dots are in fairly straight order except for perhaps the top two. If we look at these closer we can see that these are the BD and the BC terms, in addition to B, C, and D as our most important terms. Let's go back to Minitab and take out of our model the higher order interactions, (i.e. the 3-way and 4-way interactions), and produce this plot again (see below) just to see what we learn.

The normal probability plot of residuals looks okay. There is a gap in the histogram of other residuals but it doesn't seem to be a big problem.



When we look at the normal probability plot below, created after removing 3-way and 4-way interactions, we can see that now BD and BC are significant.



We can also see this in the statistical output of this model as shown below:

Loading [MathJax]/extensions/MathZoom.js

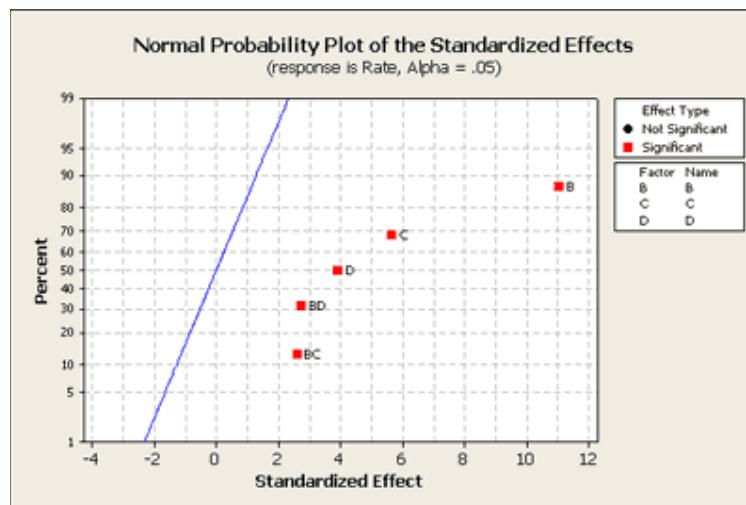
Factorial Fit: Rate versus A, B, C, D						
Estimated Effects and Coefficients for Rate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		6.15250	0.2521	24.40	0.000	
A		0.91750	0.45875	0.2521	1.82	0.128
B		6.43750	3.21875	0.2521	12.77	0.000
C		3.29250	1.64625	0.2521	6.53	0.001
D		2.29000	1.14500	0.2521	4.54	0.006
A*B		0.59000	0.29500	0.2521	1.17	0.295
A*C		0.15500	0.07750	0.2521	0.31	0.771
A*D		0.83750	0.41875	0.2521	1.66	0.158
B*C		1.51000	0.75500	0.2521	2.99	0.030
B*D		1.59250	0.79625	0.2521	3.16	0.025
C*D		0.44750	0.22375	0.2521	0.89	0.415

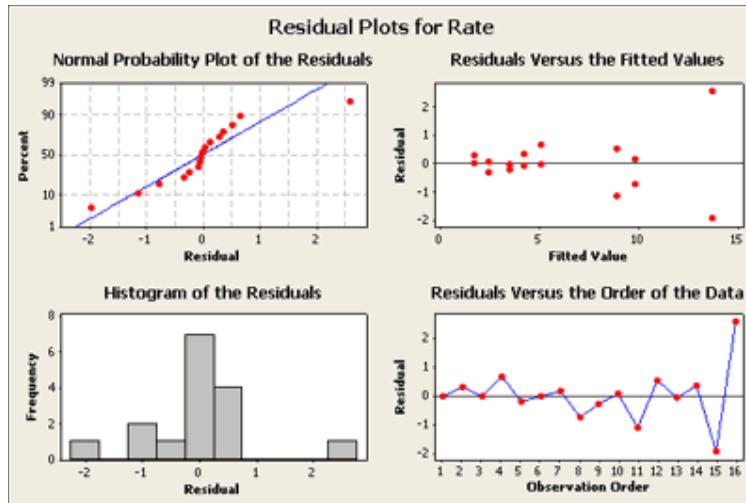
S = 1.00843 R-Sq = 98.07% R-Sq(adj) = 94.20%

Analysis of Variance for Rate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	233.471	233.471	58.368	57.40	0.000
2-Way Interactions	6	24.360	24.360	4.060	3.99	0.075
Residual Error	5	5.085	5.085	1.017		
Total	15	262.916				

The combined main effects are significant as seen in the combined summary table. And the individual terms, B, C, D, BC and BD, are all significant, just as shown on the normal probability plot above.

Now let's go one step farther and look at the completely reduced model. We'll go back into Minitab and get rid of everything except for the significant terms. Here is what you get:





What do you think?

Residuals versus the fitted values plot in the upper right-hand corner has now a very distinct pattern. It seems to be a classic as the response gets larger the residuals get more spread apart.

What does this suggest is needed? For those of you who have studied heteroscedastic variance patterns in regression models you should be thinking about possible transformations.

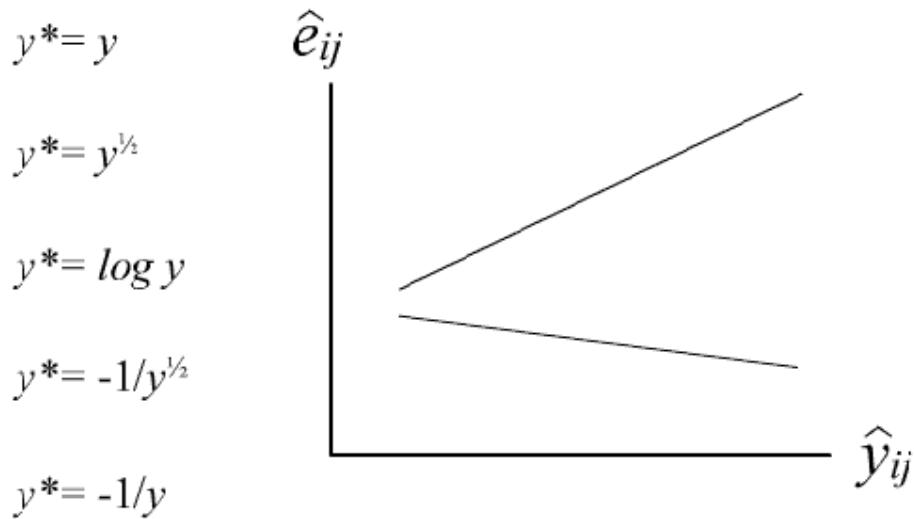
A transformation - the large values are more variable than smaller values. But why does this only show up now? Well, when we fit a full model it only has one observation per cell and there's no pure way to test for residuals. But when we fit a reduced model, now there is inherent replication and this pattern becomes apparent.

Take a look at the data set and you will find the square root and the log already added in order to analyze the same model using this transformed data. What do you find happens?

6.4 - Transformations

When you look at the graph of the residuals as shown below you can see that the variance is small at the low end and the variance is quite large on the right side producing a fanning effect. Consider the family of transformations that can be applied to the response y_{ij} .

Loading [MathJax]/extensions/MathZoom.js



Transformations towards the bottom of the list are stronger in how they shrink large values more than they shrink small values that are represented on the plot. This pattern of the residuals is one clue to get you to be thinking about the type of transformations you would select.

The other consideration and thinking about transformations of the response y_{ij} is what it does to the relationship itself. Some of you will recall from other classes the Tukey one-degree-of-freedom test for interaction. This is a test for interaction where you have one observation per cell such as with a randomized complete block design. But with one observation per cell and two treatments our model would be :

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

where,

$$\begin{aligned} i &= 1 \dots a, \\ j &= 1 \dots b, \text{ with} \\ k &= 1 \dots 1, \text{ (only have one observation per cell)} \end{aligned}$$

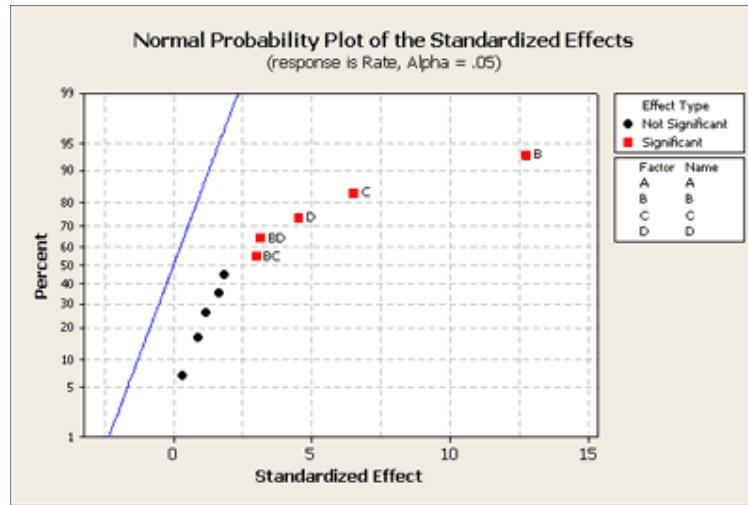
There is no estimate of pure error so we cannot fit the old model. The model proposed by Tukey's has one new parameter (γ) gamma :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma\alpha_i\beta_j + \epsilon_{ij}$$

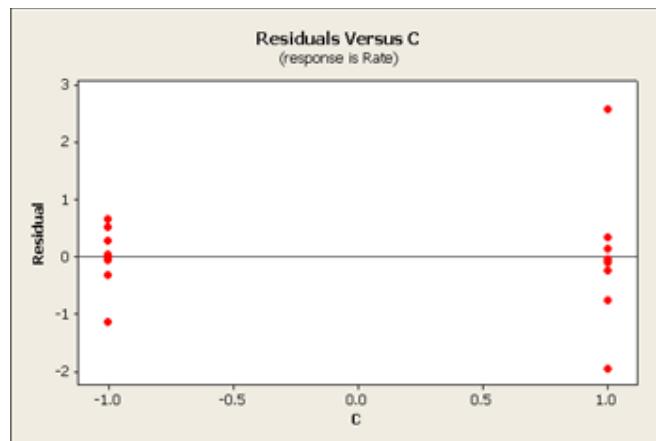
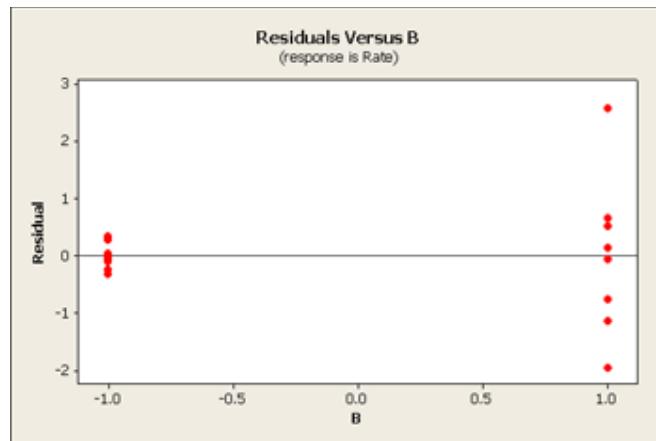
This single parameter, gamma, is the 1 degree of freedom term and so our error, ϵ_{ij} , has $(a-1)(b-1)-1$ degrees of freedom. This model allows for just a single additional parameter which is based on a multiplicative effect on the two factors.

Now, when is this applicable?

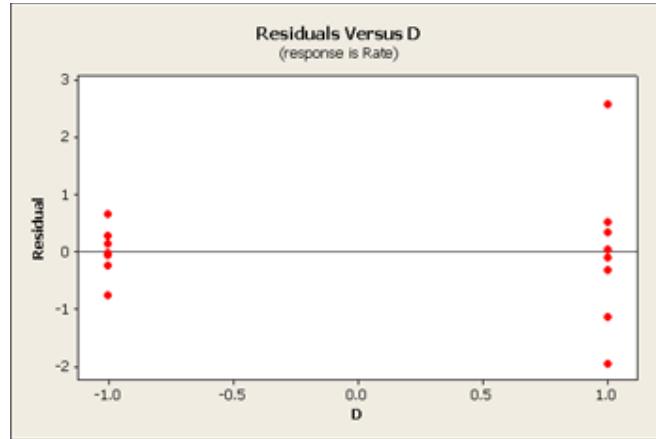
Let's go back to the drill rate example ([Ex6-3.MTW \[5\]](#)) where we saw the fanning effect in the plot of the residuals. In this example B, C and D were the three main effects and there were two interactions BD and BC. From Minitab we can reproduce the normal probability plot for the full model.



But let's first take a look at the residuals versus our main effects B, C and D.

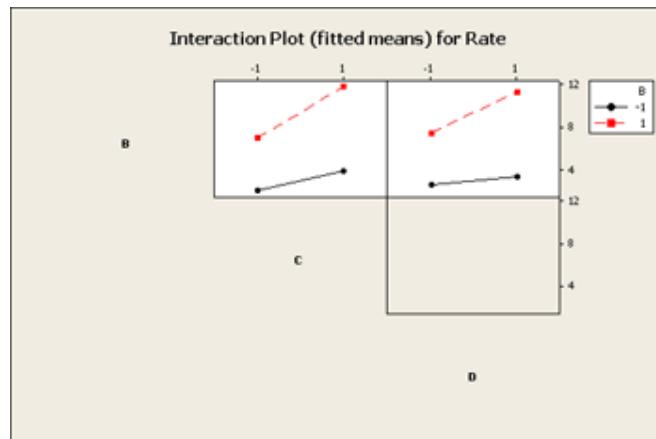
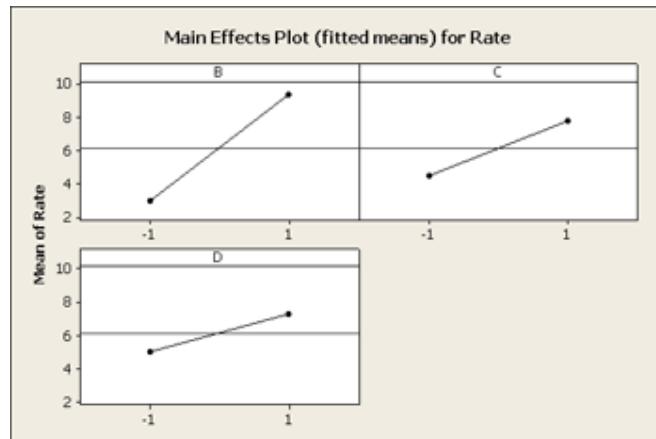


Loading [MathJax]/extensions/MathZoom.js



All three of these residuals versus the main effects show same pattern, the large predicted values tend to have larger variation.

Next, what we really want to look at is the factorial plots for these three factors, B, C and D and the interactions among these, BD and BC.

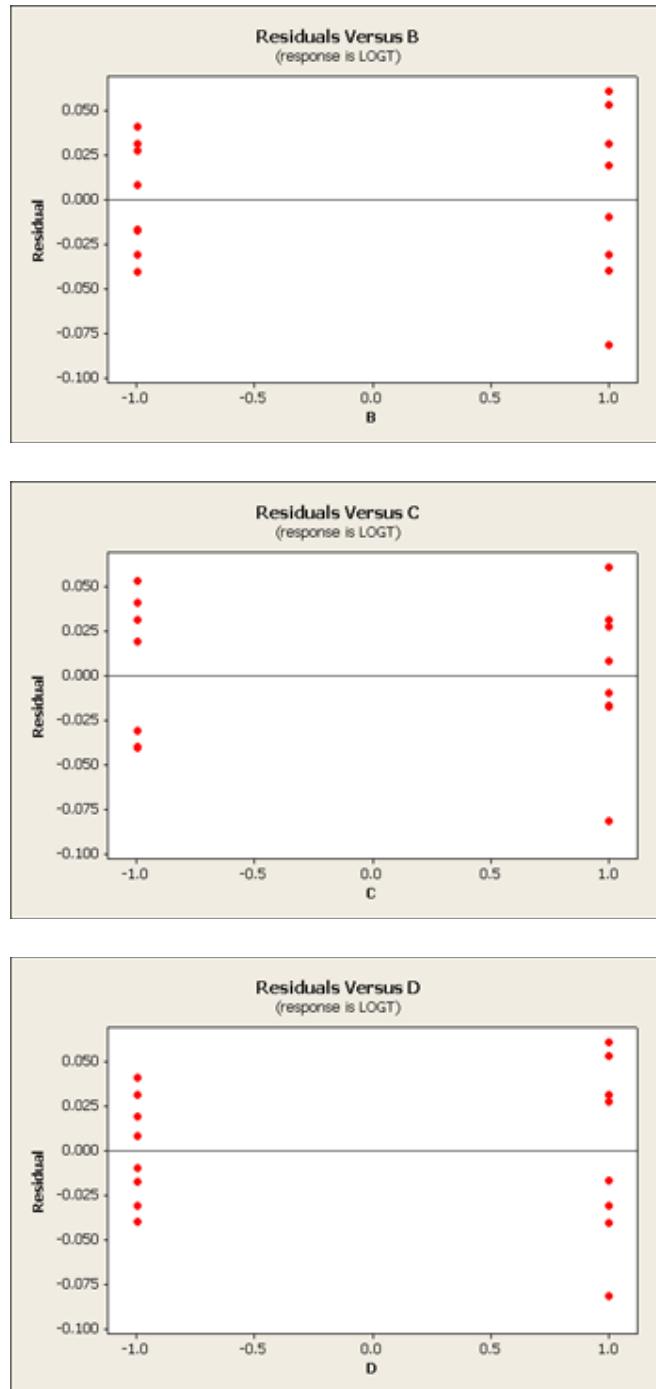


What you see in the interaction plot above is a pattern that is non-parallel showing there is interaction present. But, from what you see in the residual graph what would you expect to see on this factor plot?

Loading [MathJax]/extensions/MathZoom.js

The tell-tale pattern that is useful here is an interaction that does not have crossing lines - a fanning effect - and it is exactly the same pattern that allows the Tukey model to fit. In both cases, it is a pattern of interaction that you can remove by transformation. If we select a transformation that will shrink the large values more than it does the small values and the overall result would be that we would see less of this fan effect in the residuals.

We can look at either the square root or log transformation. It turns out that the log transformation is the one that seems to fit the best. On a log scale it looks somewhat better - it might not be perfect but it is certainly better than what we had before.



Let's also look at the analysis of variance.
Loading MathJax/extensions/MathZoom.js

Factorial Fit: LOGT versus B, C, D

Estimated Effects and Coefficients for LOGT (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		0.69355	0.01218	56.94	0.000
B		0.50204	0.01218	20.61	0.000
C		0.25126	0.01218	10.31	0.000
D		0.14248	0.01218	5.85	0.000
B*C		-0.02114	0.01218	-0.87	0.406
B*D		0.04196	0.01218	1.72	0.116

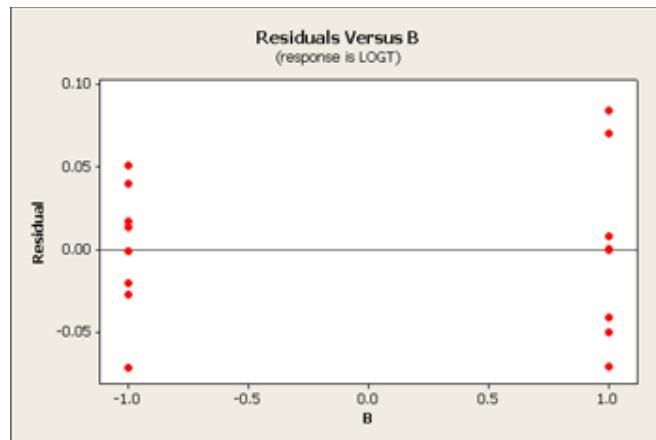
S = 0.0487213 R-Sq = 98.27% R-Sq(adj) = 97.41%

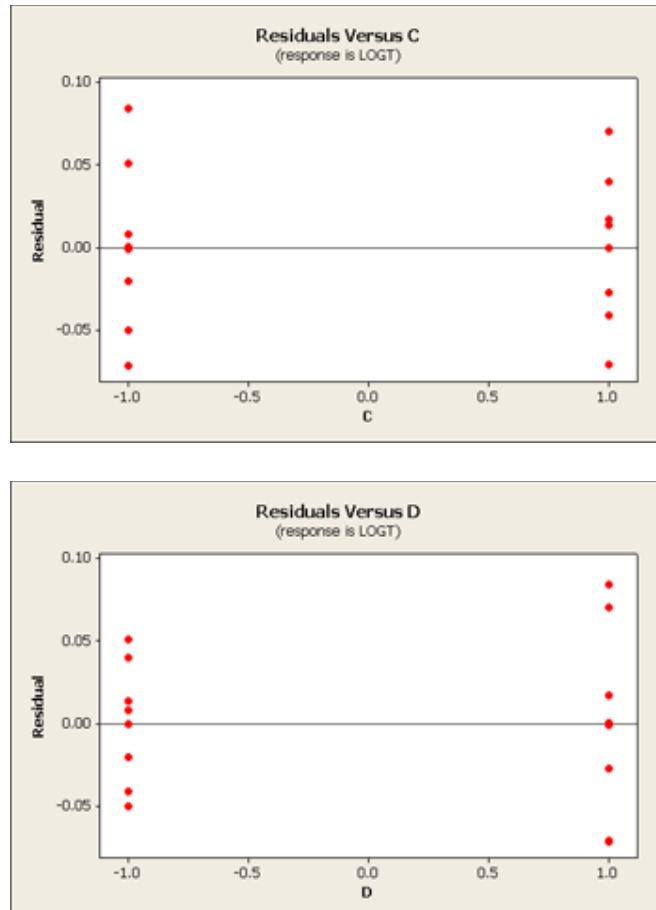
Analysis of Variance for LOGT (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	1.34190	1.34190	0.447300	188.44	0.000
2-Way Interactions	2	0.00883	0.00883	0.004414	1.86	0.206
Residual Error	10	0.02374	0.02374	0.002374		
Lack of Fit	2	0.00112	0.00112	0.000558	0.20	0.825
Pure Error	8	0.02262	0.02262	0.002828		
Total	15	1.37447				

The overall main effects are still significant. But the two 2-way interactions effects combined are no longer significant, and individually, the interactions are not significant here either. So, the log transformation which improved the unequal variances pulled the higher responses down more than the lower values and therefore resulted in more of a parallel shape. What's good for variance is good for a simple model. Now we are in a position where we can drop the interactions and reduce this model to a main effects only model.

Now our residual plots are nearly homoscedastic for B, C and D. See below...





Serendipity - good things come in packages! When you pick the correct transformation, you sometimes achieve constant variance and a simpler model.

Many times you can find a transformation that will work for your data - giving you a simpler analysis but it doesn't always work.

Transformations are typically performed to:

- Stabilize variance - to achieve equal variance
- Improve normality - this is often violated because it is easy to have an outlier when variance is large which can be 'reined in' with a transformation
- Simplify the model

Sometimes transformations will solve a couple of these problems.

Is there always a transformation that can be applied to equalize variance? Not really ... there are two approaches to solving this question. First, we could use some non-parametric method. Although non-parametric methods have fewer assumptions about the distribution, you still have to worry about how you are measuring the center of the distribution. When you have a non-parametric situation you may have a different shaped distribution in different parts of the experiment. You have to be careful about using the mean in one case, and the median in another ... but that is one approach.

The other approach is a weighted analysis, where you weight the observations according to the inverse of their variance. There are situations where you have unequal variation for maybe a

Loading [MathJax]/extensions/MathZoom.js

known reason or unknown reason, but if you have repeated observations and you can get weights, then you can do a weighted analysis.

It is this course author's experience many times you can find a transformation when you have this kind of pattern. Also, sometimes when you have unequal variance you just have a couple of bad outliers, especially when you only have one or a few observations per cell. In this case it is difficult to distinguish whether you have a couple of outliers or the data is heteroscedastic - it is not always clear.

Empirical Selection of Lambda

Prior (theoretical) knowledge or experience can often suggest the form of a transformation. However, another method for the analytical selection of lambda for the exponent used in the transformation is the Box-Cox (1964). This method simultaneously estimates the model parameters and the transformation parameter lambda.

Box-Cox method is implemented in some statistical software applications.

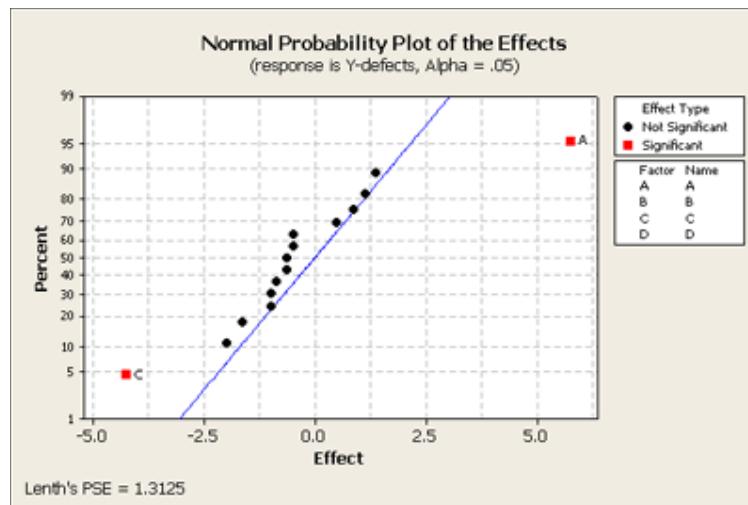
Example 6.4

This example is a four factor design in a manufacturing situation where injection molding is the focus. Injection molding is a very common application in industry; a 2^k design where you have many factors influencing the quality which is measured by how many defects are created by the process. Almost anything that you can think of which have been made out of plastic was created through the injection molding process.

See the example in ([Ex6-4.MTW](#) [6])

In this example we have four factors again: A = temperature of the material, B = clamp time for drying, C = resin flow, and D = closing time of the press. What we are measuring as the response is number of defects. This is recorded as an index of quality in terms of percent. As you look through the data in Figure 6.29 (7th edition) you can see percent of defects as high as 15.5% or as low as 0.5%. Let's analyze the full model in Minitab.

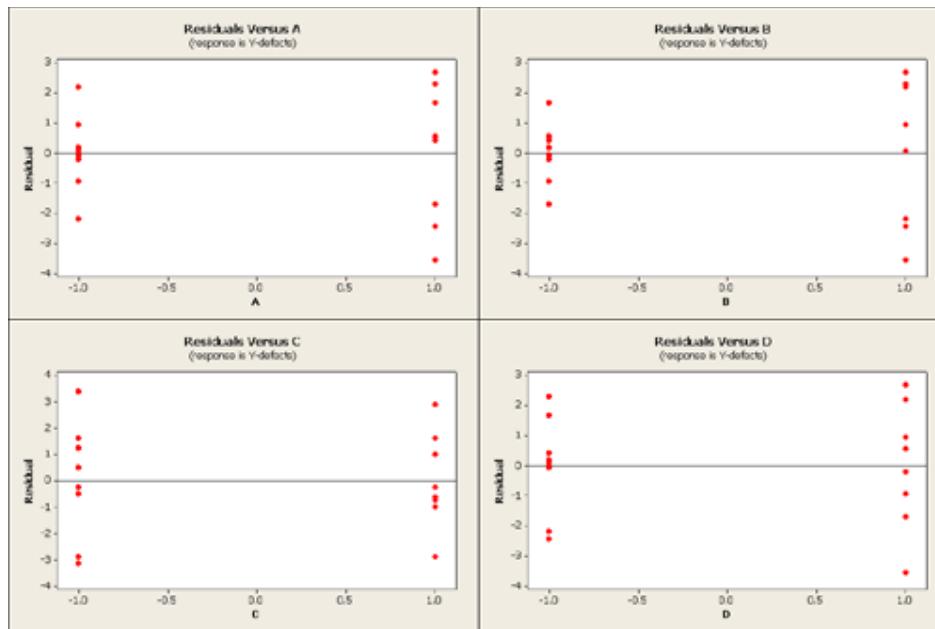
The normal probability plot of the effects shows us that two of the factors A and C are both significant and none of the two-way interactions are significant.



Loading [MathJax]/extensions/MathZoom.js

What we want to do next is look at the residuals vs. variables A, B, C, D in a reduced model with just the main effects as none of the interactions seemed important.

For each factor you see that the residuals are more dispersed (higher variance) to the right than to the left. Overall, however, the residuals do not look too bad and the normal plot also does not look too bad. When we look at the *p*-values we find that A and C are significant but B and D are not.



But there is something else that can be learned here. The point of this example is that although the B factor is not significant as it relates to the response, percentage of product defects - however, if you are looking for a recommended setting for B you should use the low level for B. A and C, are significant and will reduce the number of defects. However, by choosing B at the low level you will produce a more homogeneous product, products with less variability. What is important in product manufacturing is not only reducing the number of defects but also producing products that are uniform. This is a secondary consideration that should be taken into account after the primary considerations related to the percent of product defects.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/34>

Links:

- [1] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_factorial_design_viewlet_swf.html','I06_factorial_design', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_factorial_design_viewlet_swf.html','I06_factorial_design', 718, 668);)
- [2] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/Ex6-2.MTW>
- [3] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_unreplicated_viewlet_swf.html','I06_unreplicated', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_unreplicated_viewlet_swf.html','I06_unreplicated', 718, 668);)
- [4] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_graphical_viewlet_swf.html','I06_graphical', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/L06_graphical_viewlet_swf.html','I06_graphical', 718, 668);)
- [5] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson06/Ex6-3.MTW>
- [6] <https://onlinecourses.science.psu.edu/stat503/node/38/graphics/EX6-4.MTW>

Loading [MathJax]/extensions/MathZoom.js



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 7: Confounding and Blocking in 2^k Factorial Designs

Lesson 7: Confounding and Blocking in 2^k Factorial Designs

Introduction

In Lesson 4 we discussed blocking as a method for removing extraneous sources of variation. In this lesson we consider blocking in the context of 2^k designs. We will then make a connection to confounding, and show a surprising application of confounding where it is beneficial rather than a liability.

Learning Objectives & Outcomes

By the end of this lesson, readers are supposed to understand:

- Concept of Confounding
- Blocking of replicated 2^k factorial designs
- Confounding high order interaction effects of the 2^k factorial design in 2^p blocks
- How to choose the effects to be confounded with blocks
- That a 2^k design with a confounded main effect is actually a Split Plot design
- The concept of Partial Confounding and its importance for retrieving information on every interaction effect

Blocking in Replicated Designs

In 2^k replicated designs where we have n replications per cell and perform a completely randomized design we randomly assign all 2^k times n experimental units to the 2^k treatment combinations. Alternatively, when we have n replicates we can use these n replicates as blocks, and assign the 2^k treatments to the experimental units within each of the n blocks. If we are going to replicate the experiment anyway, at almost no additional cost, you can block the experiment, doing one replicate first, then the second replicate, etc. rather than completely randomize the n times 2^k treatment combinations to all the runs.

There is almost always an advantage to blocking when we replicate the treatments. This is true even if we only block using time due to the order of the replicates. However, there are often many other factors that we have available as potential sources of variation that we can include as a block factor, such as batches of material, technician, day of the week, or time of day, or other environmental factors. Thus if we can afford to replicate the design then it is almost always useful to block.

Typesetting math: 100%

To give a simple example, if we have four factors, the 2^k design has 16 treatment combinations, so say we plan to do just two replicates of the design. Without blocking, the ANOVA has $2^4 = 16$ treatments, but with $n = 2$ replicates, the MSE would have 16 degrees of freedom. If we included a block factor, with two levels, the ANOVA would use one of these 16 degrees of freedom for the block, leaving 15 degrees of freedom for MSE. Hence the statistical cost of blocking is really the loss of one degree of freedom for error, and the potential gain if the block explains significant variation would be to reduce the size of the MSE and thereby increase the power of the tests.

The more interesting case that we will consider next is when we have an unreplicated design. If we are only planning to do one replicate, can we still benefit from the advantage ascribed to blocking our experiment?

7.1 - Blocking in an Unreplicated Design

We begin with a very simple replicated example of blocking. Here we have 2^2 treatments and we have $n = 3$ blocks. In the graphic below the treatments are labeled using the standard Yates notation. Here the 2^2 treatments are the full set of treatment combinations so we can simply put each replicate within a block and assign them in this way.

We can use the Minitab software to construct this design as seen in the video below.

2² Design in 3 blocks of size 4 - Minitab (no sound)



Now let's consider the case when we don't have any replicates, hence when we only have one set of treatment combinations. We go back to the definition of effects that we defined before. We did this using following table, where $\{(1), a, b, ab\}$ is the set of treatment combinations, and A, B, and AB are the effect contrasts:

trt	A	B	AB
(1)	-1	-1	1
a	1	-1	-1
b	-1	1	-1
ab	1	1	1

The question is: what if we want to block this experiment? Or, more to the point, when it is necessary to use blocks, how would we block this experiment?

If our block size is less than four we are only going to consider, in this context of 2^k treatments, block sizes in the same family, i.e. 2^p number of blocks. So in the case of this example let's use blocks of size 2, which is 2^1 . If we have blocks of size two then we must put two treatments in each block. One example would be twin studies where you have two sheep from each ewe. The twins would have homogeneous genetics and the block size would be two for the two animals. Another example might be two-color micro-arrays where you have only two colors in each micro-array.

So now the question: How do we assign our four treatments to our blocks of size two?

In our example each block will be composed of two treatments. The usual rule is to pick an effect you are least interested in, and this is usually the highest order interaction, as a means of specifying how to do blocking. In this case it is the AB effect that we will use to determine our blocks. As you can see in the table below we have used the high level of AB to denote Block 1, and the low-level of AB to denote Block 2. This determines our design.

trt	A	B	AB	Block
(1)	-1	-1	1	1
a	1	-1	-1	2
b	-1	1	-1	2
ab	1	1	1	1

Now, using this design we can assign treatments to blocks. In this case treatment (1) and treatment *ab* will be in the first block, and treatment *a* and treatment *b* will be in the second block.

Blocks of size 2

Block	1	2
AB	+	-

Typesetting math: 100%

$$\begin{array}{cc} (1) & a \\ ab & b \end{array}$$

This design confounds blocks with the AB interaction. You can see this by these contrasts - the comparison between block 1 and Block 2 is the same comparison as the AB contrast. Note that the A effect and the B effect are orthogonal to the AB effect. This design gives you complete information on the A and the B main effects, but it totally confounds the AB interaction effect with the block effect.

Although our block size is fixed at size = 2 we still might want to replicate this experiment in addition. What we have above is two blocks which is one unit of the experiment. We could replicate this design additionally let's say r times and each replicate of the design would be 2 blocks of the design laid out in this way.

We show how to construct this with four replicates. Review the movie below to see how this occurs in Minitab.

2^2 Design in 8 blocks of size 2 - Minitab (no sound)



7.2 - The 2^3 Design

Typesetting math: 100%

Let's look now at the 2^3 design. Here we have 8 treatments and we could create designs with blocks of size 2^p - which could either be blocks of size 4 or 2. As before, we can write this out in a table as:

trt	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
<i>a</i>	+	+	-	-	-	-	+	+
<i>b</i>	+	-	+	-	-	+	-	+
<i>ab</i>	+	+	+	-	+	-	-	-
<i>c</i>	+	-	-	+	+	-	-	+
<i>ac</i>	+	+	-	+	-	+	-	-
<i>bc</i>	+	-	+	+	-	-	+	-
<i>abc</i>	+	+	+	+	+	+	+	+

In the table above we have defined our seven effects: three main effects {A, B, C}, three 2-way interaction effects {AB, AC, BC}, and one 3-way interaction effect {ABC}. We need to define our blocks next by selecting an effect that we are willing to give up by confounding it within the blocks. Let's first look at an example where we let the block size = 4.

Now we need to ask ourselves, what is typically the least interesting effect? The highest order interaction. Do we will use the contrast of the highest order interaction, the three-way, as the effect to guide the layout of our blocks.

trt	I	A	B	C	AB	AC	BC	ABC	Block
(1)	+	-	-	-	+	+	+	-	1
<i>a</i>	+	+	-	-	-	-	+	+	2
<i>b</i>	+	-	+	-	-	+	-	+	2
<i>ab</i>	+	+	+	-	+	-	-	-	1
<i>c</i>	+	-	-	+	+	-	-	+	2
<i>ac</i>	+	+	-	+	-	+	-	-	1
<i>bc</i>	+	-	+	+	-	-	+	-	1
<i>abc</i>	+	+	+	+	+	+	+	+	2

Under the ABC column, the - values will be placed in Block 1, and the + values will be placed in Block 2. Thus we can layout the design by defining the two blocks of four observations like this:

	Block 1	2
ABC	-	+

Typesetting math: 100%

(1)	<i>a</i>
<i>ab</i>	<i>b</i>
<i>ac</i>	<i>c</i>
<i>bc</i>	<i>abc</i>

Let's take a look at how Minitab would run this process ...

2^3 Design in 2 blocks of size 4 - Minitab (no sound)



What if we have 2^3 treatments but we want the block size to be 2?

Now for each replicate we need four blocks with only two treatments per block.

Thought Questions: How should we assign our treatments? How many and which effects must you select to confound with the four blocks?

To define the design for four blocks we need to select two effects to confound, and then we will get four combinations of those two effects.

Note: Is there a contradiction here? If we pick two effects to confound that is only two degrees of freedom. But how many degrees of freedom are there among the four blocks?

Three! So, if we confound two effects then we have automatically also confounded the interaction between those two effects. That is simply a result of the structure used here.

Typesetting math: 100%

What if we first select ABC as one of the effects? Then, it would seem logical to pick one of the 2-way interactions as the other confounding factor. Let's say we use AB. If we do this, remember, we also confound the interaction between these two effects. What is the interaction between ABC and AB. It is C. We can see this by multiplying the elements in the columns for ABC and AB. Try it and you get the same coefficients as you have in the column for C. This is called the generalized interaction. Although it intuitively seemed as though ABC and AB would be a good choice, it is not because it also confounds the main effect C.

Another choice would be to pick two of the 2-way interactions such as AB and AC. The interaction of these is BC. In this case you have not confounded a main effect, but instead have confounded the three two-way interactions. The four combinations of the AB and AC interactions define the four blocks as seen in this color coded table.

trt	I	A	B	C	AB	AC	BC	ABC	Block
(1)	+	-	-	-	+	+	+	-	4
a	+	+	-	-	-	-	+	+	1
b	+	-	+	-	-	+	-	+	3
ab	+	+	+	-	+	-	-	-	2
c	+	-	-	+	+	-	-	+	2
ac	+	+	-	+	-	+	-	-	3
bc	+	-	+	+	-	-	+	-	1
abc	+	+	+	+	+	+	+	+	4

Look under the AB and the AC columns. Where there are - values for both AB and AC these treatments will be placed in Block 1. Where there is a + value for AB and a - value for AC these treatments will be placed in Block 2. Where there is a - value for AB and a + value for AC these treatments will be placed in Block 3. And finally, where there are + values for both AB and AC these treatments will be placed in Block 4. From here we can layout the design separating the four blocks of two observations like this:

Block	1	2	3	4
AB, AC	-, -	+, -	-, +	+, +
a	ab	b	(1)	
bc	c	ac	abc	

Let's take a look at how Minitab would run this process ...

Typesetting math: 100%

2³ Design in 4 blocks of size 2 - Minitab (no sound)



For the 2³ design the only two possibilities are either block sizes of two or four. When we look at more than eight treatments or 2³, then we have more combinations possible. We typically want to confound the highest order of interaction possible remembering that all generalized interactions are also confounded. This is a property of the geometry of designs.

In the next lesson we will look at how we can analyze the data if we take replications of these basic designs, considering one replicate as just the basic building block. This is typically determined by the fact that the block size is usually imposed by some cost or size restrictions on the experiment. However, given adequate resources you can replicate that whole experiment multiple times. So then the question becomes how to analyze these designs and how do we pull out the treatment information.

7.3 - Blocking in Replicated Designs

In the previous section, we saw a 2² treatment design with 4 runs constructed in two blocks confounded with the AB contrast. We also saw a 2³ design constructed in two blocks, with ABC confounded with blocks. We say this is a 2³ design in 2¹ blocks of size 2² per replicate. And we also saw a 2³ design in 2² = 4 blocks of size 2¹ = 2 per replicate with effects AB, AC, and therefore AB × AC = A²BC = BC confounded with blocks.

Typesetting math: 100%

Now let's consider this last situation when we have $n = 3$ replicates of this basic design with $b = 4$ blocks. We can write a model:

$$Y_{ijklm} = \mu + r_i + b_{j(i)} + \alpha_k + \beta_l + \gamma_m + \dots$$

where “ i ” is the index for replicates and “ j ” is the index for blocks within the replicates. “ k ”, “ l ” and “ m ” are indices for the different treatment factors.

AOV	df	
Rep	$n-1$	$= 3 - 1 = 2$
Blk(Rep)	$n(b - 1)$	$= 3(4 - 1) = 9$
A	$2 - 1$	1
B	$2 - 1$	1
C	$2 - 1$	1
ABC	$2 - 1$	1
Error	$(n - 1)*4$	8
Total	$n*2^3 - 1$	23

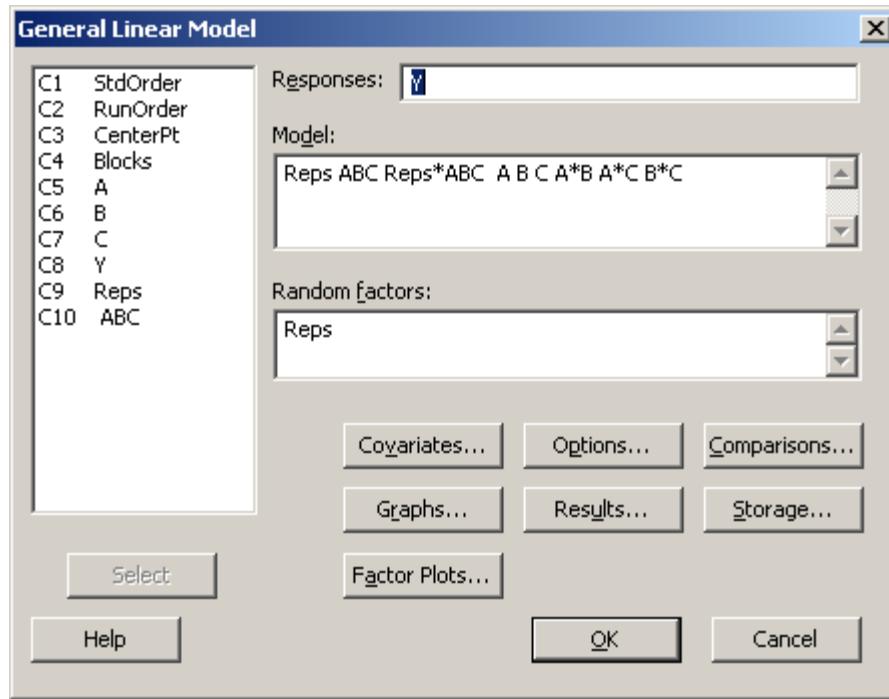
Now we consider another example: in figure 7.3 of the text we see four replicates with ABC confounded in each of the four replicates. The ANOVA for this design is seen in table 7.5 which shows that the Block effect (Block 1 vs. Block 2) is equivalent to the ABC effect and since there are four replicates of this basic design, we can extract some information about the ABC effect, and indeed test the hypothesis of no ABC effect, by using the Rep \times ABC interaction as error.

See the analysis of this design using Minitab:

Stat >> ANOVA >> General Linear Model >>

and fitting the following model:

Typesetting math: 100%



If Reps is specified as a random effects factor in the model, as above, GLM will produce the correct F-tests based on the Expected Means Squares. The reason is analogous to the RCBD with random blocks (Reps) and a fixed treatment (ABC). The topic of random factors is completely covered in chapter 13 of the text book

For Minitab Stat >> ANOVA >> GLM to analyze this data, you need to first construct a pseudo-factor called "ABC" which is constructed by multiplying the levels of A, B, and C using 'Calculator' under the 'Data' menu in Minitab. Click on the 'Inspect' button below which will walk you through this process using **Minitab v.16**.



In addition you can open this Minitab project file [2-k-confound-ABC.MPJ](#) [2] and review the steps leading to the output. The response variable Y is random data simply to illustrate the analysis.

Here is an alternative way to analyze this design using the analysis portion of the fractional factorial software in **Minitab v.16**.

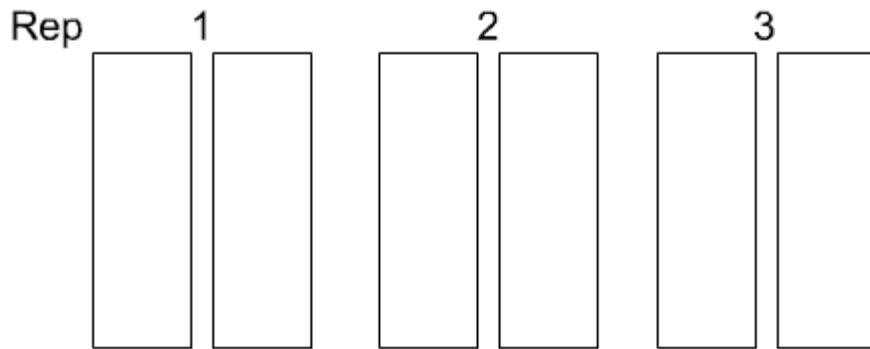


A similar exercise can be done to illustrate the confounded situation where the main effect, say A, is confounded with blocks. Again, since this is a bit nonstandard, we will need to generate a design in Minitab using the default settings and then edit the worksheet to create the confounding we desire and analyze it in GLM.

Typesetting math: 100%

7.4 - Split-Plot Example – Confounding a Main Effect with blocks

Let us consider three replicates, and at each replicate we have two large fields. A diagram of this would look like this:



We will randomly assign the low (-1) or high (+1) level of factor A to each of the two fields. In our example $A = +1$ is irrigation, and $A = -1$ is no irrigation. We will randomly assign the levels of A to the two fields in each replicate. Then the experiment layout would look like this for one replicate:

$A = -1$	$A = 1$
(1)	a
b	ab
c	ac
bc	abc

Similar to the previous example, we would then assigned the treatment combinations of factors B and C to the four experimental units in each block. We could call these experimental units plots -- or using the language of split plot designs -- the blocks are whole plots and the subplots are split plots.

The analysis of variance is similar to what we saw in the example above except we now have A rather than ABC confounded with blocks.

See the Minitab project file [2-K-Split-Plots.MPJ](#) [4] as an example. In addition, here is a viewlet that will walk you through this example using **Minitab v.16**.



Typesetting math: 100%

7.5 - Blocking in 2^k Factorial Designs

Now we will generalize what we have shown by example. We will look at 2^k designs in 2^p blocks of size 2^{k-p} . We do this by choosing k and if we want to confound the design in 2^p blocks then we need to choose p effects to confound. Then, due to the interactions among these effects, we get 2^p-1 effects confounded with blocks.

To illustrate this, if $p = 2$ then we have $2^p =$ four blocks, and thus $2^p-1 = 3$ effects confounded, i.e., the 2 effects we chose plus the interaction between these two. In general, we choose p effects and in addition to the p effects we choose, 2^p-p-1 other effects are automatically confounded. We will call these "generalized interactions" which are also confounded.

Earlier we looked at a couple of examples - for instance when $k = 3$ and $p = 2$. We chose ABC and AB. Then the $ABC \times AB = A^2B^2C = C$ which was also confounded. This shows that the generalized interaction can be a main effect, i.e. the generalized interaction affect can be a lower order term. This is not a good outcome. A better outcome that we settled on was to pick two 2-way interactions, AB and AC, which gave us $AB \times AC = A^2BC$ which = BC, another 2-way interaction. In this case we have all three 2-way interactions confounded, but all the main effects were estimable.

7.6 - Example 1

Let's take another example where $k = 4$, and $p = 2$. This is one step up in the number of treatment factors. And now we have block size = 2^{4-2} or 4. Again, we have to choose two effects to confound. We will show three cases to illustrate this.

- a. Let's try ABCD and a 3-way, ABC. This implies $ABCD \times ABC = A^2B^2C^2D = D$ is also confounded. We usually do not want to confound a main effect. It seems that if you reach too far then you fall short. So, the question is: what is the right compromise?
- b. We could try ABCD and just AB. In this case we get $ABCD \times AB = A^2B^2CD = CD$. Here we have the 4-way interaction and just two of the 2-way interactions confounded. Can we do better than this? Do you know that one or more of your 2-way interaction effects are not important? This is something you probably don't know, but you might. In this case you could pick this interaction and very carefully assign treatments based on this knowledge.
- c. One more try. How about confounding two 3-way interactions? What if we use ABC and BCD. This would give us the interactions of those, $ABC \times BCD = AB^2C^2D$ which = AD.

Which of these three attempts is better? The first try (a) is definitely not good because it confounds the main effect. So, which of the second or third do you prefer? The third (c) is probably the best because it has the fewest lower order interactions confounded. Generally

Typesetting math: 100%

it is assumed that the higher order interactions are less important, so this makes the (c) case the best choice. Both cases (b) and (c) confound 2-way interactions but the (b) case confounds two of them and the (c) case only one.

If we look at Minitab the program defaults are always set to choose the best of these options. Use this short viewlet to see how **Minitab v.17** selects these:

2^4 Design in 4 blocks of size 4 - Minitab (no sound)



7.7 - Example 2

Let's try an example where $k = 5$, and $p = 2$.

- a. If we choose to confound two 4-way interactions ABCD and BCDE, this would give us $ABCD \times BCDE = AB^2C^2D^2E = AE$, confounded as well, which is a 2-way interaction. Not so good.

If we choose ABC and CDE, this would give us $ABC \times CDE = ABC^2DE = ABDE$. So, with this choice we are confounding the higher level 4-way interaction and two 3-way interactions instead of the 2-way interaction as above.

Let's see what Minitab chooses...

Typesetting math: 100%

2^5 Design in 4 blocks of size 8 - Minitab (no sound)



If you were planning to replicate one of these designs, you would not need to use the same three factors for blocking in each replicate of the design, but instead could choose a different set of effects to use for each replicate of the experiment. More on that later.

7.8 - Alternative Method for Assigning Treatments to Blocks

We began this section by looking at the +'s and -'s that were assigned by looking at whether the treatment level was high or low. And in our simplest example we looked at our contrast as +1's and -1's and used these to determine which treatments were assigned to which blocks.

An alternative to using the -'s and +'s is to use 0 and 1. In this case, the low level is 0 and the high level is 1. You can think of this method as just another finite math procedure that can be used to determine which treatments go in which block. We introduce this here because as we will see later, this alternative method generalizes to designs with more than two levels.

Here is a 2^3 design using this notation:

Typesetting math: 100%

L_{ABC}	L_{BC}	L_{AC}	L_{AB}	X_3	X_2	X_1
C	B	A				
0	0	0	0	0	0	0
1	1	0	0	0	1	
0	1	0	1	1	0	
1	0	0	1	1	1	
1	0	1	0	0	0	
0	1	1	0	1	0	
1	1	1	1	1	0	
0	0	1	1	1	1	

Defining Contrasts

$$L_{AB} = X_1 + X_2 \pmod{2}$$

$$L_{AC} = X_1 + X_3 \pmod{2}$$

$$L_{BC} = X_2 + X_3 \pmod{2}$$

$$L_{ABC} = X_1 + X_2 + X_3 \pmod{2}$$

Note: $(\text{mod } 2)$ refers to modular arithmetic where you divide a number by 2 and keep the remainder, e.g., $(5 \pmod{2}) = 1$

If you look at L_{AB} all we are doing here is just summing the 0 and 1 combinations, therefore, L_{AB} = the sum of the row of 0's and 1's for AB (in blue for the first row only). What we are doing is defining the linear combinations using modular 2 arithmetic in this way.

If we want to construct a design for $k = 3, p = 2$ by choosing AB and AC as our defining contrasts then we would construct our design in the following manner:

4	3	2	1	Block
1, 1	0, 1	1, 0	0, 0	L_{AB}, L_{AC}
a	ab	b	(1)	
bc	c	ac	abc	

We are using L_{AB} and L_{AC} to define our blocks, so, what we need to do is exactly what we did before, but this time we are using the 0's and 1's to determine the layout for the design. We are simply using a different coding mechanism here for determining the design layout.

Why is this important?

For two level designs both methods work the same. You can either use the +'s and -'s as the two levels of the factor to divide the treatment combinations into blocks, or you can use zero and one, which is simply a different way to do this and gives us a chance to define the contrasts where

Type setting math: 100%

$$L = a_1X_1 + a_2X_2 + a_3X_3 \pmod{2}$$

where a_i is the exponent of the i th factor in the effect to be confounded (either a 0 or a 1 in each case) and X_i is the level of the i th factor appearing in a particular treatment combination.

Both approaches will give us the same set of treatment combinations in blocks. These functions translate the levels of A and B to the levels of the AB interaction.

When we get to designs with more than two levels using +'s and -'s doesn't work. Therefore, we need another method and using this 1's and 0's approach generalizes. We will come back to this method when we look at 3 level designs - but we will get to that later in Lesson 9.

Partial Confounding

In the above designs, we had to select one or more effects that we were willing to confound with blocks, and therefore not be able to estimate. Generally, we should have some prior knowledge about which effects to neglect or which effects are zero. Even if we do replicate a blocked factorial design, we would not be able to obtain good intra-block estimates the effect(s) which are confounded with blocks. To avoid this issue, there is a method of confounding called *partial confounding* which is widely used.

In partial confounding, the experimenter uses a different interaction effect to be confounded with blocks throughout different replicates. In this way, information regarding each interaction effect which is confounded in one of the replicates can be retrieved from the remaining replicates. Figure 7.7 in the text book shows a partial confounding of 2^3 design where ABC, AB, BC and AC are confounded with blocks in the first through fourth replicates, respectively. Since each interaction is unconfounded in three-quarters of replicates, $\frac{3}{4}$ is the relative information for the confounded effects. The analysis is shown in Table 7.10. Example 7.3 in the text book illustrates a 2^3 design with partial confounding.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/39>

Links:

- [1] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_2k_confnd_ABC_viewlet_swf.html', 'l07_2k_confnd_abc', 704, 652\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_2k_confnd_ABC_viewlet_swf.html', 'l07_2k_confnd_abc', 704, 652);)
- [2] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/2-k-confound-ABC.MPJ>
- [3] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_2k_confnd_default_viewlet_swf.html', 'l07_2k_confnd_default', 704, 652\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_2k_confnd_default_viewlet_swf.html', 'l07_2k_confnd_default', 704, 652);)
- [4] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/2-K-Split-Plota.MPJ>
- [5] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_split_plot_viewlet_swf.html', 'l07_split_plot', 704, 652\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson07/L07_split_plot_viewlet_swf.html', 'l07_split_plot', 704, 652);)

Typesetting math: 100%



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 8: 2-level Fractional Factorial Designs

Lesson 8: 2-level Fractional Factorial Designs

Learning Objectives & Outcomes

The learning objectives for this lesson include:

- Understanding the application of Fractional Factorial designs, one of the most important designs for screening
- Becoming familiar with the terms “design generator”, “alias structure” and “design resolution”
- Knowing how to analyze fractional factorial designs in which there aren’t normally enough degrees of freedom for error
- Becoming familiar with the concept of “foldover” either on all factors or on a single factor and application of each case
- Being introduced to “Plackett-Burman Designs” as another class of major screening designs

Introduction to Fractional Factorial Designs

What we did in the last chapter is consider just one replicate of a full factorial design and run it in blocks. The treatment combinations in each block of a full factorial can be thought of as a fraction of the full factorial.

In setting up the blocks within the experiment we have been picking the effects we know would be confounded and then using these to determine the layout of the blocks.

We begin with a simple example.

In an example where we have $k = 3$ treatments factors with $2^3 = 8$ runs, we select $2^p = 2$ blocks, and use the 3-way interaction ABC to confound with blocks and to generate the following design.

trt	A	B	C	AB	AC	BC	ABC	I
(1)	-	-	-				-	
a	+	-	-				+	
b	-	+	-				+	
ab	+	+	-				-	
c	-	-	+				+	
ac	+	-	+				-	
bc	-	+	+				-	
abc	+	+	+				+	

Here are the two blocks that result using the ABC as the generator:

Loading [MathJax]/extensions/MathMenu.js

Block	1	2
ABC	-	+
(1)		a
ab		b
ac		c
bc		abc

A fractional factorial design is useful when we can't afford even one full replicate of the full factorial design. In a typical situation our total number of runs is $N = 2^{k-p}$, which is a fraction of the total number of treatments.

Using our example above, where $k = 3$, $p = 1$, therefore, $N = 2^2 = 4$

So, in this case, either one of these blocks above is a one half fraction of a 2^3 design. Just as in the block designs where we had AB confounded with blocks - where we were not able to say anything about AB. Now, where ABC is confounded in the fractional factorial we can not say anything about the ABC interaction.

Let's take a look at the first block which is a half fraction of the full design. ABC is the generator of the 1/2 fraction of the 2^3 design. Now, take just the fraction of the full design where ABC = -1 and we place it within its own table:

trt	A	B	C	AB	AC	BC	ABC	I
(1)	-	-	-	+	+	+	-	+
ab	+	+	-	+	-	-	-	+
ac	+	-	+	-	+	-	-	+
bc	-	+	+	-	-	+	-	+

Notice the contrast defining the main effects (similar colors) - there is an aliasing of these effects. Notice that columns with the same color are just -1 times one another.

In this half fraction of the design we have 4 observations, therefore we have 3 degrees of freedom to estimate. The degrees of freedom estimate the following effects: A - BC, B - AC, and C - AB. Thus, this design is only useful if the 2-way interactions are not important, since the effects we can estimate are the combined effect of main effects and 2-way interactions.

This is referred to as a Resolution III Design. It is called a Resolution III Design because the generator ABC has three letters, but the properties of this design and all Resolution III designs are such that the main effects are confounded with 2-way interactions.

Let's take a look at how this is handled in Minitab:

1/2 Fractional Factorial of a 2^3 Design - Minitab (no sound)



This design is only useful if you can be assured that the 2-way interactions are not important. If this is the case then you will find Resolution III designs to be very useful and efficient. When runs are expensive and factors are plentiful these are popular designs.

8.1 - More Fractional Factorial Designs

We started our discussion with a single replicate of a factorial design. Then we squeezed it into blocks, whether it was replicated or not. Now we are going to construct even more sparse designs. There will be a large number of factors, k , but the total number of observations will be $N = 2^{k-p}$, so we keep the total number of observations relatively small as k gets large.

The goal is to create designs that allow us to screen a large number of factors but without having a very large experiment. In the context where we are screening a large number of factors, we are operating under the assumption that only a few are very important. This is called sparsity of effects. We want an efficient way to screen the large number of factors knowing in advance that there will likely be only two or three factors that will be the most important ones. Hopefully we can detect those factors even with a relatively small experiment.

We started this chapter by looking at the 2^{3-1} fractional factorial design. This only has four observations. This is totally unrealistic but served its purpose in illustrating how this design works. ABC was the generator, which is equal to the Identity, ($I = ABC$ or $I = -ABC$). This defines the generator of the design and from this we can determine which effects are confounded or aliased with which other effects.

Let's use the concept of the generator and construct a design for the 2^{4-1} fractional factorial. This gives us a one half fraction of the 2^4 design. Again, we want to pick a high order interaction. Let's

select ABCD as the generator ($I = ABCD$) and by hand we can construct the design. $I = ABCD$ implies that $D = ABC$. First of all, $2^{4-1} = 2^3 = 8$. So, we will have eight observations in our design. Here is a basic 2^3 design in standard Yates notation defined by the levels of A, B, and C:

trt	A	B	C	D=ABC
(1)	-	-	-	-
a	+	-	-	+
b	-	+	-	+
ab	+	+	-	-
c	-	-	+	+
ac	+	-	+	-
bc	-	+	+	-
abc	+	+	+	+

We can then construct the levels of D by using the relationship where $D = ABC$. Therefore, in the first row where all the treatments are minus, $D = -1*-1*-1 = -1$. In the second row, +1, and so forth. As before we write - and + as a shorthand for -1 and +1.

This is a one half fraction of the 2^4 design. A full 2^4 design would have 16 factors.

This 2^{4-1} design is a **Resolution IV** design. The resolution of the design is based on the number of the letters in the generator. If the generator is a four letter word, the design is Resolution IV. The number of letters in the generator determines the confounding or aliasing properties in the resulting design.

We can see this best by looking at the expression $I = ABCD$. We obtain the alias structure by multiplying $A \times I = A \times ABCD = A^2BCD$ which implies $A = BCD$. If we look at the aliasing that occurs we would see that A is aliased with BCD, and similarly all of the main effects are aliased with a three-way interaction:

$$\begin{aligned}B &= ACD \\C &= ABD \\D &= ABC\end{aligned}$$

Main effects are aliased with three-way interactions. Using the same process, we see that two-way interactions are aliased with other two-way interactions:

$$\begin{aligned}AB &= CD \\AC &= BD \\AD &= BC\end{aligned}$$

In total, we have seven effects, the number of degrees of freedom in this design. The only effects that are estimable from this design are the four main effects assuming the 3-way interactions are zero and the three 2-way interactions that are confounded with other 2-way interactions. All 16 effects are accounted for with these seven contrasts plus the overall mean.

Let's take a look at how this type of design is generated in Minitab...

1/2 Fractional Factorial of a 2^4 Design - Minitab (no sound)



Resolution IV Designs

What you need to know about Resolution IV designs:

- the main effects are aliased with the 3-way interactions. This is just the result of the fact that this is a four letter effect that we are using as the generator.
- the 2-way interactions are aliased with each other. Therefore, we can not determine from this type of design which of the 2-way interactions are important because they are confounded or aliased with each other.

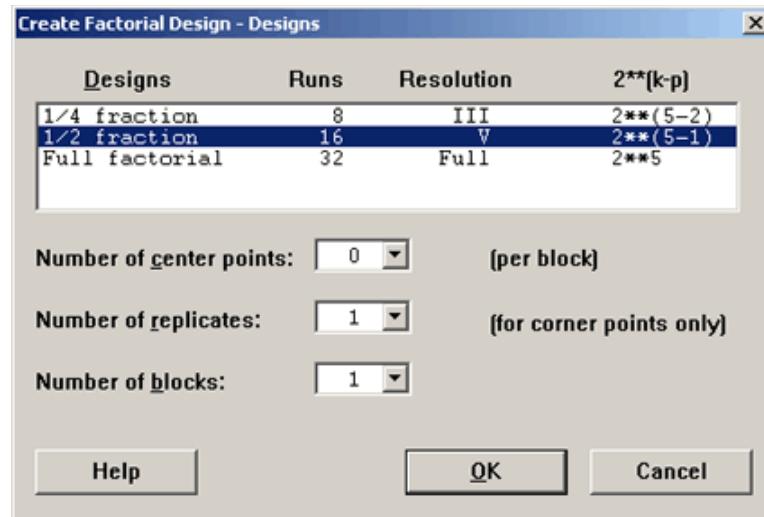
Resolution IV designs are preferred over Resolution III designs. Resolution III designs do not have as good properties because main effects are aliased with two-way interactions. Again, we work from the assumption that the higher order interactions are not as important. We want to keep our main effects clear of other important effects.

The 5 Factor Design

Here we let $k = 5$ and $p = 1$, again, so that we have a one half fraction of a 2^5 design. Now we have five factors, A, B, C, D and E, each at two levels. What would we use as our generator? Since we are only picking one generator, we should choose the highest order interaction as possible. So we will choose $I = ABCDE$, the five-way interaction.

Let's use Minitab to set this up. Minitab gives us a choice of a one half or one fourth fraction. We will select the one half fraction. It says it is a Resolution V design because it has a five letter generator $I = ABCDE$ or $(E = ABCD)$.

Loading [MathJax]/extensions/MathMenu.js



We get a 2^{5-1} , so there are 16 observations. Taking a look at the design:

```
Fractional Factorial Design

Factors: 5 Base Design: 5, 16 Resolution: V
Runs: 16 Replicates: 1 Fraction: 1/2
Blocks: 1 Center pts (total): 0

Design Generators: E = ABCD

Alias Structure

I + ABCDE
A + BCDE
B + ACDE
C + ABDE
D + ABCE
E + ABCD
AB + CDE
AC + BDE
AD + BCE
AE + BCD
BC + ADE
BD + ACE
BE + ACD
CD + ABE
CE + ABD
DE + ABC
```

$E = ABCD$ gives us the basis for the resolution of the design.

Let's look at the properties of a Resolution V design. We can see that:

- the main effects are 'clear' of 2-way and 3-way interactions.
- the main effects are only confounded with 4-way interactions or higher, so this gives us really good information, and
- the 2-way interactions are 'clear' of each other but are aliased with 3-way interactions.

The Resolution V designs are everybody's favorite because you can estimate main effects and two-way interactions if you are willing to assume that three-way interactions and higher are not important.

You can go higher, with Resolution VI, VII etc. designs, however, Resolution III is more or less the minimum, and Resolution IV and V are increasing in good properties in terms of being able to estimate the effects.

A One-Fourth Fractional Design, or a $1/2^2$ Fraction of a 2^k design

Let's try to construct a $1/4$ fractional design using the previous example where $k = 4$ factors. In this case $p = 2$, therefore we will have to pick 2 generators in order to construct this type of design.

As in the previous example $k = 4$, but now $p = 2$, so this would give us $2^{4-2} = 4$ observations. A problem that we can foresee here is that we only have a total of 3 degrees of freedom to estimate. But we have four main effects, so a main effect is going to have to be confounded or aliased with another main effect. Hence, this design is not even a Resolution III. Let's go ahead anyway.

Let's pick ABCD, as we did before, as one generator and ABC as the other. So we would have $ABCD \times ABC = D$ as our third generator.

This is not good ... now we have a main effect as a generator which means the main effect would be confounded with the mean we can do better than that.

Let's pick ABCD and then AB as a second generator, this would give us $ABCD \times AB = CD$ as our third generator. We pick two but we must also include a generalized interaction.

Now the smallest word in our generator set is a two letter word - so this means that this is a Resolution II design. But we found out that a Resolution II designs tell us that the main effects are aliased with each other, ... hence not a good design if we want to learn which main effects are important.

Let's try another example...

Let's say we have $k = 5$ and $p = 2$. We have five factors, so again we need to pick two generators. We want to pick the generators so that the generators and their interactions are each as large a word as possible. This is very similar to what we were doing when we were confounding in blocks.

Let's pick the 4-way interaction ABCD, and CDE. Then the generalized interaction is $ABCD \times CDE = ABE$. In this case, in the way we picked them the smallest number of letters is 3 so this is a Resolution III design.

We can construct this design in the same way we had previously. We begin with $2^{5-2} = 2^3 = 8$ observations which are constructed from all combinations of A, B, and C, then we'll use our generators to define D and E. Note that $I = ABCD$ tells us that $D = ABC$, and the other generator $I = CDE$ tells us that $E = CD$. Now we can define the new columns $D = ABC$ and $E = CD$. Although D and E weren't a part of the original design, we were able to construct them from the two generators as shown below:

trt	A	B	C	D=ABC	E=CD
(1)	-	-	-	-	+
a	+	-	-	+	-
b	-	+	-	+	-
ab	+	+	-	-	+
				-	+

Loading [MathJax]/extensions/MathMenu.js

ac	+	-	+	-	-
bc	-	+	+	-	-
abc	+	+	+	+	+

Now we have a design with eight observations, 2^3 , with five factors. Our generator set is: I = ABCD = CDE = ABE. This is a Resolution III design because the smallest word in the generator set has only three letters. Let's look at this in Minitab ...

1/4 Fractional Factorial of a 2^5 Design - Minitab (no sound)



A Resolution IV Design Example

Let's take $k = 6$ and $p = 2$, now we again have to choose two generators with the highest order possible, such that the generalized interaction is also as high as possible. We have factors A, B, C, D, E and F to choose from. What should we choose as generators?

Let's try ABCD and CDEF. The generalized interaction of these two = ABEF. We have strategically chosen two four letter generators whose generalized interaction is also four letters. This is the best that we can do. This results in a 2^{6-2} design, which is sometimes written like this, 2^{6-2}_{IV} , because it is a Resolution IV design.

In Minitab we can see the available designs for six factors in the table below:

Available Factorial Designs (with Resolution)															
	Factors														
Runs	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
4	Full	II													
8	Full	IV	II	II	III										
16		Full	V	IV	IV	IV	IV	III	III	II	II	II	III	II	
32			Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
64				Full	VII	V	IV	IV	IV	IV	IV	IV	IV	IV	IV
128					Full	VIII	VI	V	V	IV	IV	IV	IV	IV	IV

Available Resolution III Plackett-Burman Designs

Factors	Runs	Factors	Runs	Factors	Runs
2-7	12,20,24,28,...,48	20-23	24,28,32,36,...,48	36-39	40,44,48
8-11	12,20,24,28,...,48	24-27	28,32,36,40,44,48	40-43	44,48
12-15	20,24,28,36,...,48	28-31	32,36,40,44,48	44-47	48
16-19	20,24,28,32,...,48	32-35	36,40,44,48		

Help **OK**

... with six factors, a $2^{6-2} = 2^4$ design, which has 16 observations, is located in the six factor column, the 16 observation row. This tells us that this design is a Resolution IV, (in yellow). We know from this table that this type of design exists, so in Minitab we can specify this design.

Create Factorial Design - Designs

Designs	Runs	Resolution	2^{k-p}
1/8 fraction	8	III	$2^{*(6-3)}$
1/4 fraction	16	IV	$2^{*(6-2)}$
1/2 fraction	32	VI	$2^{*(6-1)}$
Full factorial	64	Full	2^{**6}

Number of center points: (per block)

Number of replicates: (for corner points only)

Number of blocks:

Help **OK** **Cancel**

... which results in the following output.

Fractional Factorial Design					
Factors:	6	Base Design:	6, 16	Resolution:	IV
Runs:	16	Replicates:	1	Fraction:	1/4
Blocks:	1	Center pts (total):	0		
Design Generators: E = ABC, F = BCD					
Alias Structure					
I + ABCE + ADEF + BCDF					
A + BCE + DEF + ABCDF					
B + ACE + CDF + ABDEF					
C + ABE + BDF + ACDEF					
D + AEF + BCF + ABCDE					
E + ABC + ADF + BCDEF					
F + ADE + BCD + ABCEF					
AB + CE + ACDF + BDEF					
AC + BE + ABDF + CDEF					
AD + EF + ABCF + BCDE					
AE + BC + DF + ABCDEF					
AF + DE + ABCD + BCEF					
BD + CF + ABEF + ACDE					
BF + CD + ABDE + ACEF					
ABD + ACF + BEF + CDE					
ABF + ACD + BDE + CEF					

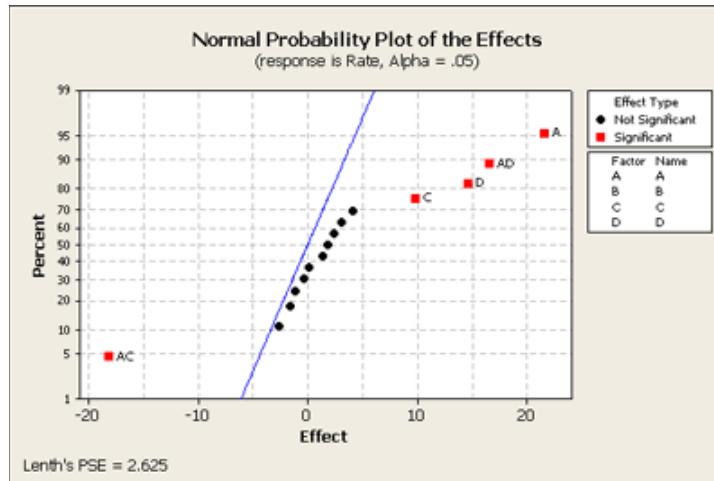
In Minitab by default ABCE and BCDF were chosen as the design generators. The design was constructed by starting with the full factorial of factors A, B, C, and D. Minitab then generated E by using the first three columns, A, B and C. Then it could choose F = BCD.

Because the generator set, I = ABCE = ADEF = BCDF, contains only four letter words, this is classified as a Resolution IV design. All the main effects are confounded with 3-way interactions and a 5-way interaction. The 2-way interactions are aliased with each other. Again, this describes the property of the Resolution IV design.

8.2 - Analyzing a Fractional Factorial Design

We discussed designing experiments, but now let's discuss how we would analyze these experiments. We take an example we saw before. The response Y is filtration rate in a chemical pilot plant and the four factors are: A = temperature, B = pressure, C = concentration and D = stirring rate. (Example 2 from Chapter 6, [Ex6-2.MTW](#) [1])

This experimental design has 16 observations, a 2^4 with one complete replicate. This is the example we looked at with one observation per cell when we introduced a normal scores plot.

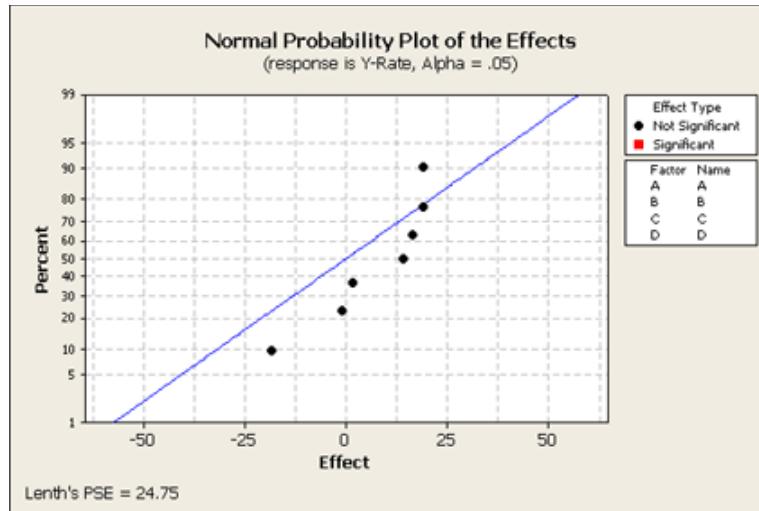


Our final model ended up with three factors, A, C and D, and two of their interactions, AC and AD. This was based on one complete replicate of this design. What might we have learned if we had done an experiment half this size, $N = 8$? If we look at the fractional factorial - one half of this design - where we have D = ABC or I = ABCD as the generator - this creates a design with 8 observations.

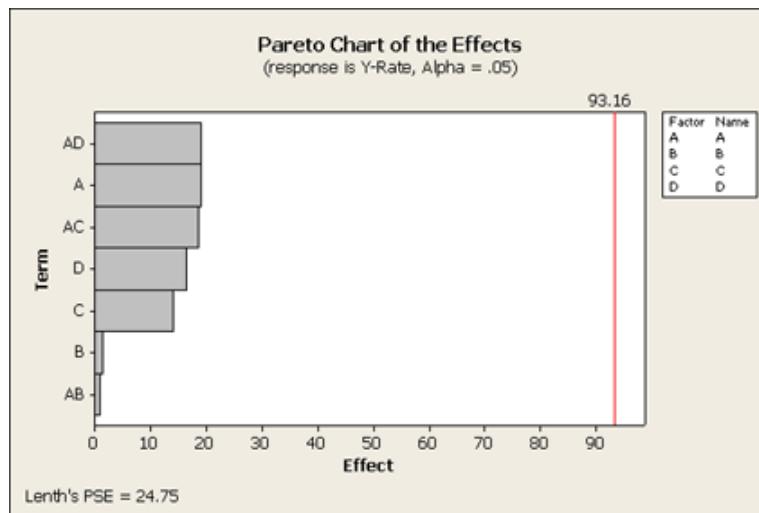
Fractional Factorial Design					
Factors:	4	Base Design:	4, 8	Resolution:	IV
Runs:	8	Replicates:	1	Fraction:	1/2
Blocks:	1	Center pts (total):	0		
Design Generators: D = ABC					
Alias Structure					
I + ABCD					
A + BCD					
B + ACD					
C + ABD					
D + ABC					
AB + CD					
AC + BD					
AD + BC					

The alias structure is a four letter word, therefore this is a Resolution IV design, A, B, C and D are each aliased with a 3-way interaction, (so we can't estimate them any longer), and the two way interactions are aliased with each other.

If we look at the analysis of this 1/2 fractional factorial design and we put all of the terms in the model, (of course some of these are aliased with each other), and we will look at the normal scores plot. What do we get? (The data are in Ex6_2Half.MTW)



We only get seven effects plotted, since there were eight observations. The overall mean does not show up here. These points are labeled but because there are only seven of them there is no estimate of error. Let's look at another plot that we haven't used that much yet - the Pareto plot. This type of plot looks at the effects and orders them from largest to smallest showing you the relative sizes of the effects. Although we do not know what is significant and what is not significant, this still might be a helpful plot to look at to better understand the data.



This Pareto plot shows us that the three main effects A, C, and D that were most significant in the full design are still important as well as the two interactions, AD and AC. However, B and AB are clearly not as large. (You can do this using the Stat >> DOE >> Factorial >> Analyze and click on graph.)

What can we learn from this? Let's try to fit a reduced model from the information that we gleaned from this first step. We will include all the main effects and the AC and AD interactions.

In the analysis, we have four main effects ...

Factorial Fit: Y-Rate versus A, B, C, D						
Estimated Effects and Coefficients for Y-Rate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		70.750	0.5000	141.50	0.004	
A	19.000	9.500	0.5000	19.00	0.033	
B	1.500	0.750	0.5000	1.50	0.374	
C	14.000	7.000	0.5000	14.00	0.045	
D	16.500	8.250	0.5000	16.50	0.039	
A*C	-18.500	-9.250	0.5000	-18.50	0.034	
A*D	19.000	9.500	0.5000	19.00	0.033	
 S = 1.41421 R-Sq = 99.93% R-Sq(adj) = 99.54%						
Analysis of Variance for Y-Rate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	1663.00	1663.00	415.750	207.88	0.052
2-Way Interactions	2	1406.50	1406.50	703.250	351.63	0.038
Residual Error	1	2.00	2.00	2.000		
Total	7	3071.50				

... overall they are almost significant, (.052), and the overall two-way interactions, (.038) but we only have one degree of freedom of error - so this makes this a very low-power test. However, this is the price that you would pay with a fractional factorial. If we look above at the individual effects, B as we saw on the plot appears to be not important, we have further evidence that we should drop this from the analysis.

Back to Minitab and let's drop the B term because it doesn't show up as a significant main effect nor as part of any of the interactions.

Factorial Fit: Y-Rate versus A, C, D						
Estimated Effects and Coefficients for Y-Rate (coded units)						
Term	Effect	Coef	SE Coef	T	P	
Constant		70.750	0.6374	111.00	0.000	
A	19.000	9.500	0.6374	14.90	0.004	
C	14.000	7.000	0.6374	10.98	0.008	
D	16.500	8.250	0.6374	12.94	0.006	
A*C	-18.500	-9.250	0.6374	-14.51	0.005	
A*D	19.000	9.500	0.6374	14.90	0.004	
 S = 1.80278 R-Sq = 99.79% R-Sq(adj) = 99.26%						
Analysis of Variance for Y-Rate (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	1658.50	1658.50	552.833	170.10	0.006
2-Way Interactions	2	1406.50	1406.50	703.250	216.38	0.005
Residual Error	2	6.50	6.50	3.250		
Total	7	3071.50				

Now the overall main effects and 2-way interactions are significant. Residual error still only has 2 degrees of freedom, but this gives us an estimate at least and we can also look at the individual effects.

Loading [MathJax]/extensions/MathMenu.js

So, fractional factorials are useful when you hope or expect that not all of the factors are going to be significant. You are screening for factors to drop out of the study. In this example, we started with a 2^{4-1} design but when we dropped B we ended up with a 2^3 design with 1 observation per cell.

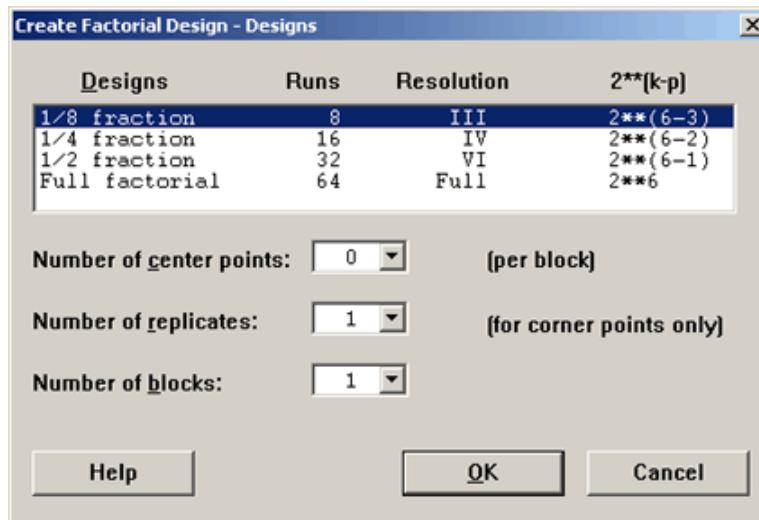
This is a typical scenario, you begin by screening a large number of factors and end up with a smaller set. We still don't know much about the factors and this is still a pretty thin or weak design but it gives you the information that you need to take the next step. You can now do a more complete experiment on fewer factors.

8.3 - Foldover Designs

Foldover designs are useful when you are involved in sequentially testing a set of factors. You begin with very small experiments and proceed in stages. We consider this type of design through two examples.

1/8th fractional factorial of a 2^6 design

First, we will look at an example with 6 factors and we select a 2^{6-3} design, or a 1/8th fractional factorial of a 2^6 design.



In order to select a 1/8 fraction of the full factorial, we will need to choose 3 generators and make sure that the generalized interactions among these three generators are of sufficient size to achieve the higher resolution. In this case it will be a Resolution III as Minitab shows us above.

Let's remind ourselves how we do this. We can choose I = ABD = ACE = BCF as the generators.

Since $N = 2^{6-3} = 2^3$ observations, we start with a basic 2^3 design which would be set up using the following framework. First write down the complete factorial for factors A, B, and C. From that we can generate additional factors based on the available interactions, i.e. we will make D = AB, E = AC, and F = BC. Complete the table below ...

trt	A	B	C	D=AB	E=AC	F=BC	ABC
(1)	-	-	-				
a	+	-	-				

Loading [MathJax]/extensions/MathMenu.js

b	-	+	-
ab	+	+	-
c	-	-	+
ac	+	-	+
bc	-	+	+
abc	+	+	+

Our generators are ABD, ACE and BCF. So, our alias structure is created by this equivalence:

$$I = ABD = ACE = BCF$$

If these are our generators then all of the generalized interactions among these terms are also part of the generator set. Let's take a look at them:

The 2-way interactions:

- $ABD \times ACE = BCDE$,
- $ABD \times BCF = ACDF$,
- $ACE \times BCF = ABEF$,

and the 3-way interaction:

- $ABD \times ACE \times BCF = DEF$

We still have a Resolution III design because the generator set is composed of words, the smallest of which has 3 letters. So you could fill in the framework above for these factors just by multiplying from the basic design, the pluses and minuses.

Minitab does this for you. And the worksheet will look like this:

StdOrder	RunOrder	CenterPt	Blocks	A	B	C	D	E	F
1	1	1	1	-1	-1	-1	1	1	1
2	2	1	1	1	-1	-1	-1	-1	1
3	3	1	1	-1	1	-1	-1	1	-1
4	4	1	1	1	1	-1	1	-1	-1
5	5	1	1	-1	-1	1	1	-1	-1
6	6	1	1	1	-1	1	-1	1	-1
7	7	1	1	-1	1	1	-1	-1	1
8	8	1	1	1	1	1	1	1	1

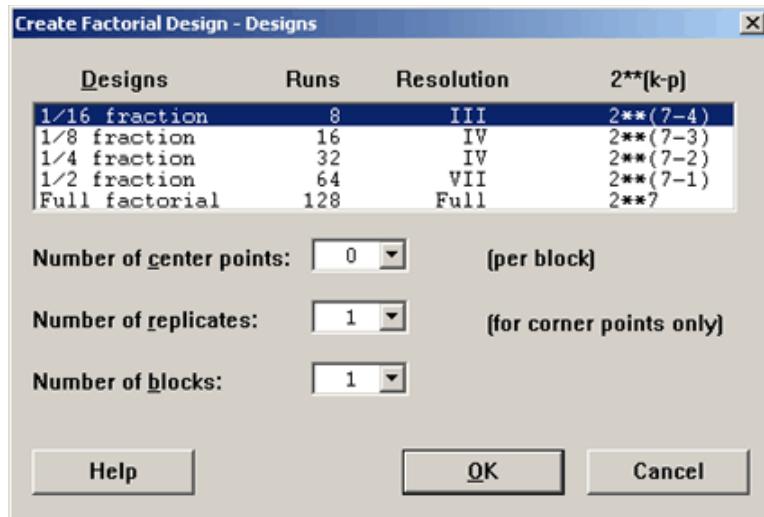
We can estimate all of the main effects and one of the aliased two-way interactions. What this also suggests is that there is one more factor that we could include in a design of this size, $N = 8$.

1/16th fraction of a 2^7 design

Now we consider a 1/16th fraction of a 2^7 design, or a 2^{7-4} design. Again, we will have only $N = 2^3 = 8$ observations but now we have seven factors. Thus $k = 7$ and $p = 4$.

Let's look at this in Minitab - for seven factors here are the design options ...

Loading [MathJax]/extensions/MathMenu.js



The 1/16 fraction design is a Resolution III design and it is the smallest possible one. Here is what the design looks like:

```
Fractional Factorial Design

Factors: 7   Base Design:    7, 8   Resolution: III
Runs:     8   Replicates:   1   Fraction: 1/16
Blocks:   1   Center pts (total): 0

* NOTE * Some main effects are confounded with two-way interactions.

Design Generators: D = AB, E = AC, F = BC, G = ABC

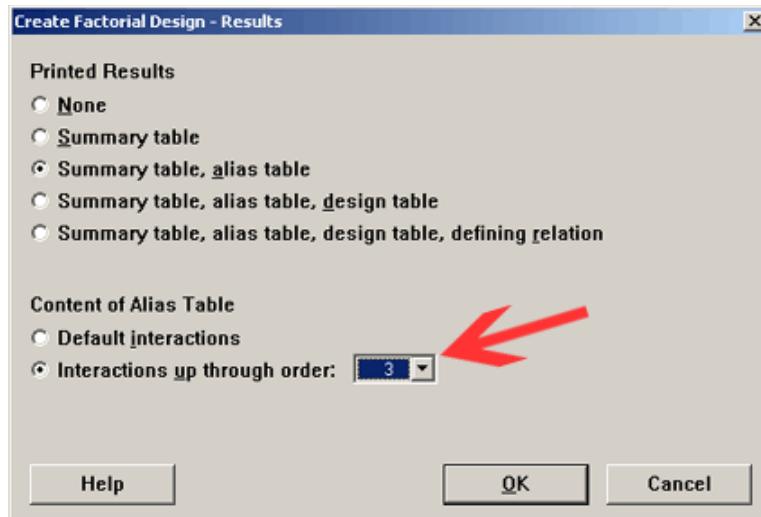
Alias Structure

I + ABD + ACE + AFG + BCF + BEG + CDG + DEF + ABCG + ABEF + ACDF + ADEG + BCDE
+ BDFG + CEGF + ABCDEFG

A + BD + CE + FG + BCG + BEF + CDF + DEG + ABCF + ABEG + ACDG + ADEF + ABCDE +
ABDFG + ACEFG + BCDEFG
B + AD + CF + EG + ACG + AEF + CDE + DFG + ABCE + ABFG + BCDG + BDEF + ABCDF +
ABDEG + BCEFG + ACDEFG
C + AE + BF + DG + ABC + ADF + BDE + EFG + ABCD + ACFG + BCEG + CDEF + ABCEF +
ACDEG + BCDFG + ABDEFG
D + AB + CG + EF + ACF + AEG + BCE + BFG + ACDE + ADFG + BCDF + BDEG + ABCDG +
ABDEF + CDEFG + ABCEFG
E + AC + BG + DF + ABF + ADG + BCD + CFG + ABDE + AEFG + BCEF + CDEG + ABCEG +
ACDEF + BDEFG + ABCDFG
F + AG + BC + DE + ABE + ACD + BDG + CEG + ABDF + ACEF + BEFG + CDFG + ABCFG +
ADEFG + BCDEF + ABCDEG
G + AF + BE + CD + ABC + ADE + BDF + CEF + ABDG + ACEG + BCFG + DEFG + ABEFG +
ACDFG + BCDEG + ABCDEF
```

The generators are listed on top. The same first three are as before and then G = ABC, the only one left. The alias structure gets quite convoluted. The reason being that if we were taking a complete replicate of this design, 2^7 , we could put it into 16 blocks. In this case, we are only looking at one of the 16 blocks in the complete design. In these 16 blocks there are 15 degrees of freedom among these blocks. So, you see I + the 15 effects.

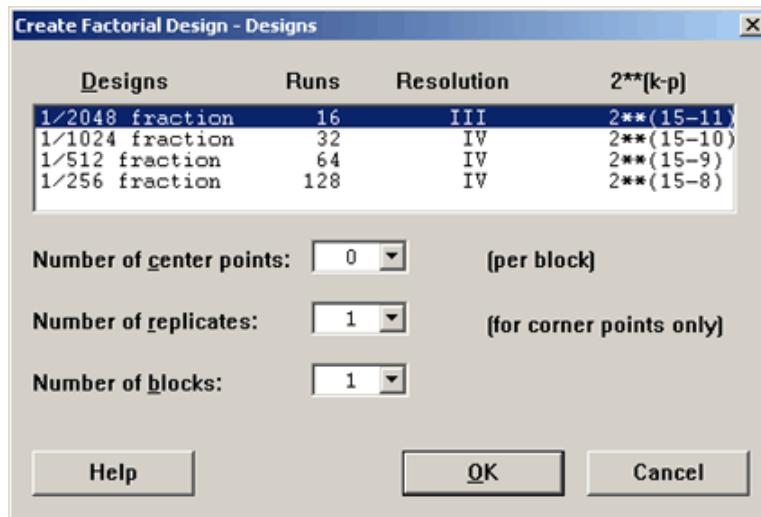
Sometimes people are not interested in seeing all of these higher order interactions, after all five way interactions are not all that interesting. You can clean up this output a bit by using this option found in the 'Results...' dialog box in Minitab:



Notice now that the only thing you find in the table are main effects. No 2-way interactions are available. This is a unique design called a **Saturated Design**. This is the smallest possible design that you could use for 7 factors. Another way to look at this is that for a design with eight observations the maximum number of factors you can include in that design is seven. We are using every degree of freedom to estimate the main effects.

If we moved to the next smallest design where $N = 16$, then what would the saturated design be? 15 factors. You would have a 2^{15-11} , which would give us a 2^4 basic design. Then we could estimate up to 15 main effects.

So, you can see with fairly small designs, only 16 observations, we can test for a lot of factors if we are only interested in main effects, using a Resolution III design. Let's see what the options are in Minitab.



Notice that the largest design shown has 128 runs which is already a very large experiment for 15 factors. You probably wouldn't want more than that.

Folding a Design

Loading [MathJax]/extensions/MathMenu.js

We will come back to Saturated designs - but first let's consider the 2^{7-4} design, which is saturated and is a Resolution III design and let's fold it over.

Let's assume we ran this design, we found some interesting effects but we have no degrees of freedom for error. So we want to look at another replicate of this design. Rather than repeating this exact **same** design we can fold it over.

We can fold it over on all factors, or specify a single factor for folding.

What folding means is to take the design and reverse the sign on all the factors. This would be a fold on all factors.

Now instead of eight observations we have 16. And if you look at the first eight and compare these with the second set of eight you will see that the signs have simply been reversed

StdOrder	RunOrder	CenterPt	Blocks	A	B	C	D	E	F	G
1	1	1	1	1	-1	-1	-1	1	1	-1
2	2	2	1	1	1	-1	-1	-1	-1	1
3	3	3	1	1	-1	1	-1	-1	1	-1
4	4	4	1	1	1	1	-1	1	-1	-1
5	5	5	1	1	-1	-1	1	1	-1	1
6	6	6	1	1	1	-1	1	-1	1	-1
7	7	7	1	1	-1	1	1	-1	-1	1
8	8	8	1	1	1	1	1	1	1	1
9	9	9	1	1	1	1	1	-1	-1	1
10	10	10	1	1	-1	1	1	1	1	-1
11	11	11	1	1	1	-1	1	1	-1	1
12	12	12	1	1	-1	-1	1	-1	1	1
13	13	13	1	1	1	1	-1	-1	1	-1
14	14	14	1	1	-1	1	-1	1	-1	1
15	15	15	1	1	1	-1	-1	1	1	-1
16	16	16	1	1	-1	-1	-1	-1	-1	-1

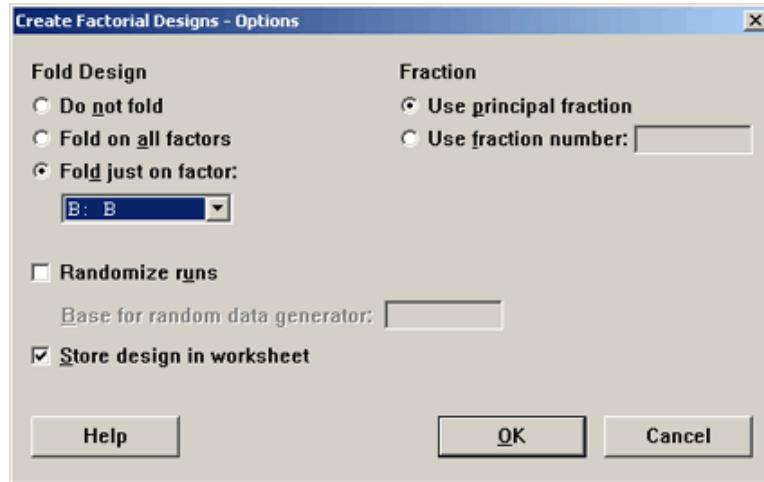
Look at row 1 and row 9 and you will see that they have the exact opposite signs. Thus you double the basic design with all factors exchanged. Or, you can think of this somewhat as taking one replicate and putting it in blocks, we've now taken two of the blocks to create our design.

These designs are used to learn how to proceed from a basic design, where you might have learned something about one of the factors that looks promising, and you want to give more attention to that factor. This would suggest folding, not on all factors, but folding on that one particular factor. Let's examine why you might want to do this.

In our first example above we started with a Resolution III design, and by folding it over on all factors, we have increased the resolution by one number, in this case it goes from resolution III to IV. So, instead of the main effects being confounded with two-way interactions, which they were before, now they are all clear of the two-way interactions. We still have the two-way interactions confounded with each other however.

Now, let's look at the situation where after the first run we were mostly intrigued by factor B.

Now, rather than fold on all factors we want to fold on just factor B.



Notice now that in the column for B, the folded part is exactly the opposite. None of the other columns change, just the column for factor B. All of the other columns stayed the same.

StdOrder	RunOrder	CenterPt	Blocks	A	B	C	D	E	F	G
1	1	1	1	-1	-1	-1	1	1	1	-1
2	2	1	1	1	-1	-1	-1	-1	1	1
3	3	1	1	-1	1	-1	-1	1	-1	1
4	4	1	1	1	1	-1	1	-1	-1	-1
5	5	1	1	-1	-1	1	1	-1	-1	1
6	6	1	1	1	-1	1	-1	1	-1	-1
7	7	1	1	-1	1	1	-1	-1	1	-1
8	8	1	1	1	1	1	1	1	1	1
9	9	1	1	-1	1	-1	1	1	1	-1
10	10	1	1	1	1	-1	-1	-1	1	1
11	11	1	1	-1	-1	-1	-1	1	-1	1
12	12	1	1	1	-1	-1	1	-1	-1	-1
13	13	1	1	-1	1	1	1	-1	-1	1
14	14	1	1	1	1	1	-1	1	-1	-1
15	15	1	1	-1	-1	1	-1	-1	1	-1
16	16	1	1	1	-1	1	1	1	1	1

Now look at the alias structure for this design...

Results for: Worksheet 8**Fractional Factorial Design**

```
Factors: 7  Base Design: 7, 8  Resolution: III
Runs: 16  Replicates: 1  Fraction: 1/8
Blocks: 1  Center pts (total): 0
```

* NOTE * Some main effects are confounded with two-way interactions.

Design Generators (before folding): D = AB, E = AC, F = BC, G = ABC

Folded on Factors: B

Alias Structure

I + ACE + AFG + CDG + DEF + ACDF + ADEG + CEFG

```
A + CE + FG + CDF + DEG + ACDG + ADEF + ACEFG
B + ABCE + ABFG + BCDG + BDEF + ABCDF + ABDEG + BCEFG
C + AE + DG + ADF + EFG + ACFG + CDEF + ACDEG
D + CG + EF + ACF + AEG + ACDE + ADFG + CDEFG
E + AC + DF + ADG + CFG + AEFG + CDEG + ACDEF
F + AG + DE + ACD + CEG + ACEF + CDFG + ADEFG
G + AF + CD + ADE + CEF + ACEG + DEFG + ACDFG
AB + BCE + BFG + BCDF + BDEG + ABCDG + ABDEF + ABCEFG
AD + CF + EG + ACG + AEF + CDE + DFG + ACDEFG
BC + ABE + BDG + ABDF + BEFG + ABCFG + BCDEF + ABCDEG
BD + BCG + BEF + ABCF + ABEG + ABCDE + ABDFG + BCDEFG
BE + ABC + BDF + ABDG + BCFG + ABEG + BCDEG + ABCDEF
BF + ABG + BDE + ABCD + BCEG + ABCEF + BCDFG + ABDEFG
BG + ABF + BCD + ABDE + BCEF + ABCEG + BDEFG + ABCDFG
ABD + BCF + BEG + ABCG + ABEF + BCDE + BDFG + ABCDEFG
```

This is still a Resolution III, (we haven't folded on all factors so we don't jump a resolution number). But look at the B factor which we folded. The main effect, B, is aliased with only the four way interactions and higher. Also, notice that all of the 2-way interactions with B are clear of other two-way interactions, so they become estimable. So by only folding on one factor, you get very good information on that factor and its interactions. However, it is still a resolution three design.

There are two purposes for folding; one is taking on another replicate for the purpose of moving to a higher resolution number. The other reason would be to isolate the information on a particular factor. Both of these would be done in the context of doing a sequential experiment, doing an analysis of that and then doing a second stage experiment. If you do this two stage experiment, performing a second stage based on the first experiment, you should also use stage as the block factor in the analysis.

All of these designs, even though they are fractions of an experiment, should be blocked, if they are done in stages.

One more example ...

Let's go to 8 factors. The minimal design now can not be eight observations but must be 16. This is a Resolution IV design.

Fractional Factorial Design

```
Factors: 8 Base Design: 8, 16 Resolution: IV
Runs: 16 Replicates: 1 Fraction: 1/16
Blocks: 1 Center pts (total): 0
```

Design Generators: E = BCD, F = ACD, G = ABC, H = ABD

Alias Structure (up to order 4)

```
I + ABCG + ABDH + ABEF + ACDF + ACEH + ADEG + AFGH + BCDE + BCFH + BDFG + BEGH
+ CDGH + CEGF + DEFH

A + BCG + BDH + BEF + CDF + CEH + DEG + FGH
B + ACG + ADH + AEF + CDE + CFH + DFG + EGH
C + ABG + ADF + AEH + BDE + BFH + DGH + EFG
D + ABH + ACF + AEG + BCE + BFG + CGH + EFH
E + ABF + ACH + ADG + BCD + BGH + CFG + DFH
F + ABE + ACD + AGH + BCH + BDG + CEG + DEH
G + ABC + ADE + AFH + BDF + BEH + CDH + CEF
H + ABD + ACE + AFG + BCF + BEG + CDG + DEF
AB + CG + DH + EF + ACDE + ACFH + ADFG + AEHG + BCDF + BCEH + BDEG + BFGH
AC + BG + DF + EH + ABDE + ABFH + ADGH + AEFG + BCDH + BCEF + CDEG + CFGH
AD + BH + CF + EG + ABCE + ABFG + ACGH + AEFH + BCDG + BDEF + CDEH + DFGH
AE + BF + CH + DG + ABCD + ABGH + ACFG + ADFH + BCEG + BDEH + CDEF + EFGH
AF + BE + CD + GH + ABCH + ABDG + ACEG + ADEH + BCFG + BDFH + CEFH + DEFG
AG + BC + DE + FH + ABDF + ABEH + ACDH + ACEF + BDGH + BEFG + CDFG + CEGH
AH + BD + CE + FG + ABCF + ABEG + ACDG + ADEF + BCGH + BEFH + CDFH + DEGH
```

This design has 4 generators BCDE, ACDF, ABCG and ABDH. It is a Resolution IV design and it is a design with 16 observations. OK, now we are going to assume that we can only run these experiments eight at a time so we have to block. We will use two blocks, and we will still have the same fractional design, eight factors in 16 runs but now we want to have two blocks.

We let Minitab pick the blocks:

Fractional Factorial Design

```
Factors: 8 Base Design: 8, 16 Resolution with blocks: III
Runs: 16 Replicates: 1 Fraction: 1/16
Blocks: 2 Center pts (total): 0
```

* NOTE * Blocks are confounded with two-way interactions.

Design Generators: E = BCD, F = ACD, G = ABC, H = ABD

Block Generators: AB

Alias Structure (up to order 4)

```
I + ABCG + ABDH + ABEF + ACDF + ACEH + ADEG + AFGH + BCDE + BCFH + BDFG + BEGH
+ CDGH + CEGF + DEFH

Blk = AB + CG + DH + EF + ACDE + ACFH + ADFG + AEHG + BCDF + BCEH + BDEG + BFGH

A + BCG + BDH + BEF + CDF + CEH + DEG + FGH
B + ACG + ADH + AEF + CDE + CFH + DFG + EGH
C + ABG + ADF + AEH + BDE + BFH + DGH + EFG
D + ABH + ACF + AEG + BCE + BFG + CGH + EFH
E + ABF + ACH + ADG + BCD + BGH + CFG + DFH
F + ABE + ACD + AGH + BCH + BDG + CEG + DEH
G + ABC + ADE + AFH + BDF + BEH + CDH + CEF
H + ABD + ACE + AFG + BCF + BEG + CDG + DEF
AC + BG + DF + EH + ABDE + ABFH + ADGH + AEFG + BCDH + BCEF + CDEG + CFGH
AD + BH + CF + EG + ABCE + ABFG + ACGH + AEFH + BCDG + BDEF + CDEH + DFGH
AE + BF + CH + DG + ABCD + ABGH + ACFG + ADFH + BCEG + BDEH + CDEF + EFGH
AF + BE + CD + GH + ABCH + ABDG + ACEG + ADEH + BCFG + BDFH + CEFH + DEFG
AG + BC + DE + FH + ABDF + ABEH + ACDH + ACEF + BDGH + BEFG + CDFG + CEGH
AH + BD + CE + FG + ABCF + ABEG + ACDG + ADEF + BCGH + BEFH + CDFH + DEGH
```

In this design, we have eight factors, 16 runs, and the same generators but now we need an additional generator, the block generator. Minitab is using AB as the block generator. Notice in the alias structure that the blocks are confounded with the AB term.

Notice also that the AB term does not show up as an estimable effect below. It would have been an effect we could have estimated but it is now confounded with blocks. So, one additional degree of freedom is used in this confounding with blocks.

The only choice the program had was to select one of these effects that were previously estimable and confound them with blocks. The program picked one of those 2-way interactions and this means blocks are now confounded with a 2-way interaction.

We can still block these fractional designs and it is useful to do this if you can only perform a certain number at a time. However, if you are doing sequential experimentation you should block just because you are doing it in stages.

In summary, when you fold over a Resolution III design **on all factors**, then you get a Resolution IV design. Look at the table of all possible designs in Minitab below:

Available Factorial Designs (with Resolution)															
Runs	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
4	Full	III													
8		Full	IV	II	III	III									
16			Full	V	IV	IV	IV	III	III	III	II	II	III	II	
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV	
64					Full	VII	V	IV	IV	IV	IV	IV	IV	IV	
128						Full	VIII	VI	V	V	IV	IV	IV	IV	

Available Resolution III Plackett-Burman Designs							
Factors	Runs	Factors	Runs	Factors	Runs	Factors	Runs
2-7	12,20,24,28,...,48	20-23	24,28,32,36,...,48	36-39	40,44,48		
8-11	12,20,24,28,...,48	24-27	28,32,36,40,44,48	40-43	44,48		
12-15	20,24,28,36,...,48	28-31	32,36,40,44,48	44-47	48		
16-19	20,24,28,32,...,48	32-35	36,40,44,48				

Help

OK

If you fold any of the red Resolution III designs you go to the next level, it has twice as many observations and becomes a Resolution IV design. If you fold many of the Resolution IV designs, even though you double the number of observations by folding, you are still at the Resolution IV.

Whereas, if you fold a Resolution III or IV design on one factor, you get better information on that factor and all its 2-way interactions would be clear of other 2-way interactions. Therefore, it serves that purpose well for Resolution III or IV designs.

8.4 - Plackett-Burman Designs

We looked at 2^{k-p} designs, which give us designs that have 8, 16, 32, 64, 128, etc. number of runs. We noted that all of these numbers are some fraction of $1/2^p$ of a 2^k design.

Loading [MathJax]/extensions/MathMenu.js

However, when you look at these numbers there is a pretty big gap between 16 and 32, 32 to 64, etc. We sometimes need other alternative designs besides these with a different number of observations.

A class of designs that allows us to create experiments with some number between these fractional factorial designs are the Plackett-Burman designs. Plackett-Burman designs exist for

$N = 12, [16], 20, 24, 28, [32], 36, 40, 44, 48, \dots$

... any number which is divisible by four. These designs are similar to Resolution III designs, meaning you can estimate main effects clear of other main effects. Main effects are clear of each other but they are confounded with other higher interactions.

Look at the table of available designs in Minitab. The Plackett-Burman designs are listed below:

Available Factorial Designs (with Resolution)														
Runs	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Full	II												
8		Full	IV	III	II	III								
16			Full	V	IV	IV	IV	II	III	III	II	II	III	II
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV
64					Full	VII	V	IV	IV	IV	IV	IV	IV	IV
128						Full	VIII	VI	V	V	IV	IV	IV	IV

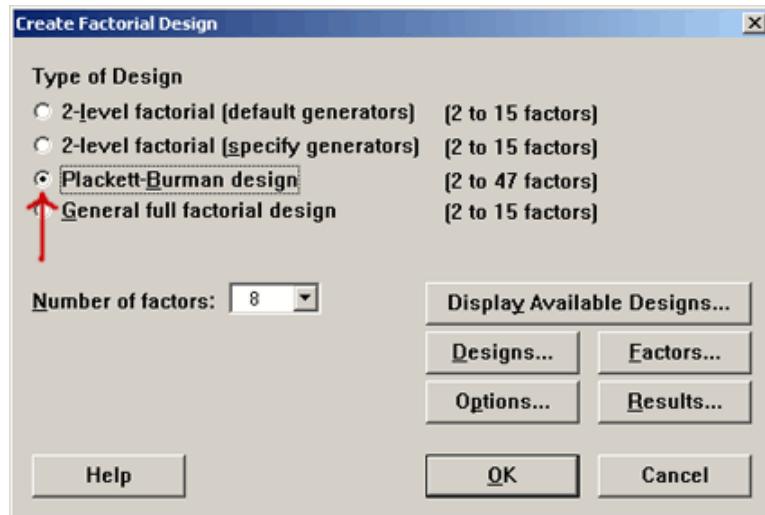
Available Resolution III Plackett-Burman Designs											
Factors	Runs	Factors	Runs	Factors	Runs						
2-7	12,20,24,28,...,48	20-23	24,28,32,36,...,48	36-39	40,44,48						
8-11	12,20,24,28,...,48	24-27	28,32,36,40,44,48	40-43	44,48						
12-15	20,24,28,36,...,48	28-31	32,36,40,44,48	44-47	48						
16-19	20,24,28,32,...,48	32-35	36,40,44,48								

Help

OK

So, if you have 2 to 7 factors you can create a Plackett-Burman design with 12, 20, 24, ... up to 48 observations. Of course, if you have 7 factors with eight runs then you have a saturated design.

In the textbook there is a brief shortcut way of creating these designs, but in Minitab we simply select the Plackett-Burman option.



You specify how many runs and how many factors are in your experiment. If we specified eight factors and 12 runs, we get a design that looks like this:

StdOrder	RunOrder	PtType	Blocks	A	B	C	D	E	F	G	H
1	1	1	1	1	-1	1	-1	-1	-1	1	1
2	2	1	1	1	1	-1	1	-1	-1	-1	1
3	3	1	1	-1	1	1	-1	1	-1	-1	-1
4	4	1	1	1	-1	1	1	-1	1	-1	-1
5	5	1	1	1	1	-1	1	1	-1	1	-1
6	6	1	1	1	1	1	-1	1	1	-1	1
7	7	1	1	-1	1	1	1	-1	1	1	-1
8	8	1	1	-1	-1	1	1	1	-1	1	1
9	9	1	1	-1	-1	-1	1	1	1	-1	1
10	10	1	1	1	-1	-1	-1	1	1	1	-1
11	11	1	1	-1	1	-1	-1	-1	1	1	1
12	12	1	1	-1	-1	-1	-1	-1	-1	-1	-1

This look very much like the designs we had before. In this case we have eight factors, A through H, each with two levels. And each factor is defined by a 12 run design, 6 pluses and 6 minuses. Again, these are contrasts. Half of the observations at a high level and half at the low-level, and if you take any two columns they are orthogonal to each other. So, these are an orthogonal set of columns just as we had for the 2^{k-p} design. If you take the product of any two of these and add them up, the sum of the products you get is zero.

Because these are orthogonal contrasts we get clean information on all main effects. The main effects are not confounded as required by the orthogonality of those columns.

Here is a quick way to manually create this type of design. First of all, one would fill out the first column of the design table, this would be column A. Then you can create the B column by taking the last element for permuting and then slide everything down. This process can be repeated for each column of factors needed in the design. Click the '**Create Design**' button below to see how this works:

You can generate these designs by just knowing the first 11 elements, permuting these into the next column and adding an additional row of minuses across the bottom. It has this cyclical pattern and it works for most of these types of designs, (12, 20, 24, 36, but not for 28!). Here is what it looks like for 20 runs with 16 factors:

Plackett - Burman Design																	
Factors:	16								Replicates:	1							
Base runs:	20								Total runs:	20							
Base blocks:	1								Total blocks:	1							
Design Table																	
Run	Blk	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q
1	1	+	-	+	+	-	-	-	+	-	+	-	+	+	+	+	
2	1	+	+	-	+	+	-	-	-	+	-	+	-	+	+	+	
3	1	-	+	+	-	+	+	-	-	-	+	-	+	-	+	+	
4	1	-	-	+	+	-	+	-	-	-	-	+	-	+	-	+	
5	1	+	-	-	+	-	+	-	-	-	-	-	+	-	+	-	
6	1	+	+	-	-	+	+	-	-	-	-	-	-	+	-	+	
7	1	+	+	+	-	+	+	-	+	-	-	-	-	-	+	-	
8	1	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	
9	1	-	+	+	+	-	-	+	-	+	+	-	-	-	-	-	
10	1	+	-	+	+	+	-	-	+	+	-	+	+	-	-	-	
11	1	-	+	-	+	+	+	-	-	+	+	-	+	+	-	-	
12	1	+	-	+	-	+	+	+	-	+	+	-	+	+	-	-	
13	1	-	-	+	-	+	+	+	-	-	+	+	-	+	+	-	
14	1	-	-	-	+	-	+	+	+	-	-	+	+	-	+	-	
15	1	-	-	-	-	+	-	+	+	+	-	-	+	+	-	-	
16	1	-	-	-	-	+	-	+	+	+	+	-	-	+	+	-	
17	1	+	-	-	-	-	+	-	+	+	+	+	-	-	-	+	
18	1	+	+	-	-	-	-	+	-	+	-	+	+	+	-	-	
19	1	-	+	+	-	-	-	+	-	+	-	+	+	+	+	-	
20	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

The cyclical pattern is a result of number theory properties that generate these orthogonal arrays. There is a lot of mathematical research behind these designs to achieve a matrix with orthogonal columns which is what we need.

We point out that these designs are a little different than the 2^{k-p} designs. When you have a 2^{k-p} design you have an alias structure that confounds some factors with other factors. Let's look at two examples to illustrate this.

Example: [FF2LevelCorr.MPJ](#) [2]

The first is a fractional factorial, 4 factor design, Resolution IV with one generator ABCD or D = ABC. From this design we get an alias structure that we are familiar with. Main effects are aliased with 3-way interactions which means that they are completely confounded with those factors. Two-way interactions are confounded with each other.

Fractional Factorial Design					
Factors:	4	Base Design:	4, 8	Resolution:	IV
Runs:	8	Replicates:		1	Fraction: 1/2
Blocks:	1	Center pts (total):	0		
Design Generators: D = ABC					
Alias Structure					
I + ABCD					
A + BCD					
B + ACD					
C + ABD					
D + ABC					
AB + CD					
AC + BD					
AD + BC					

Let's look at the correlation among these factors, A, B, C and D, and then a couple of interaction columns.

Correlations: A, B, C, D, AB, ABC, CD						
	A	B	C	D	AB	ABC
B	0.000					
	1.000					
C	0.000	0.000				
	1.000	1.000				
D	0.000	0.000	0.000			
	1.000	1.000	1.000			
AB	0.000	0.000	0.000	0.000		
	1.000	1.000	1.000	1.000		
ABC	0.000	0.000	0.000	1.000	0.000	
	1.000	1.000	1.000	*	1.000	
CD	0.000	0.000	0.000	0.000	1.000	0.000
	1.000	1.000	1.000	1.000	*	1.000

This is just a simple Pearson correlation. What Minitab gives us is the coefficient and the p-value. We can ignore the p-values because we are not really interested in testing, however a correlation between A and B = 0, A and C = 0, A and D = 0, etc. The correlation between **all** these factors = 0 because of the orthogonality.

Look back up at the alias structure and you will see the D is confounded with ABC. As we look back at the correlation table the correlation between D and ABC = 1. The correlation between two factors that are confounded = 1. This is appropriate because they are completely correlated with each other. Therefore, in these 2^{k-p} designs we can see through correlation that factors are either orthogonal (correlation = 0) or they are completely confounded (correlation = 1).

Next, let's look at the Plackett-Burman designs and see how this differs. Below, we have created a design for 9 factors, 12 runs and we are looking at the correlation among the main effects, A, B, C, D, and E.

Plackett - Burman Design				
Factors:	9	Replicates:	1	
Base runs:	12	Total runs:	12	
Base blocks:	1	Total blocks:	1	
Correlations: A, B, C, D, E				
	A	B	C	
B	0.000			
	1.000			
C	0.000	0.000		
	1.000	1.000		
D	0.000	0.000	0.000	
	1.000	1.000	1.000	
E	0.000	0.000	0.000	0.000
	1.000	1.000	1.000	1.000

These are the main factors themselves already set in orthogonal columns so these correlations = 0. If we look at the next design, however, in this case we have the 12 runs and then we have created new 2-way interactions through multiplication of the factors already determined. Again, ignoring the p-values we produced a correlation matrix, (partially displayed below).

Correlations: AB, A, B, C, D, F, G, H, J, AB, AC, AD									
	AB	A	B	C	D	F	G	H	J
A	0.000								
	1.000								
B	0.000	0.000							
	1.000	1.000							
C	-0.333	0.000	0.000						
	0.290	1.000	1.000						
D	0.333	0.000	0.000	0.000					
	0.290	1.000	1.000	1.000					
F	-0.333	0.000	0.000	0.000	0.000				
	0.290	1.000	1.000	1.000	1.000				
G	-0.333	0.000	0.000	0.000	0.000	0.000			
	0.290	1.000	1.000	1.000	1.000	1.000			
H	0.333	0.000	0.000	0.000	0.000	0.000	0.000		
	0.290	1.000	1.000	1.000	1.000	1.000	1.000		
J	-0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	0.290	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
AB	1.000	0.000	0.000	-0.333	0.333	-0.333	-0.333	0.333	-0.333
*		1.000	1.000	0.290	0.290	0.290	0.290	0.290	0.290
AC	0.000	0.000	-0.333	0.000	-0.333	0.333	-0.333	0.333	-0.333
	1.000	1.000	0.290	1.000	0.290	0.290	0.290	0.290	0.290
AD	0.000	0.000	0.333	-0.333	0.000	-0.333	-0.333	-0.333	-0.333
	1.000	1.000	0.290	0.290	1.000	0.290	0.290	0.290	0.290

A is orthogonal to every other factor, the correlation value = 0. B is not correlated with all the other main effects, where correlation = 0, but with some of these two-way interactions the correlation = 0.333. This shows partial confounding with the two-way interaction. Likewise, C has partial confounding with AB and AD. D is partially confounded with AB and AC. F is partially confounded with AB and AC and AD, and so forth.

Loading [MathJax]/extensions/MathMenu.js

Plackett-Burman designs have partial confounding, not complete confounding, with the 2-way and 3-way and higher interactions. Although they have this property that some effects are orthogonal they do not have the same structure allowing complete or orthogonal correlation with the other two way and higher order interactions.

Like other Resolution II designs, these designs are also good for screening for important factors. But remember, in a Resolution II design a main effect might look important, because some combination of interactions is important and the main effect itself might not be the important effect.

If you assume that your interactions = 0 or are not important these are great designs. If your assumption is wrong and there are interactions, then it could show up as influencing one or the other main effects. These designs are very efficient with small numbers of observations and useful, but remember the caveat, you are assuming that the main effects are going to show up as larger effects than interactions so that they will dominate the interaction effects.

Using Minitab we can ask for up to 47 factors. In doing so you want to select a sufficient number of runs over the number of factors so that you have a reasonable number of degrees of freedom for error. At this stage a statistical test really isn't that important, you are just screening for a place to start.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/48>

Links:

- [1] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson08/Ex6-2.MTW>
- [2] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson08/FF2levelCorr.MPJ>

Loading [MathJax]/extensions/MathMenu.js



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 9: 3-level and Mixed-level Factorials and Fractional Factorials

Lesson 9: 3-level and Mixed-level Factorials and Fractional Factorials

Learning Objectives and Outcomes

By the end of this lesson, students are assumed to know

- Application of 3^k factorial designs, the interaction components and relative degrees of freedom
- How to perform blocking of 3^k designs in 3^p number of blocks and how to choose the effect(s) which should be confounded with blocks
- Concept of “Partial Confounding” in replicated blocked designs and its advantages
- How to generate reasonable 3^{k-p} fractional factorial designs and understand the alias structure
- The fact that Latin square and Graeco-Latin square designs are special cases of 3^k fractional factorial design
- Mixed level factorial designs and their applications

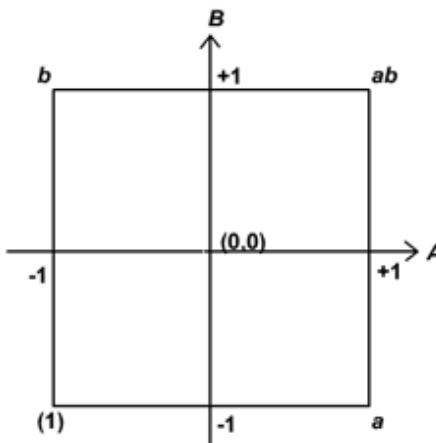
Introduction

Basic material

These designs are a generalization of the 2^k designs. We will continue to talk about coded variables so we can describe designs in general terms, but in this case we will be assuming in the 3^k designs that the factors are all quantitative. With 2^k designs we weren't as strict about this because we could have either qualitative or quantitative factors. Most 3^k designs are only useful where the factors are quantitative. With 3^k designs we are moving from screening factors to analyzing them to understand what their actual response function looks like.

With 2 level designs, we had just two levels of each factor. This is fine for fitting a linear, straight line relationship. With three level of each factor we now have points at the middle so we will be able to fit curved response functions, i.e. quadratic response functions. In two dimensions with a square design space, using a 2^k design we simply had corner points, which defined a square that looked like this:

Loading [MathJax]/extensions/MathZoom.js



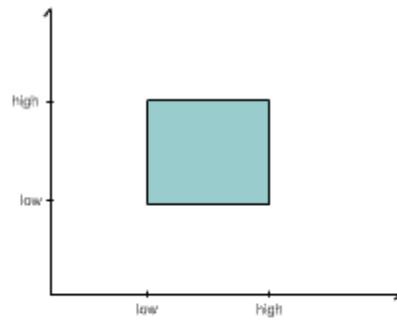
In three dimensions the design region becomes a cube and with four or more factors it is a hypercube which we can't draw.

We can label the design points, similar to what we did before – see the columns on the left. However for these design we prefer the other way of coding, using $\{0,1,2\}$ which is a generalization of the $\{-1,1\}$ coding that we used in the 2^k designs. This is shown in the columns on the right in the table below:

A	B	A	B
-	-	0	0
0	-	1	0
+	-	2	0
-	0	0	1
0	0	1	1
+	0	2	1
-	+	0	2
0	+	1	2
+	+	2	2

For either method of coding, the treatment combinations represent the actual values of X_1 and X_2 , where there is some high level, a middle level and some low level of each factor. Visually our region of experimentation or region of interest is highlighted in the figure below when $k = 2$:

Loading [MathJax]/extensions/MathZoom.js



If we look at the analysis of variance for a $k = 2$ experiment with n replicates, where we have three levels of both factors we would have the following:

AOV	df
A	2
B	2
A x B	4
Error	9(n-1)
Total	9n-1

Important idea used for confounding and taking fractions

How we consider three level designs will parallel what we did in two level designs, therefore we may confound the experiment in incomplete blocks or simply utilize a fraction of the design. In two-level designs, the interactions each have 1 d.f. and consist only of +/- components, so it is simple to see how to do the confounding. Things are more complicated in 3 level designs, since a p-way interaction has 2^p d.f. If we want to confound a main effect (2 d.f.) with a 2-way interaction (4 d.f.) we need to partition the interaction into 2 orthogonal pieces with 2 d.f. each. Then we will confound the main effect with one of the 2 pieces. There will be 2 choices. Similarly, if we want to confound a main effect with a 3-way interaction, we need to break the interaction into 4 pieces with 2 d.f. each. Each piece of the interaction is represented by a psuedo-factor with 3 levels. The method given using the Latin squares is quite simple . There is some clever modulus arithmetic in this section, but the details are not important. The important idea is that just as with the 2^k designs, we can purposefully confound to achieve designs that are efficient either because they do not use the entire set of 3^k runs or because they can be run in blocks which do not disturb our ability to estimate the effects of most interest.

Following the text, for the A*B interaction, we define the pseudo factors, which are called the AB component and the AB² component. These components could be called pseudo-interaction effects. The two components will be defined as a linear combination as follows, where X_1 is the level of factor A and X_2 is the level of factor B using the {0,1,2} coding

Let the AB component be defined as

$$L_{AB} = X_1 + X_2 \pmod{3}$$

and the AB^2 component will be defined as:

$$L_{AB^2} = X_1 + 2X_2 \pmod{3}$$

Using these definitions we can create the pseudo-interaction components. Below you see that the AB levels are defined by L_{AB} and the AB^2 levels are defined by L_{AB^2} .

A	B	AB	AB^2
0	0	0	0
1	0	1	1
2	0	2	2
0	1	1	2
1	1	2	0
2	1	0	1
0	2	2	1
1	2	0	2
2	2	1	0

This table has entries {0, 1, 2} which allow us to confound a main effect or either component of the interaction A^*B . Each of these main effects or pseudo interaction components have three levels and therefore 2 degrees of freedom.

This section will also discuss partitioning the interaction SS's into 1 d.f. sums of squares associated with a polynomial, however this is just polynomial regression. This method does not seem to be readily applicable to creating interpretable confounding patterns.

9.1 - 3^k Designs in 3^p Blocks

Let's begin by taking the 3^k designs and we will describe partitioning where you take one replicate of the design and put it into blocks. We will then take that structure and look at 3^{k-p} factorials. These designs are not used for screening as the 2^k designs were; rather with three levels we begin to think about response surface models. Also, 3^k designs become very large as k gets large. With just four factors a complete factorial is already 81 observations, i.e. $N = 3^4$. In general we won't consider these designs for very large k , but we will point out some very interesting connections that these designs reveal.

Reiterating what was said in the introduction, consider the two factor design 3^2 with factors A and B, each at 3 levels. We denote the levels 0, 1, and 2. The $A \times B$ interaction, with 4 degrees of freedom, can be split into two orthogonal components. One way to define the components is that AB component will be defined as a linear combination as follows:

Loading [MathJax]/extensions/MathZoom.js

$$L_{AB} = X_1 + X_2 \pmod{3}$$

and the AB^2 component will be defined as:

$$L_{AB^2} = X_1 + 2X_2 \pmod{3}$$

A	B	AB	AB²
0	0	0	0
1	0	1	1
2	0	2	2
0	1	1	2
1	1	2	0
2	1	0	1
0	2	2	1
1	2	0	2
2	2	1	0

In the table above for the AB and the AB^2 components we have 3 0's, 3 1's and 3 2's, so this modular arithmetic gives us a balanced set of treatments for each component. Note that we could also find the A^2B and A^2B^2 components but when you do the computation you discover that $AB^2=A^2B$ and $AB=A^2B^2$.

We will use this to construct the design as shown below.

We will take one replicate of this design and partition it into 3 blocks. Before we do, let's consider the analysis of variance table for this single replicate of the design.

AOV	df
A	3-1=2
B	3-1=2
A × B	2*2=4
AB	3-1=2
AB²	3-1=2

We have partitioned the $A \times B$ interaction into AB and AB^2 , the two components of the interaction, each with 2 degrees of freedom. So, by using modular arithmetic, we have partitioned the 4 degrees of freedom into two sets, and these are orthogonal to each other. If you create two dummy variables for each of these factors, A, B, AB and AB^2 you would see that each of these sets of dummy variables are orthogonal to the other.

Loading [MathJax]/extensions/MathZoom.js

These pseudo components can also be manipulated using a symbolic notation. This is included here for completeness, but it is not something you need to know to use or understand confounding. Consider the interaction between AB and AB^2 . Thus $AB \times AB^2$ which gives us A^2B^3 which using modular (3) arithmetic gives us $A^2B^0 = A^2 = (A^2)^2 = A$.

Therefore, the interaction between these two terms gives us the main effect. If we wanted to look at a term such as A^2B or A^2B^2 , we would reduce it by squaring it which would give us: $(A^2B)^2 = AB^2$ and likewise $(A^2B^2)^2 = AB$. We never include a component that has an exponent on the first letter because by squaring it we obtain an equivalent component. This is just a way of partitioning the treatment combinations and these labels are just an arbitrary identification of them.

Let's now look at the one replicate where we will *confound the levels of the AB component with our blocks*. We will label these 0, 1, and 2 and we will put our treatment pairs in blocks from the following table.

A	B	AB	AB^2
0	0	0	0
1	0	1	1
2	0	2	2
0	1	1	2
1	1	2	0
2	1	0	1
0	2	2	1
1	2	0	2
2	2	1	0

Now we assign the treatment combinations to the blocks, where the pairs represent the levels of factors A and B.

L_{AB}		
0	1	2
0, 0	1, 0	2, 0
2, 1	0, 1	1, 1
1, 2	2, 2	0, 2

This is how we get these three blocks confounded with the levels of the L_{AB} component of interaction.

Now, let's assume that we have four reps of this experiment - all the same - with AB confounding with blocks using the L_{AB} . (each replicate is assigned to 3 blocks with AB confounded with blocks). We have defined one rep by confounding the AB component, and ~~then we will do the same with more reps.~~

Let's take a look at the AOV resulting from this experiment:

AOV	df
Rep	4-1=3
Blk = AB	3-1=2
Rep × AB	3*2=6
<i>Inter-block Total</i>	11
A	3-1=2
B	3-1=2
A × B = AB²	3-1=2
Error	(2+2+2)* (4-1)=18
Total	3*3*4-1=35

Note that Rep as an overall block has 3 *df*. Within reps we have variation among the 3 blocks, which are the AB levels - this has 2 *df*. Then we have Rep by blk or Rep by AB which has 6 *df*. This is the inter-block part of the analysis. These 11 degrees of freedom represents the variation among the 12 blocks (3*4).

Next we consider the intra-block part: A with 2 *df*, B with 2 *df* and the A × B or AB² component that also has 2 *df*. Finally we have error, which we can get by subtraction, (36 observations = 35 total *df*, 35 - 17 = 18 *df*). Another way to think about the Error is the interaction between the treatments and reps which is 6 × 3 = 18, which is the same logic as in a randomized block design, where the SSE is (a-1)(b-1). A possible confusion here is using the terminology of blocks at two levels, the reps are at an overall level, and then within each rep we have the smaller blocks which are confounded with the AB component.

We now examine another experiment, this time confounding the AB² factor. We can construct another design using this component as our generator to confound with blocks.

A	B	AB	AB²
0	0	0	0
1	0	1	1
2	0	2	2
0	1	1	2
1	1	2	0
2	1	0	1
0	2	2	1
1	2	0	2

Loading [MathJax]/extensions/MathZoom.js

2 2 1 0

Using the AB^2 then gives us the following treatment pairs (A,B) assigned to 3 blocks:

L_{AB^2}		
0	1	2
0, 0	1, 0	2, 0
1, 1	2, 1	0, 1
2, 2	0, 2	1, 2

This partitions all nine of the treatment combinations into the three blocks.

Partial Confounding (optional section)

We now consider a combination of these experiments, in which we have 2 reps confounding AB and 2 reps confounding AB^2 . We again will have 4 reps but our AOV will look a little different:

AOV	df
Rep	$4-1=3$
Blk = AB	$3-1=2$
Blk = AB^2	$3-1=2$
Rep \times AB	$(2-1)*2=2$
<i>Inter-block Error</i>	4
Rep \times AB^2	$(2-1)*2=2$
<i>Inter-block Total</i>	11
A	$3-1=2$
B	$3-1=2$
A \times B	$2*2=4$
AB	$3-1=2$
AB^2	$3-1=2$
Error	$2*(4-1)+2*$ $(4-1)+2*$ $(2-1)+2*$ $(2-1)=16$
Total	$3*3*4-1=35$

There are only two reps with AB confounded, so $Rep \times AB = (2-1) * (3-1) = 2 df$. The same is true for the AB^2 component. This gives us the same 11 df among the 12 blocks. In the intra-block section, we can estimate A and B, so they will have 2 df. A \times B will have 4 df

Loading [MathJax]/extensions/Math2.js

now, and if we look at what this is in terms of the AB and the AB^2 component each accounts for 2 df . Then we have Error with 16 df and the total stays the same. The 16 df comes from the unconfounded effects - (A: $2 \times 3 = 6$ and B: $2 \times 3 = 6$) - that's 12 of these df , plus each of the AB and the AB^2 components which are confounded in two reps, and unconfounded in the other two reps - ($2 * (2-1) = 2$ for AB and $2 * (2-1) = 2$ for AB^2) - which accounts for the remaining 4 of the total 16 df for error.

We could determine the Error df simply by subtracting from the Total df , but, if it is helpful to think about randomized block designs where you have blocks and treatments and the error is the interaction between them. Note that here we use the term replicates instead of blocks, so actually we consider replicates as sort of super-blocks. In this case error would be the interaction between replicates and unconfounded treatments. This RCBD framework is a foundational structure that we use again and again in experimental design.

This is a good example of the benefit of partial confounding because the interaction of the pseudo factors are confounded in only half of the design, so we can estimate the interaction A*B from the other half. You get overall exactly half the information on the interaction from this partially confounded design.

Confounding a main effect (an important idea)

Now let's think further outside of the box. What if we confound the main effect A? What would this do to our design? What kind of experimental design would this be?

Now we define or construct our blocks by using levels of A from the table above. A single replicate of the design would look like this.

A	0	1	2
0, 0	1, 0	2, 0	
0, 1	1, 1	2, 1	
0, 2	1, 2	2, 2	

Then we could replicate this design four times. Let's consider an agricultural application and say that A = irrigation method, B = crop variety, and the Blocks = whole plots of land to which we apply the irrigation type. By confounding a main effect we're going to get a split-plot design in which the analysis will look like this:

AOV	df
Reps	3
A	2
Rep × A	6
Inter-block Total	11

Loading [MathJax]/extensions/MathZoom.js

B	2
A × B	4
Error	18
Total	35

In this design there are four reps (*3 df*), and the blocks within reps are actually the levels of A which has *2 df*, Rep × A has *6 df*. The interblock part of the analysis here is just a randomized complete block analysis of four reps, three treatments and their interactions. The intra-block part contains B which has *2 df*, and the A × B interaction which has *4 df*. Therefore this is another way to understand a split plot design, where you confound one of the main effects .

More examples of confounding.

Let's look at the $k = 3$ case - an increase in the number of treatments by one. Here we will look at a 3^3 design confounded in 3^1 blocks, or we could look at a 3^3 design confounded in 3^2 blocks. In a 3^3 design confounded in three blocks, each block would have nine observations now instead of three.

To create the design shown in Figure 9-7 below, follow the following commands:

Stat > DOE > Factorial > Create Factorial Design

- click on General full factorial design,
- set Number of factors to 3
- set Number of levels of each factor to 3
- under options, unclick the randomize button
- Then use Calc menu and subtract 1 from each of column A, B, and C (We could have initially made levels 0, 1 and 2).

Now the levels of the three factors are coded with (0, 1, 2). We are ready to calculate the pseudo factor, AB^2C^2 , which we will abbreviate as AB2C2.

Label the next blank column, AB2C2. Again, using the Calc menu, let $AB2C2 = \text{Mod}(A + 2*B + 2*C, 3)$, which creates the levels of the pseudo factor $L_{AB^2C^2}$ described on the page 371.

Here is a link to a Minitab project file that implements this: [Figure-9-7.MPJ \[1\]](#)

Let's look at the $k = 3$ case - a 3^3 design confounded in 3^1 blocks. In a 3^3 design confounded in three blocks, each block would have nine observations now.

0	0	0
1	0	0
2	0	0
0	1	0
1	1	0
2	1	0
0	2	0
1	2	0
2	2	0
0	0	1
1	0	1
2	0	1
0	1	1
1	1	1
2	1	1
0	2	1
1	2	1
2	2	1
0	0	2
1	0	2
2	0	2
0	1	2
1	1	2
2	1	2
0	2	2
1	2	2
2	2	2

With 27 possible combinations, without even replicating, we have 26 *df*. These can be broken down in the following manner:

AOV	<i>df</i>
A	2
B	2

Loading [MathJax]/extensions/MathZoom.js

C	2
A × B	4
A × C	4
B × C	4
A × B × C	8
Total	26

The main effects all have 2 *df*, the three two-way interactions all have 4 *df*, and the three-way interaction has 8 *df*. If we think about what we might confound with blocks to construct a design we typically want to pick a higher order interaction.

The three way interaction $A \times B \times C$ can be partitioned into four orthogonal components labeled, ABC, AB^2C , ABC^2 and AB^2C^2 . These are the only possibilities where the first letter has exponent = 1. When the first letter has an exponent higher than one, for instance A^2BC , to reduce it we can first square it, $A^4B^2C^2$, and then using mod 3 arithmetic on the exponent get AB^2C^2 , i.e. a component we already have in our set. These four components partition the 8 degrees of freedom and we can define them just as we have before. For instance:

$$L_{ABC} = X_1 + X_2 + X_3 \pmod{3}$$

This column has been filled out in the table below in two steps, the first column carries out the arithmetic (sum) and the next column applies the mod 3 arithmetic:

A	B	C	A+B+C	L_{ABC}
0	0	0	0	0
1	0	0	1	1
2	0	0	2	2
0	1	0	1	1
1	1	0	2	2
2	1	0	3	0
0	2	0	2	2
1	2	0	3	0
2	2	0	4	1
0	0	1	1	1
1	0	1	2	2
2	0	1	3	0
0	1	1	2	2
1	1	1	3	0
Loading [MathJax]/extensions/MathZoom.js		1	4	1

0	2	1	3	0
1	2	1	4	1
2	2	1	5	2
0	0	2	2	2
1	0	2	3	0
2	0	2	4	1
0	1	2	3	0
1	1	2	4	1
2	1	2	5	2
0	2	2	4	1
1	2	2	5	2
2	2	2	6	0

Using the L_{ABC} component to assign treatments to blocks we could write out the following treatment combinations for one of the reps:

L_{ABC}		
0	1	2
0, 0, 0	1, 0, 0	2, 0, 0
2, 1, 0	0, 1, 0	1, 1, 0
1, 2, 0	2, 2, 0	0, 2, 0
2, 0, 1	0, 0, 1	1, 0, 1
1, 1, 1	2, 1, 1	0, 1, 1
0, 2, 1	1, 2, 1	2, 2, 1
1, 0, 2	2, 0, 2	0, 0, 2
0, 1, 2	1, 1, 2	2, 1, 2
2, 2, 2	0, 2, 2	1, 2, 2

This partitions the 27 treatment combinations into three blocks. The ABC component of the three-way interaction is confounded with blocks.

If we performed one block of this design perhaps because we could not complete 27 runs in one day - we might be able to accommodate nine runs per day. So perhaps on day one we use the first column of treatment combinations, on day two we used the second column of treatment combinations and on day three we use the third column of treatment combinations. This would conclude one complete replicate of the experiment. We can then continue a similar approach in the next three days to complete the second replicate. So, in twelve days four reps would have been performed.

Loading [Mathjax]/extensions/MathZoom.js

How would we analyze this? We would use the same structure.

AOV	df
Rep	4-1=3
ABC = Blk	2
Rep × ABC	6
A	2
B	2
C	2
A × B	4
A × C	4
B × C	4
A × B × C	6
AB^2C	2
ABC^2	2
AB^2C^2	2
Error	72
Total	108-1=107

We have (4 - 1) or 3 *df* for Rep, ABC is confounded with blocks so the ABC component of blocks has 2 *df*, the Rep by ABC (3×2) has 6 *df*. In summary to this point we have twelve of these blocks in our 4 reps so there are 11 *df* in our inter-block section of the analysis. Everything else follows below. The main effects have 2 *df*, the two-way interactions have 4 *df*, and the $A \times B \times C$ would have 8 *df*, but it only has 6 *df* because the ABC component is gone, leaving the other three components with 2 *df* each.

Error will be the unconfounded terms times the number of reps -1, or $24 \times (4 - 1) = 72$.

Likewise, $L_{AB^2C} = X_1 + 2X_2 + X_3 \pmod{3}$ can also be defined as another pseudo component in a similar fashion.

9.2 - 3^k Designs in 3^p Blocks cont'd.

We again start out with a 3^3 design which has 27 treatment combinations and assign them to 3 blocks. What we want to do in this lesson, going beyond the 3^2 design, is to describe the AOV for this 3^3 design. Then we also want to look at the connection between confounding in blocks and 3^{k-p} fractional factorials. This story will be very similar to what we did in the 2^{k-p} designs previously. There is a direct analogue here that you will see.

Loading [MathJax]/extensions/MathZoom.js

From the previous section we had the following design, 3^3 treatments in 3 blocks with the ABC pseudo factor confounded with blocks, i.e.,

L_{ABC}		
0	1	2
0, 0, 0	1, 0, 0	2, 0, 0
2, 1, 0	0, 1, 0	1, 1, 0
1, 2, 0	2, 2, 0	0, 2, 0
2, 0, 1	0, 0, 1	1, 0, 1
1, 1, 1	2, 1, 1	0, 1, 1
0, 2, 1	1, 2, 1	2, 2, 1
1, 0, 2	2, 0, 2	0, 0, 2
0, 1, 2	1, 1, 2	2, 1, 2
2, 2, 2	0, 2, 2	1, 2, 2

The three (color coded) blocks are determined by the levels of the ABC component of the three-way interaction which is confounded with blocks. If we only had one replicate of this design we would have 26 degrees of freedom. So, let's pretend that this design is Rep 1 and we will add Reps 2, 3, 4, just as we did with the two factor case. This would result in a total of 12 blocks.

If we did this as our basic design and replicate it three more times our AOV would look like the following:

AOV	df
Reps	3
Blocks(Rep)	$4 \times (3-1) = 8$
ABC	2
Rep × ABC	6
A	2
B	2
C	2
A × B	4
A × C	4
B × C	4
A × B × C	6
Error	72

Loading [MathJax]/extensions/MathZoom.js

Total	107
--------------	-----

We would have Reps with 3 *df*, blocks nested in Reps with 2 *df* \times 4 Reps = 8 *df*, then we would have all of the unconfounded effects as shown above. The A \times B \times C would only have 6 *df* because one component (ABC) is confounded with blocks. Error is $24 \times 3 = 72$ *df* and our total is $(3^3 \times 4) - 1 = 107$ *df*.

Now, we have written Blocks(Rep) with 8 *df* equivalently (in the blue font above) as ABC with 2 *df*, and Rep \times ABC with 6 *df*, but now we are considering the 0, 1, and 2 as levels of the ABC factor. In this case ABC is one component of the interaction and still has meaning in terms of the levels of ABC, just not very interesting since it is part of the three way interaction. Had we confounded a main effect with blocks, we certainly would have wanted to analyze it, as seen above where a main effect was confounded with blocks. Then it had an important meaning and you certainly would want to pull this out and be able to test it.

Now we have a total of $3 \times 4 = 12$ blocks and the 11 *df* among them are the interblock part of the analysis. If we averaged the nine observations in each block and got a single number, we could analyze those 12 numbers and this would be the inter-block part of this analysis.

How do we accomplish this in Minitab? If you have a set of data labeled by rep, blocks, and A, B, and C, then you would have everything you need and you can fit a general linear model:

$$Y = \text{Rep} \text{ } \text{Blocks}(\text{Rep}) \text{ } A | B | C$$

This would generate the analysis since A | B | C expands to all main effects and all interactions in GLM of Minitab.

An Alternate Design - Partial Confounding

In thinking about how this design should be implemented a good idea would be to follow this first Rep with a second Rep that confounds L_{AB^2C} , confound L_{ABC^2} in Rep three, and finally confound $L_{AB^2C^2}$ in fourth Rep. Now we could estimate all four components of the three-way interactions because in three of the Reps they would be unconfounded. There is no information available in the way we had approached it previously. There is lots of information available using this partial confounding strategy of the three-way interactions.

3^{k-p} designs - Fractional Factorial 3-level Designs

The whole point of looking at this structure is because sometimes we want to only conduct a fractional factorial. We sometimes can't afford 27 runs, certainly not 108 runs. Often we can only afford a fraction of the design. So, let's construct a 3^{3-1} design which is a 1/3 fraction of a 3^3 design. In this case, $N = 3^{3-1} = 3^2 = 9$, the total number of runs. This is a small, compact design. For the case where we use the L_{ABC} pseudo factor to create the design, we would use just one block of the design above, and below here is the alias structure:

$$I = ABC$$

$$A = A \times ABC = (A^2BC) = AB^2C^2$$

Loading [MathJax]/extensions/MathZoom.js

$$\begin{aligned}
 A &= A \times (ABC)^2 = A^3B^2C^2 = (B^2C^2)^2 = BC \\
 B &= B \times ABC = AB^2C \\
 B &= B \times (ABC)^2 = A^2B^3C^2 = AC \\
 C &= C \times ABC = ABC^2 \\
 C &= C \times (ABC)^2 = A^2B^2C^3 = (A^2B^2)^2 = AB
 \end{aligned}$$

Here A is confounded with part of the 3-way and part of the 2-way interaction, likewise for B and for C. This design only has 9 observations. It has A, B and C main effects estimable and if we look at the AOV we only have nine observations so we can only include the main effects:

AOV	df
A	2
B	2
C	2
Error	2
Total	8

Below is the 3^3 design where we partitioned the treatment combinations for one Rep of the experiment using the levels of L_{ABC} . It is of interest to notice that a 3^{3-1} fractional factorial design is also a design we previously discussed. Can you guess what it is?

If we look at the first light blue column, we can call A the row effect, B the column effect and C the Latin letters, or in this case 0, 1, 2. We would use this procedure to assign the treatments to the square. This is how we get a 3×3 Latin square. So, a one third fraction of a 3^3 design is the same as a 3×3 Latin square design that we saw earlier in this course. Click on the 'Start' button above to see how this works.

It is important to see the connection here. We have three factors, A, B, C, and before when we talked about Latin squares, two of these were blocking factors and the third was the treatment factor. We could estimate all three main effects and we could not estimate any of the interactions. And now you should be able to see why. The interactions are all aliased with the main affects.

Let's look at another component L_{AB^2C} of the three factor interaction: $A \times B \times C$:

$$L_{AB^2C} = X_1 + 2X_2 + X_3 \pmod{3}$$

We can now fill out the table by first plugging in the levels of X_1 , X_2 and X_3 from the levels of A, B and C to generate the column L_{AB^2C} . When you assign treatments to the level of $L_{AB^2C} = 0$ you get an arrangement that follows (only the principle block filled in):

L_{AB^2C}		
0	1	2
0	0	0

Loading [MathJax]/extensions/MathZoom.js,

1, 1, 0
 2, 2, 0
 2, 0, 1
 0, 1, 1
 1, 2, 1
 1, 0, 2
 2, 1, 2
 0, 2, 2

Then it also generates its own Latin square using the same process that we used above. You should be able to follow how this Latin square was assigned to the nine treatment combinations from the table above.

		B			
		C	0	1	2
		0	0	1	2
A	1	2	0	1	
	2	1	2	0	

The benefit of doing this is to see that this one third fraction is also a Latin square. This is a Resolution III design, (it has a three letter word generator), and so it has the same properties that we saw at the two level designs, i.e. the main effects are clear of each other and estimable and aliased with higher order interactions including two-way. In fact, since the ABC and the AB²C are orthogonal to each other - they partition the A×B×C interaction - the two Latin squares we constructed are orthogonal Latin Squares.

The Next Level Example - Four Factors

Now let's take a look at the 3⁴⁻² design. How do we create this design? In this case we would have to pick 2 generators. We have four factors, A, B, C and D. So, let's say we will begin (trial and error) by selecting I = ABC = BCD as our generators then we will also have the generalized interactions between those generators which are also included. Thus we will also confound:

$$\begin{aligned} ABC \times BCD &= AB^2C^2D, \text{ and} \\ ABC \times (BCD)^2 &= AD^2 \end{aligned}$$

This is a Resolution II design - there are only two letters in the second component and we should be able to do better.

Let's try again, how about

Loading [MathJax]/extensions/MathZoom.js

$I = ABC = BC^2D$ as our generators. This confounds:

$$\begin{aligned}ABC \times BC^2D &= AB^2D \\ABC \times (BC^2D)^2 &= AC^2D^2\end{aligned}$$

This is much better because there is nothing smaller than a three letter word in the generator set so this is a Resolution III design. Now, how do we generate the design? It is a design with four factors but how many observations are there? Nine. It is still a design with only nine observations, or a 1/9th fraction of a 3^4 design or 81 observations. If we can write out the basic design with nine observations, which we can do with A and B, it gives us the basic design, and then we use our generators to give us C and D. We can use ABC such that:

$$L_{ABC} = 0 \text{ this principle fraction implies that } X_3 = 2X_1 + 2X_2 \pmod{3}.$$

$$L_{BC^2D} = 0 \text{ this implies that } X_4 = 2X_2 + X_3 \pmod{3}$$

If we were confounding this in blocks we will want a principal block where these two defining relationships are both zero. You will see that by defining X_3 and X_4 in this way results in ABC being equal to zero. Take a look and make sure that you understand how column C was generated by the function $X_3 = 2X_1 + 2X_2 \pmod{3}$ yet still preserves the principle implied where $L_{ABC} = 0$. Also, by the same process column D was generated using the function $X_4 = 2X_2 + X_3 \pmod{3}$ in such a way that it preserves the principle implied where in $L_{BC^2D} = 0$.

And so, the 3^{4-2} design is equivalent to the Graeco-Latin square. There are two Latin squares, one for each component, C and D, superimposed as shown below:

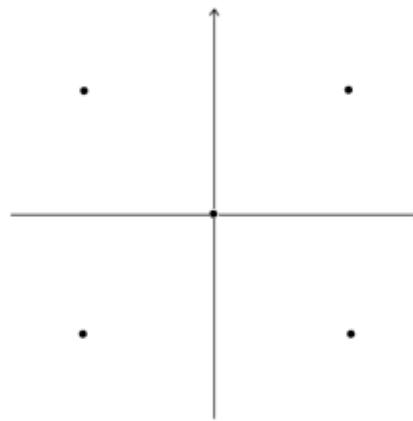
So we can see that the Graeco-Latin Square with three treatments is simply a fractional factorial of this 3^4 design!

9.3 - Mixed Factorials

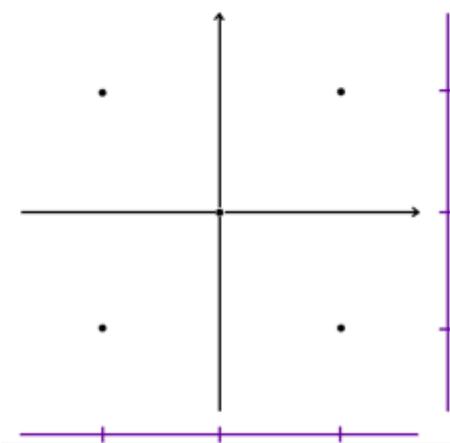
We have been talking about 2-level designs and 3-level designs. 2 level designs for screening factors and 3 level designs analogous to the 2 level designs, but the beginning of our discussion of response surface designs.

Since a 2 level design only has two levels of each factor, we can only detect linear effects. We have been mostly thinking about quantitative factors but especially when screening two level designs the factors can be presence/absence, or two types and you can still throw it into that framework and decide whether that's an important factor. If we go to three level designs we are almost always thinking about quantitative factors. But, again, it doesn't always have to be, it could be three types of something. However, in the general application we are talking about quantitative factors.

If we take a 2 level design that has center points.



Then, if you project into the A axis or the B axis, you have three distinct values, -1, 0, and +1.



In the main effect sense, a two level design with center points gives you three levels. This was our starting point towards moving to a three level design. Three-level designs require a whole lot more observations. With just two factors, i.e., $k = 2$, you have $3^k = 9$ observations, but as soon as we get to $k = 4$, now you already have $3^4 = 81$ observations, and with $k = 5$ becomes out of reach - $3^5 = 243$ observations. These designs grow very fast so obviously we are going to look for more efficient designs.

Mixed Level Designs

When we think of next level designs we think of factors with 4 or 5 levels, or designs with combinations of 2, 3, 4, or 5 levels of factors. In an Analysis of Variance course, which most of you have probably taken, it didn't distinguish between these factors. Instead, you looked at general machinery for factors with any numbers of level. What is new here is thinking about writing efficient designs. Let's say you have a $2^3 \times 3^2$ - this would be a mixed level design with $8 \times 9 = 72$ observations in a single replicate. So this is growing pretty rapidly! As this gets even bigger we could trim the size of this by looking at fractions for instance, 2^{3-1} , a fractional factorial of the first part. And, as these numbers of observations get larger you could look at crossing fractions of factorial designs.

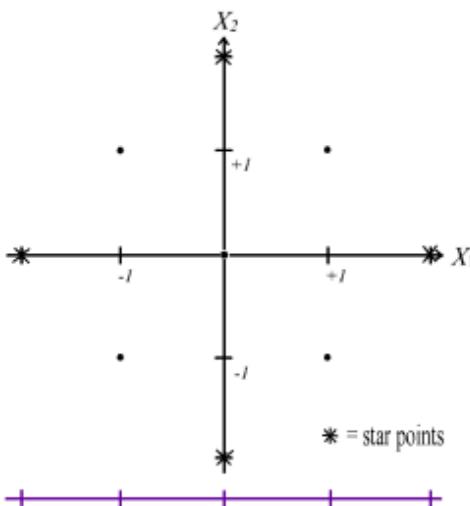
Loading [MathJax]/extensions/MathZoom.js

A Note about Factors with 4 levels - 2^2

This design is 2^2 , so in some sense there is nothing new here. By using the machinery of the 2^k designs you can always take a factor with four levels and call it the four combinations of 2^2 .

A Note about Factors with 5 levels

Design with factors with 5 levels... Think quantitative - if it is quantitative then you have five levels, and we should then be thinking about fitting a polynomial regression function.



This leads us to a whole new class of designs that we will look at next - Response Surface Designs.

What we have plotted here is a 2^2 design, which are the four corners of a 2^2 . We have center points. And then to achieve what we will refer to as a **central composite design** we will add what are called star points (axial points). These are points that are outside the range of -1 and 1 in each dimension. If you think in terms of projecting, we now have 5 levels of each of these 2 factors obtained in some automatic way. Instead of having 25 points which is what a 5×5 requires, we only have 9 points. It is a more efficient design but still in a projection we have five levels in each direction. What we want is enough points to estimate a response surface but at the same time keep the design as simple and with as few observations as possible.

The primary reason that we looked at the 3^k designs is to understand the confounding that occurs. When we have quantitative variables we will generally not use a 3 level designs. We use this more for understanding of what is going on. In some sense 3 level designs are not as practical as CCD designs. We will next consider response surface designs to address to goals of fitting a response surface model.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/53>

Links:

[1]

Loading [MathJax]/extensions/MathZoom.js

<https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson09/Figure-9-7.MPJ>

Loading [MathJax]/extensions/MathZoom.js



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 10: Simple Linear Regression

Lesson 10: Simple Linear Regression

This lesson corresponds to Chapter 10: Fitting Regression Models.

A course on fitting regression models is a prerequisite for this course. Chapter 10 covers standard topics in regression. Please read the Chapter if you feel you need to review and then proceed to Chapter 11.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/91>



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 11: Response Surface Methods and Designs

Lesson 11: Response Surface Methods and Designs

Overview of Response Surface Methods

We are now going to shift from screening designs where the primary focus of previous lessons was factor screening – (two-level factorials, fractional factorials being widely used), to trying to optimize an underlying process and look for the factor level combinations that give us the maximum yield and minimum costs. In many applications, this is our goal. However in some cases we are trying to hit a target or aim to match some given specifications - but this brings up other issues which we will get to later.

Here the objective of Response Surface Methods (RSM) is optimization, finding the best set of factor levels to achieve some goal. This lesson aims to cover the following goals:

Learning Objectives & Outcomes

- Response Surface Methodology and its sequential nature for optimizing a process
- First order and second order response surface models and how to find the direction of steepest ascent (or descent) to maximize (or minimize) the response
- How to deal with several responses simultaneously (Multiple Response Optimization)
- Central Composite Designs (CCD) and Box-Behnken Designs as two of the major Response Surface Designs and how to generate them using Minitab
- Design and Analysis of Mixture Designs for cases where the sum of the factor levels equals a constant, i.e. 100% or the totality of the components
- Introductory understanding of designs for computer models

RSM dates from the 1950's. Early applications were found in the chemical industry. We have already talked about Box. Box and Draper have some wonderful references about building RSMs and analyzing them which are very useful.

RSM as a Sequential Process

The text has a graphic depicting a response surface method in three dimensions, though actually it is four dimensional space that is being represented since the three factors are in 3-dimensional space the response is the 4th dimension.

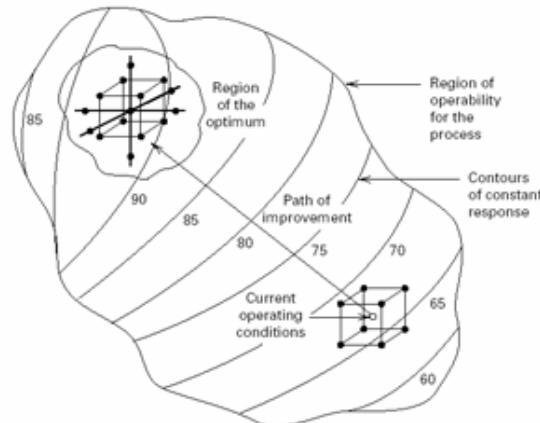
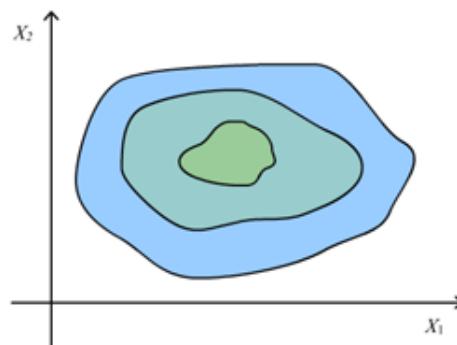


Figure 11-3 The sequential nature of RSM.

Instead, let's look at 2 dimensions - this is easier to think about and visualize. There is a response surface and we will imagine the ideal case where there is actually a 'hill' which has a nice centered peak. (If only reality were so nice, but it usually isn't!). Consider the geologic ridges that exist here in central Pennsylvania, the optimum or highest part of the 'hill' might be anywhere along this ridge. There's no clearly defined centered high point or peak that stands out. In this case there would be a whole range of values of X_1 and X_2 that would all describe the same 'peak' -- actually the points lying along the top of the ridge. This type of situation is quite realistic where there does not exist a predominate optimum.



But for our purposes let's think of this ideal 'hill' and the problem is that you don't know where this is and you want to find factor level values where the response is at its peak. This is your quest, to find the values X_1^{optimum} and X_2^{optimum} , where the response is at its peak. You might have a hunch that the optimum exists in certain location. This would be good area to start - some set of conditions, perhaps the way that the factory has always been doing things - and then perform an experiment at this starting point.

The actual variables in their natural units of measurement are used in the experiment. However, when we design our experiment we will use our coded variables, X_1 and X_2 which will be centered on 0, and extend +1 and -1 from the center of the region of experimentation. Therefore, we will take our natural units and then center and rescale them to the range from -1 to +1.

Our goal is to start somewhere using our best prior or current knowledge and search for the optimum spot where the response is either maximized or minimized.

Here are the models that we will use.

Screening Response Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \quad (1)$$

The screening model that we used for the first order situation involves linear effects and a single cross product factor, which represents the linear x linear interaction component.

Steepest Ascent Model

If we ignore cross products which gives an indication of the curvature of the response surface that we are fitting and just look at the first order model this is called the steepest ascent model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (2)$$

Optimization Model

Then, when we think that we are somewhere near the 'top of the hill' we will fit a second order model. This includes in addition the two second-order quadratic terms.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \quad (3)$$

Steepest Ascent - The First Order Model

Let's look at the first order situation - the method of steepest ascent. Now, remember, in the first place we don't know if the 'hill' even exists so we will start somewhere where we think the optimum exists. We start somewhere in terms of the natural units and use the coded units to do our experiment. Consider the example 11.1 in the textbook. We want to start in the region where x_1 = reaction time (30 - 40 seconds) and x_2 = temperature (150 - 160 degrees), and we want to look at the yield of the process as a function of these factors. In a sense, for the purpose of illustrating this concept, we can superimpose this region of experimentation on to our plot of our unknown 'hill'. We obviously conduct the experiment in its natural units but the designs will be specified in the coded units so we can apply them to any situation.

Specifically, here we use a design with four corner points, a 2^2 design and five center points. We now fit this first-order model and investigate it.

We put in the actual data for A and B and the response measurements Y.

We fit a full model first: See: [Ex-11-1-output.doc \[1\]](#)

We fit the surface. The model has two main effects, one cross product term and then one additional parameter as the mean for the center point. The residuals in this case have four *df* ***which come from replication of the center points***. Since there are five center points, i.e., four *df* among the five center points. This is a measure of pure error.

We start by testing for curvature. The question is whether the mean of the center points is different from the values at $(x_1, x_2) = (0,0)$ predicted from the screening response model (main effects plus interaction). We are testing whether the mean of the points at the center are on the plane fit by the four corner points. If the p-value had been small, this would have told you that a mean of the center points is above or below the plane indicating curvature in the response surface. The fact that, in this case, it is not significant indicates there is no curvature. Indeed the center points fall exactly on the plane that fits the quarter points.

There is just one degree of freedom for this test because the design only has one additional location in terms of the x's.

Next we check for significant effects of the factors. We see from the ANOVA that there is no interaction. So, let's refit this model without the interaction term, leaving just the A and B terms.

We still have the average of the center points and our AOV now shows 5 *df* for residual error. One of these is lack of fit of the additive model and there are 4 *df* of pure error as before. We have 1 *df* for curvature, and lack of fit in this case is just the interactions from the model.

What do we do with this? See the Minitab analysis and redo these results in [EX11-1.MPJ](#) [2]

Our estimated model is: $\hat{y} = 40.43 + 0.775x_1 + 0.325x_2$

So, for any X_1 and X_2 we can predict y . This fits a flat surface and it tells us that the predicted y is a function of X_1 and X_2 and the coefficients are the gradient of this function. We are working in coded variables at this time so these coefficients are unitless.

If we move 0.775 in the direction of X_1 and then 0.325 in the direction of X_2 this is the direction of steepest ascent. All we know is that this flat surface is one side of the 'hill'.

The method of steepest ascent tells us to do a first order experiment and find the direction that the 'hill' goes up and start marching up the hill taking additional measurements at each (x_1, x_2) until the response starts to decrease. If we start at 0, in coded units, then we can do a series of single experiments on this path up the 'hill' of the steepest descent. If we do this at a step size of $x_1 = 1$, then:

$$1 / 0.775 = x_2 / 0.325 \rightarrow x_2 = 0.325 / 0.775 = 0.42$$

and thus our step size of $x_1 = 1$ determines that $x_2 = 0.42$, in order to move in the direction determined to be the steepest ascent. If we take steps of 1 in coded units, this would be five minutes in terms of the time units. And for each step along that path we would go up 0.42 coded units in x_2 or approximately 2° on the temperature scale.

Table 11-3 Steepest Ascent Experiment for Example 11-1

Steps	Coded Variables		Natural Variables		Response <i>y</i>
	x_1	x_2	ξ_1	ξ_2	
Origin	0	0	35	155	
Δ	1.00	0.42	5	2	
Origin + Δ	1.00	0.42	40	157	41.0
Origin + 2Δ	2.00	0.84	45	159	42.9
Origin + 3Δ	3.00	1.26	50	161	47.1
Origin + 4Δ	4.00	1.68	55	163	49.7
Origin + 5Δ	5.00	2.10	60	165	53.8
Origin + 6Δ	6.00	2.52	65	167	59.9
Origin + 7Δ	7.00	2.94	70	169	65.0
Origin + 8Δ	8.00	3.36	75	171	70.4
Origin + 9Δ	9.00	3.78	80	173	77.6
Origin + 10Δ	10.00	4.20	85	175	80.3
Origin + 11Δ	11.00	4.62	90	179	76.2
Origin + 12Δ	12.00	5.04	95	181	75.1

Here is the series of steps in additional measures of five minutes and 2° temperature. The response is plotted and shows an increase that drops off towards the end.

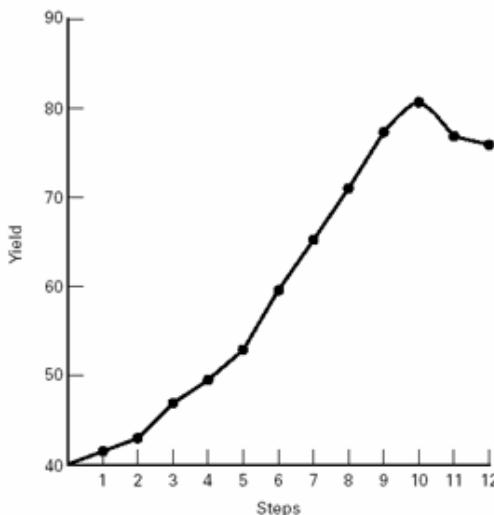


Figure 11-5 Yield versus steps along the path of steepest ascent for Example 11-1.

This is a pretty smooth curve and in reality you probably should go a little bit more beyond the peak to make sure you are at the peak. But all you are trying to do is to find out approximately where the top of the 'hill' is. If your first experiment is not exactly right you might have gone off in a wrong direction!

So you might want to do another first-order experiment just to be sure. Or, you might wish to do a second order experiment, assuming you are near the top. This is what we will discuss in the next section. The second order experiment will help find a more exact location of the peak.

The point is, this is a fairly cheap way to 'scout around the mountain' to try to find where the optimum conditions are. Remember, this example is being shown in two dimensions but you may be working in three or four dimensional space! You can use the same method, fitting a first-order model and then moving up the response surface in k dimensional space until you think you are close to where the optimal conditions are.

If you are in more than 2 dimensions, you will not be able to get a nice plot. But that is OK. The method of steepest ascent tells you where to take new measurements, and you will know the response at those points. You might move a few steps and you may see that the response continued to move up or perhaps not - then you might do another first order experiment and redirect your efforts. The point is, when we do the experiment for the second order model, we hope that the optimum will be in the range of the experiment - if it is not, we are extrapolating to find the optimum. In this case, the safest thing to do is to do another experiment around this estimated optimum. Since the experiment for the second order model requires more runs than experiments for the first order model, we want to move into the right region *before* we start fitting second order models.

Steepest Ascent - The Second Order Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

This second order model includes linear terms, cross product terms and a second order term for each of the x 's. If we generalize this to k x 's, we have k first order terms, k second order terms and then we have all possible pairwise first-order interactions. The linear terms just have one subscript. The quadratic terms have two subscripts. There are $k*(k-1)/2$ interaction terms. To fit

this model, we are going to need a response surface design that has more runs than the first order designs used to move close to the optimum.

This second order model is the basis for response surface designs under the assumption that although the hill is not a perfect quadratic polynomial in k dimensions, it provides a good approximation to the surface near the maximum or a minimum.

Assuming that we have 'marched up this hill' and if we re-specified the region of interest in our example, we are now between 80-90 in terms of time and 170-180 in terms of temperature. We would now translate these natural units into our coded units and if we fit the first order model again, hopefully we can detect that the middle is higher than the corner points so we would have curvature in our model, and could now fit a quadratic polynomial.

After using the Steepest Ascent method to find the optimum location in terms of our factors, we can now go directly to the second order response surface design. A favorite design that we consider is sometimes referred to as a central composite design. The central composite design is shown on Figure 11.3 above and in more detail in the text in Figure 11.10. The idea is simple - take the 2^k corner points, add a center point, and then create a star by drawing a line through the center point orthogonal to each face of the hypercube. Pick a radius along this line and place a new point at that radius. The effect is that each factor is now measured at 5 levels - center, 2 corners and the 2 star points. This gives us plenty of unique treatments to fit the 2nd order model with treatment degrees of freedom left over to test the goodness of fit. Replication is still usually done only at the center point.

11.1 - Multiple Responses

In many experiments more than one response is of interest for the experimenter. Furthermore, we sometimes want to find a solution for controllable factors which result in the best possible value for each response. This is the context of multiple response optimization, where we seek a compromise between the responses; however, it is not always possible to find a solution for controllable factors which optimize all of the responses simultaneously. Multiple response optimization has an extensive literature in the context of multiple objective optimization which is beyond the scope of this course. Here, we will discuss the basic steps in this area.

As expected, multiple response analysis starts with building a regression model for each response separately. For instance, in Example 11.2 we can fit three different regression models for each of the responses which are Yield, Viscosity and Molecular Weight based on two controllable factors: Time and Temperature.

One of the traditional methods way to analyze and find the desired operating condition one is **overlaid contour plots**. This method is mainly useful when we have two or maybe three controllable factors but in higher dimensions it loses its efficiency. This method simply consists of overlaying contour plot for each of the responses one over another in the controllable factors space and finding the area which makes the best possible value for each of the responses. Figure 11.16 (Montgomery, 7th Edition) shows the overlaid contour plots for example 11.2 in Time and Temperature space.

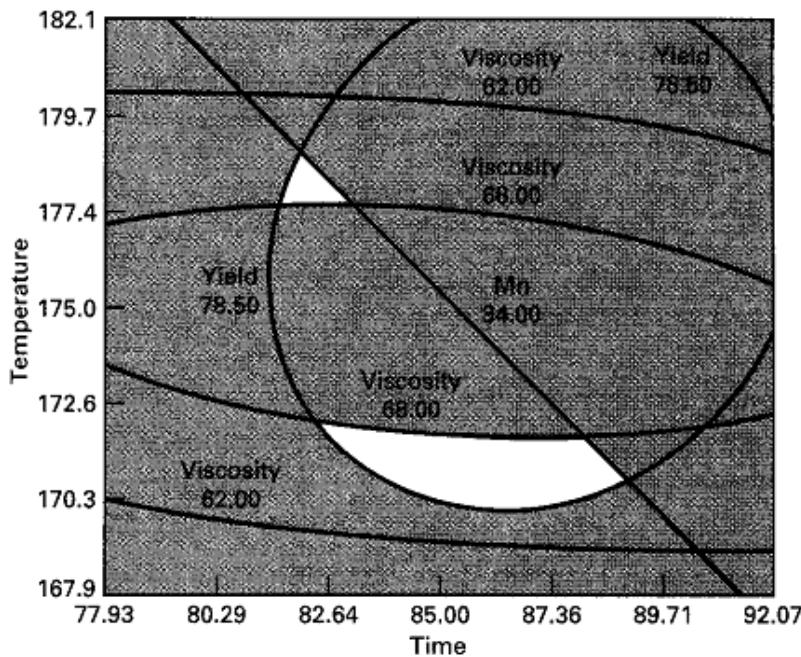


Figure 11.16 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)

The unshaded area is where *yield* > 78.5, 62 < *viscosity* < 68, and *molecular weight* < 3400. This area might be of special interest for the experimenter because they satisfy given conditions on the responses.

Another dominant approach for dealing with multiple response optimization is to form a **constrained optimization problem**. In this approach we treat one of the responses as the objective of a constrained optimization problem and other responses as the constraints where the constraint's boundary is to be determined by the decision maker (DM). The Design-Expert software package solves this approach using a direct search method.

Another important procedure that we will discuss here, also implemented in Minitab, is the **desirability function** approach. In this approach the value of each response for a given combination of controllable factors is first translated to a number between zero and one known as **individual desirability**. Individual desirability functions are different for different objective types which might be Maximization, Minimization or Target. If the objective type is maximum value, the desirability function is defined as

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^r & L \leq y \leq T \\ 1 & y > T \end{cases}$$

When the objective type is a minimum value the, the individual desirability is defines as

$$d = \begin{cases} 1 & y < T \\ \left(\frac{U-y}{U-T}\right)^r & T \leq y \leq U \\ 0 & y > U \end{cases}$$

Finally the two-sided desirability function with target-the-best objective type is defined as

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^{r_1} & L \leq y \leq T \\ \left(\frac{U-y}{U-T}\right)^{r_2} & T \leq y \leq U \\ 0 & y > U \end{cases}$$

Where the r_1 , r_2 and r define the shape of the individual desirability function (Figure 11.17 in the text shows the shape of individual desirability for different values of shape parameter). Individual desirability is then used to calculate the overall desirability using the following formula:

$$D = (d_1 d_2 \dots d_m)^{1/m}$$

where m is the number of responses. Now, the design variables should be chosen so that the overall desirability will be maximized. Minitab's Stat > DOE > Response Surface > Response Optimizer routine uses the desirability approach to optimize several responses, simultaneously.

11.2 - Response Surface Designs

After using the Steepest Ascent method to find the optimum location in terms of our factors, we can now go directly to the second order response surface design. A favorite design that we consider for a second order model is referred to as a central composite design.

We give here an example in two dimensions, Example 11.2 in the text. We have 2^k corner points and we have some number of center points which generally would be somewhere between 4 and 7, (five here). In two dimensions there are 4 star points, but in general there are $2k$ star points in k dimensions. The value of these points is something greater than 1. Why is it something greater than 1? If you think about the region of experimentation, we have up to now always defined a box, but if you think of a circle the star points are somewhere on the circumference of that circle, or in three dimensions on the ball enclosing the box. All of these are design points around the region where you expect the optimum outcome to be located. Typically the only replication, in order to get some measure of pure error, is done at the center of the design.

The data set for the Example 11.2 is found in the Minitab worksheet, [Ex11-2.MTW](#) [3]. The analysis using the Response Surface Design analysis module is shown in the [Ex11-2.MPJ](#) [4].

11.2.1 - Central Composite Designs

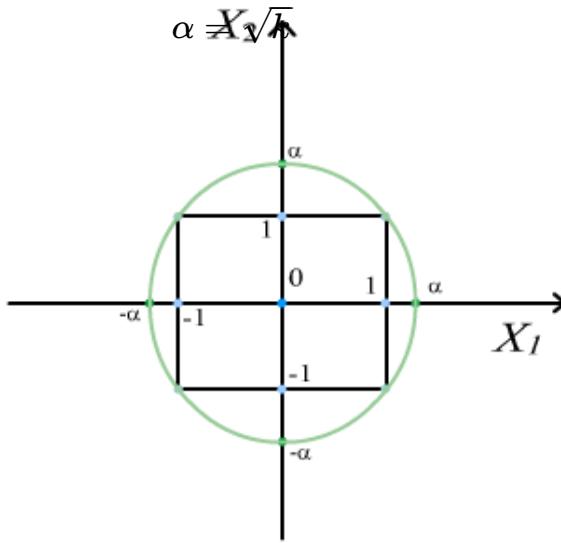
In the last section we looked at the example 11.2 which was in coded variables and was a central composite design.

In this section we examine a more general central composite design. For $k = 2$ we had a 2^2 design with center points, which was required for our first order model; then we added 2^*k star points. The star or axial points are, in general, at some value α and $-\alpha$ on each axis.

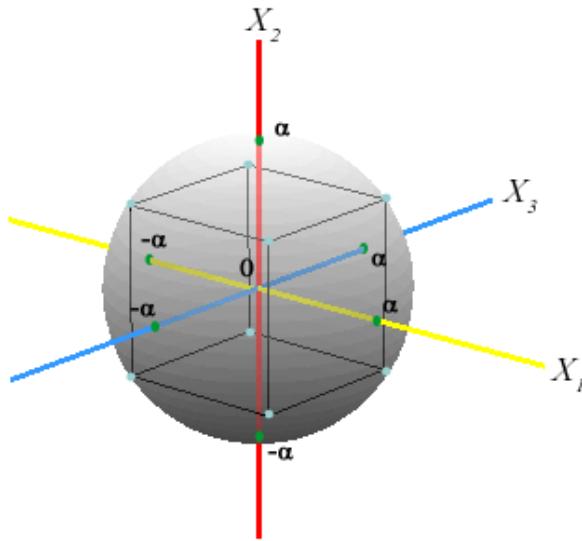
There are various choices of α . If $\alpha = 1$, the star points would be right on the boundary, and we would just have a 3^2 design. Thus $\alpha = 1$ is a special case, a case that we considered in the 3^k

designs. A more common choice of α is below.

which gives us a spherical design as shown below.



Our 2^2 design gives us the box, and adding the axial points (in green) outside of the box gives us a spherical design where $\alpha = \sqrt{k}$. The corner points and the axial points at α , are all points on the surface of a ball in three dimensions, as we see below.



This design in $k = 3$ dimensions can also be referred to as a central composite design, chosen so that the design is spherical. This is a common design. Much of this detail is given in Table 11.11 of the text.

An alternative choice where $\alpha = (n_F)^{\frac{1}{4}}$, or the fourth root of the number of points in the factorial part of the design, gives us a rotatable design.

If we have k factors, then we have, 2^k factorial points, $2*k$ axial points and n_c center points. Below is a table that summarizes these designs and compares them to 3^k designs:

--	--	--	--	--	--	--

		$k = 2$	$k = 3$	$k = 4$	$k = 5$
Central Composite Designs	Factorial points 2^k	4	8	16	32
	Star points 2^k	4	6	8	10
	Center points n_c (varies)	5	5	6	6
	Total	13	19	30	48
3^k Designs		9	27	81	243
Choice of α	Spherical design ($\alpha = \sqrt{k}$)	1.4	1.73	2	2.24
	Rotatable design ($\alpha = (n_F)^{\frac{1}{4}}$)	1.4	1.68	2	2.38

Compare the total number of observations required in the central composite designs versus the 3^k designs. As the number of factors increases you can see the efficiencies that are brought to bear.

The spherical designs are rotatable in the sense that the points are all equidistant from the center. Rotatable refers to the variance of the response function. A rotatable design exists when there is an equal prediction variance for all points a fixed distance from the center, 0. This is a nice property. If you pick the center of your design space and run your experiments, all points that are equal distance from the center in any direction, have equal variance of prediction.

You can see in the table above that the difference in the variation between the spherical and rotatable designs are slight, and don't seem to make much difference. But both ideas provide justification for selecting how far away the star points should be from the center.

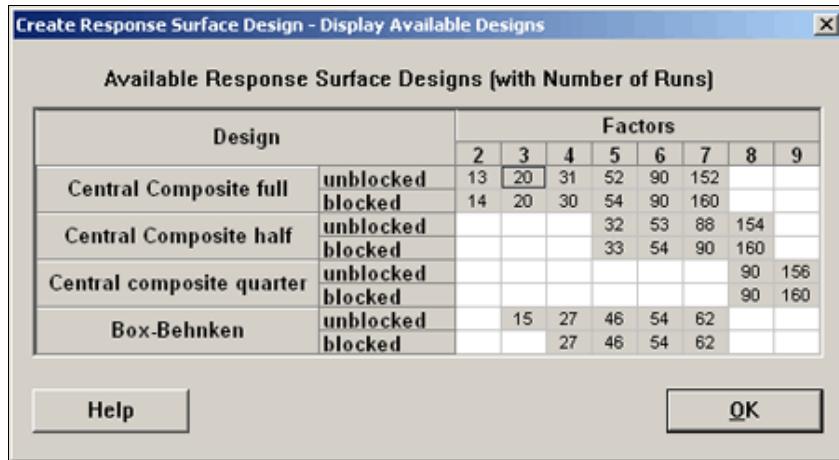
Why do we take about five or six center points in the design? The reason is also related to the variance of a predicted value. When fitting a response surface you want to estimate the response function in this design region where we are trying to find the optimum. We want the prediction to be reliable throughout the region, and especially near the center since we hope the optimum is in the central region. By picking five to six center points, the variance in the middle is approximately the same as the variance at the edge. If you only had one or two center points, then you would have less precision in the middle than you would have at the edge. As you go farther out beyond a distance of 1 in coded units, you get more variance and less precision. What we are trying to do is to balance the precision at the edge of the design relative to the middle.

How do you select the region where you want to run the experiment? Remember, for each factor X we said we need to choose the lower level is and the upper level for the region of experimentation. We usually picked the -1 and 1 as the boundary. If the lower natural unit is really the lowest number that you can test, because the experiment won't work lower than this, or the lower level is zero and you can't put in a negative amount of something, then, the star point is not possible because it is outside the range of experimentation.

If this is the case, one choice that you could make would be to use the $-\alpha$ as the lowest point. Generally, if you are not up against a boundary then this is not an issue and the star points are a way to reach beyond the region that you think the experiment should be run in. The issue isn't selecting the coding of the design relative to the natural units. You might lose some of these exact properties, but as long as you have the points nicely spread out in space you can fit a regression function. The penalty for not specifying the points exactly, would be seen in the variance, and it would be actually very slight.

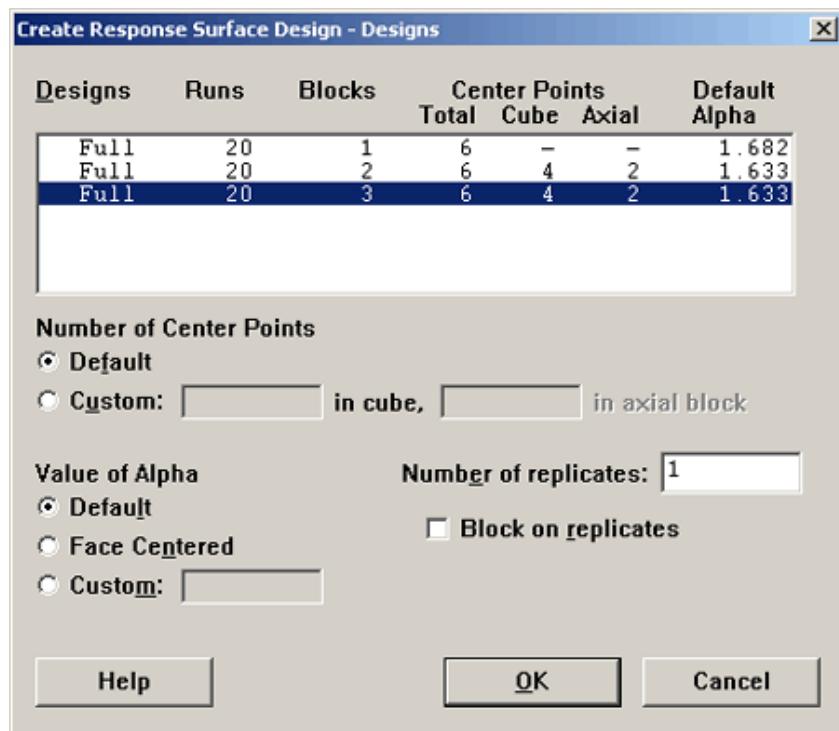
Generating These Designs in Minitab

Minitab will show you the available designs and how to generate these designs.



We can create central composite designs using a full factorial, central composite designs with fractional factorials, half fraction and a quarter fraction, and they can be arranged in blocks. Later, we will look at the Box-Behnken designs.

As an example, we look at the $k = 3$ design, set up in Minitab using a full factorial, completely randomized, in two blocks, or three blocks with six center points and the default $\alpha = 1.633$ (or $\alpha = 1.682$ for a rotatable design).



If you do not want the default α you can specify your own in the lower left. A Face Centered design is obtained by putting the α at +1 and -1 on the cube. Here is the design that results:

The first block is a one half fraction of the 2^3 plus 2 center points. Block 2 is the second half fraction of the factorial part with two center points. The third block consists of 6 star points, plus two center points. Each of the three blocks contains 2 center points and the first two blocks have half of the corner points each. The third block contains the star points and is of size 8.

Rollover the words 'Block 1', 'Block 2', and 'Block 3' in the graphic above. Do you see how they use center points strategically to tie the blocks together? They are represented in each block and they keep the design connected.

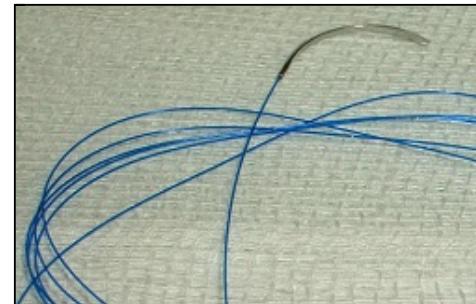
The corner points all have +1 or -1 for every dimension, because they're at the corners. They are either up or down, in or out, right or left. The axial points have $+\alpha$, or $-\alpha$ (+1.6330 or -1.6330) for A, but are 0 for factors B and C. The center points have zero on all three axes, truly the center of this region. We have designed this to cover the space in just the right way so that we can estimate a quadratic equation. Using a Central Composite Design, we can't estimate cubic terms, and we can't estimate higher order interactions. If we had utilized a 3^k design, one that quickly becomes unreasonably large, then we would have been able to estimate all of the higher order interactions.

However, we would have wasted a lot of resources to do it. The CCD allows us to estimate just linear and quadratic terms and first order interactions.

An Example - Polymers and Elasticity

This example is from the Box and Draper (1987) book and the data from Tables 9.2 and 9.4 are in Minitab ([BD9-1.MTW](#) [5]).

This example has three variables and they are creating a polymer, a kind of plastic that has a quality of elasticity. The measurement in this experiment is the level of elasticity. We created the design in Minitab for this experiment, however the data only has two center points:



	C1	C2	C3	C4	C5	C6	C7	C8
	StdOrder	RunOrder	PtType	Blocks	A	B	C	Y-elasticity
1	1	1	1	1	-1	-1	-1	25.74
2	2	2	1	1	1	-1	-1	48.98
3	3	3	1	1	-1	1	-1	42.78
4	4	4	1	1	1	1	-1	35.94
5	5	5	1	1	-1	-1	1	41.50
6	6	6	1	1	1	-1	1	50.10
7	7	7	1	1	-1	1	1	46.06
8	8	8	1	1	1	1	1	27.70
9	9	9	-1	1	-2	0	0	35.50
10	10	10	-1	1	2	0	0	44.18
11	11	11	-1	1	0	-2	0	38.58
12	12	12	-1	1	0	2	0	28.46
13	13	13	-1	1	0	0	-2	33.50
14	14	14	-1	1	0	0	2	42.02
15	15	15	0	1	0	0	0	57.52
16	16	16	0	1	0	0	0	59.68
17	17	17	0	1	0	0	0	*
18	18	18	0	1	0	0	0	*
19	19	19	0	1	0	0	0	*
20	20	20	0	1	0	0	0	*

Variables A and B are the concentration of two ingredients that make up the polymer, and C is the temperature, and the response is elasticity. There are 8 corner points, a complete factorial, 6 star points and 2 center points.

Let's go right to the analysis stage now using Minitab ...

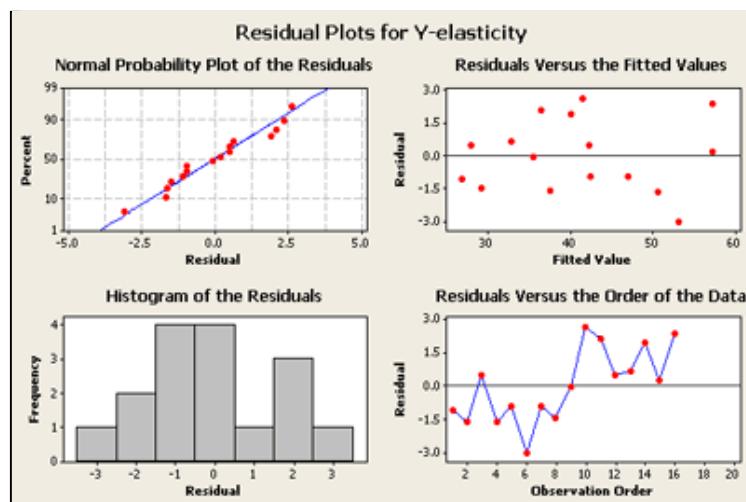
A video demonstration is given here: <https://screencast.com/t/yfyiPXdPq> [6]

or you can view the flash animation...



[7]

Before we move on I would like to go back and take a look again at the plot of the residuals. Wait a minute! Is there something wrong with this residual plot?



Residual plot for the polynomial fit

Look at the plot in the lower right. The first eight points tend to be low, and then the next eight points are at a higher level. This is a clue, that something is influencing the response that is not being fit by the model. This looks suspicious. What happened? My guess is that the experiment was run in two phases. They first ran the 2^k part - (block 1). And then they noticed the response and added the star points to make a responsive surface design in the second part. This is often how these experiments are conducted. You first perform a first-order experiment, and then you add center points and star points and then fit the quadratic.

Add a block term and rerun the experiment to see if this makes a difference.

Two Types of Central Composite Designs

The central composite design has 2^k star points on the axial lines outside of the box defined by the corner points. There are two major types of central composite designs: the spherical central composite design where the star points are the same distance from the center as the corner points, and the rotatable central composite design where the star points are shifted or placed such that the variances of the predicted values of the responses are all equal, for x 's which are an equal distance from the center.

When you are choosing, in the natural units, the values corresponding to the low and high, i.e. corresponding to -1 and 1 in coded units, keep in mind that the design will have to include points further from the center in all directions. You are trying to fit the design in the middle of your region of interest, the region where you expect the experiment to give the optimal response.

11.2.2 - Box-Behnken Designs

Box-Behnken Designs

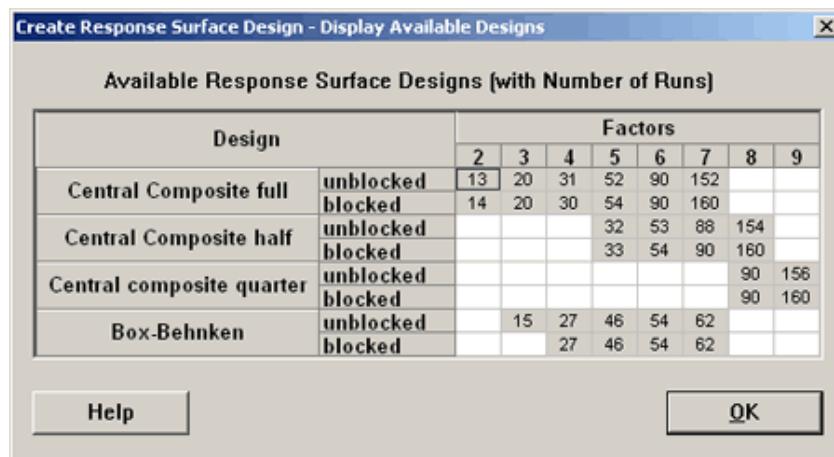
Another class of response surface designs are called Box-Behnken designs. They are very useful in the same setting as the central composite designs. Their primary advantage is in addressing the issue of where the experimental boundaries should be, and in particular to avoid treatment combinations that are extreme. By extreme, we are thinking of the corner points and the star points, which are extreme points in terms of region in which we are doing our experiment. The Box-Behnken design avoids all the corner points, and the star points.

One way to think about this is that in the central composite design we have a ball where all of the corner points lie on the surface of the ball. In the Box-Behnken design the ball is now located inside the box defined by a 'wire frame' that is composed of the edges of the box. If you blew up a balloon inside this wire frame box so that it just barely extends beyond the sides of the box, it might look like this, in three dimensions. Notice where the balloon first touches the wire frame; this is where the points are selected to create the design.

Box-Behnken Design														
Factors: 3			Replicates: 1											
Base runs: 15			Total runs: 15											
Base blocks: 1			Total blocks: 1											
Center points: 3														
Design Table														
Run	Blk	A	B	C										
1	1	-	-	0										
2	1	+	-	0										
3	1	-	+	0										
4	1	+	+	0										
5	1	-	0	-										
6	1	+	0	-										
7	1	-	0	+										
8	1	+	0	+										
9	1	0	-	-										
10	1	0	+	-										
11	1	0	-	+										
12	1	0	+	+										
13	1	0	0	0										
14	1	0	0	0										
15	1	0	0	0										

Therefore the points are still on the surface of a ball, but the points are never further out than the low and high in any direction. In addition, there would be multiple center points as before. In this type of design you do not need as many center points because points on the outside are closer to the middle. The number of center points are again chosen so that the variance of is about the same in the middle of the design as it is on the outside of the design.

In Minitab we can see the different designs that are available. Listed at the bottom are the Box-Behnken Designs.



A Box-Behnken (BB) design with two factors does not exist. With three factors the BB design by default will have three center points and is given in the Minitab output shown above. The last three observations are the center points. The other points, you will notice, all include one 0 for one of the factors and then a plus or minus combination for the other two factors.

If you consider the BB design with four factors, you get the same pattern where we have two of the factors at + or - 1 and the other two factors are 0. Again, this design has three center points, and a total of 27 observations.

Comparing the central composite design with 4 factors, which has 31 observations, a Box-Behnken design only includes 27 observations. For 5 factors, the Box-Behnken would have 46 observations, and a central composite would have 52 observations if you used a complete factorial, but this is where the central composite also allows you to use a fractional factorial as a means of making this experiment more efficient. Likewise for six factors, the Box-Behnken requires 54 observations, and this is the minimum of the central composite design.

Both the CCD and the BB design can work, but they have different structures, so if your experimental region is such that extreme points are a problem then there are some advantages to the Box-Behnken. Otherwise, they both work well.

The central composite design is one that I favor because even though you are interested in the middle of a region, if you put all your points in the middle you do not have as much leverage about where the model fits. So when you can move your points out you get better information about the function within your region of experimentation.

However, by moving your points too far out, you get into boundaries or could get into extreme conditions, and then enter the practical issues which might outweigh the statistical issues. The central composite design is used more often but the Box-Behnken is a good design in the sense that you can fit the quadratic model. It would be interesting to look at the variance of the predicted values for both of these designs. (This would be an interesting research question for somebody!) The question would be which of the two designs gives you smaller average variance over the region of experimentation.

The usual justification for going to the Box-Behnken is to avoid the situation where the corner points in the central composite design are very extreme, i.e. they are at the highest level of several factors. So, because they are very extreme, the researchers may say these points are not very typical. In this case the Box Behnken may look a lot more desirable since there are more points in the middle of the range and they are not as extreme. The Box-Behnken might feel a little 'safer' since the points are not as extreme as all of the factors.

The Variance of the Predicted Values

Let's look at this a little bit. We can write out the model:

$$\hat{y}_x = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2$$

Where the b_0 , b_1 , etc are the estimated parameters. This is a quadratic model with two x's. The question we want to answer is how many center points should there be so that the variance of the predicted value, $\text{var}(\hat{y}_x)$ when x is at the center is the same as when x is at the outside of the region?

See handout [Chapter 11: Supplemental Text Material](#) [8]. This shows the impact on the variance of a predicted value in the situation with $k = 2$, full factorial and the design has only 2 centerpoints rather than the 5 or 6 that the central composite design would recommend.

What you see (S11-3) is that in the middle of the region the variance is much higher than further out. So, by putting more points in the center of the design, collecting more information there, (replicating a design in the middle), you see that the standard error is lower in the middle and

roughly the same as farther out. It gets larger again in the corners and continues growing as you go out from the center. By putting in enough center points you balance the variance in the middle of the region relative to further out.

Another example (S11-4) is a central composite design where the star points are on the face. It is not rotatable design and the variance changes depending on which direction you're moving out from center of the design.

It also shows another example (S11-4), also a face-centered design with zero center points, which shows a slight hump in the middle on the variance function.

Notice that we only need two center points for the face centered design. Rather than having our star points farther out, if we move them closer into the face we do not need as many center points because we already have points closer to the center. A lot of factors affect the efficiencies of these designs.

Rotatability

Rotatability is determined by our choice of alpha. A design is *rotatable* if the prediction variance depends only on the distance of the design point from the center of the design. This is what we were observing previously. Here in the supplemental material (S11-5) is an example with a rotatable design, but the variance contours are based on a reduced model. It only has one quadratic term rather than two. As a result we get a slightly different shape, the point being that rotatability and equal variance contours depend both on the design and on the model that we are fitting. We are usually thinking about the full quadratic model when we make that claim.

11.3 - Mixture Experiments

This is another class of response surface designs where the components are not just levels of factors but a special set where the x_1, x_2, \dots are coded and are the components of the mixture such that the sum of the $x_i = 1$. So, these make up the proportions of the mixture.

Examples

If you are making any kind of product it usually involves mixtures of ingredients. A classic example is gasoline which is a mixture of various petrochemicals. In polymer production, polymers are actually mixtures of components as well. My favorite classroom example is baking a cake. A cake is a mixture of flour, sugar, eggs, and other ingredients depending on the type of cake. It is a mixture where the levels of x are the proportions of the ingredients.

This constraint that the sum of the x 's sum to 1, i.e.,

$$0 \leq x_i \leq 1$$

has an impact on how we analyze these experiments.

Here we will write out our usual linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

where,

If you want to incorporate this constraint then we can write:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

in other words, if we drop the β_0 , this reduces the parameter space by 1 and then we can fit a reduced model even though the x 's are each constrained.

In the quadratic model:

$$Y_h = \sum_{i=1}^k \beta_i x_{hi} + \sum_{i < j} \beta_{ij} x_{hi} x_{hj} + \varepsilon_h$$

This is probably the model we are most interested in and will use the most. Then we can generalize this into a cubic model which has one additional term.

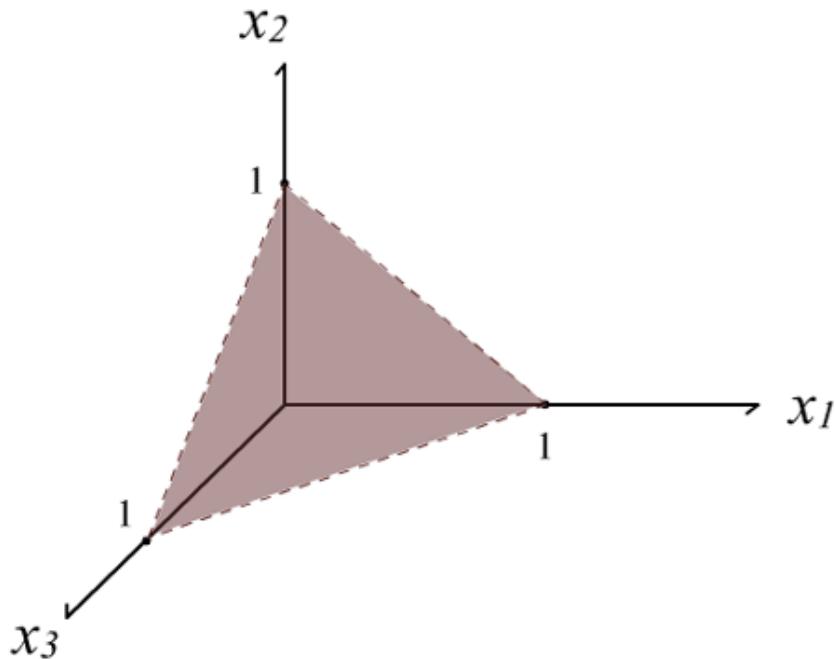
A Cubic Model

$$Y_h = \sum_{i=1}^k \beta_i x_{hi} + \sum_{i < j} \beta_{ij} x_{hi} x_{hj} + \sum_{i < j < l} \sum_{h=1}^k \beta_{ijl} x_{hi} x_{hj} x_{hl} + \varepsilon_h$$

These models are used to fit response surfaces.

Let's look at the parameter of space. Let's say that $k = 2$. The mixture is entirely made up of two ingredients, x_1 and x_2 . The sum of both ingredients is a line plotted in the parameter space below: An experiment made up of two components is either all of x_1 or all of x_2 or something in between, a proportion of the two. Use your mouse to *click and drag the intersection point* along the line that serves as a boundary to this region of experimentation.

Let's take a Look at the parameter space in three dimensions. Here we have three components: x_1 , x_2 and x_3 . As we satisfy our constraint that the sum of all the components equal 1 and then our parameter space is the plane that cuts the three-dimensional surface, intersecting these three points in the graph below scratch that in the plot below.



Plot showing the plane where the sum: $x_1+x_2+x_3 = 1$

The triangle represents the full extent of the region of experimentation in this case with the points sometimes referred to as the Barycentric coordinates. The design question we want to address is where do we do our experiment? We are not interested in any one of the corners of the triangle where only one ingredient is represented, we are interested in some way on the middle where there is a proportion of all three of the ingredients included. We will restrict it to a feasible region of experimentation somewhere in the middle area.

Let's look at an example, for instance, producing cattle feed. The ingredients might include the following: corn, oats, hay, soybean, grass, ... all sorts of things.

In some situations it might work where you might have 100% of one component, but many instances of mixtures we try to partition off a part of the space in the middle where we think the combination is optimal.

In $k = 4$ the region of experimentation can be represented by the shape of a tetrahedron where each of the four sides of the tetrahedron is an equilateral triangle and has its own set of Barycentric coordinates. Each face of the tetrahedron corresponds to design region where one coordinate is zero, and the remaining three must sum to 1.

11.3.1 - Two Major Types of Mixture Designs

Simplex Lattice Design

A $\{p,m\}$ simplex lattice design for p factors (components) is defined as all possible combination of factor levels defined as

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad i = 1, 2, \dots, p$$

As an example, the simplex lattice design factor levels for the case of $\{3,2\}$ will be

Which results in the following design points:

$$(x_1, x_2, x_3) = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, 0, \frac{1}{2}), (0, \frac{1}{2}, \frac{1}{2})\}$$

$$x_i = 0, \frac{1}{3}, 1 \quad i = 1, 2, 3$$

Simplex Centroid Design

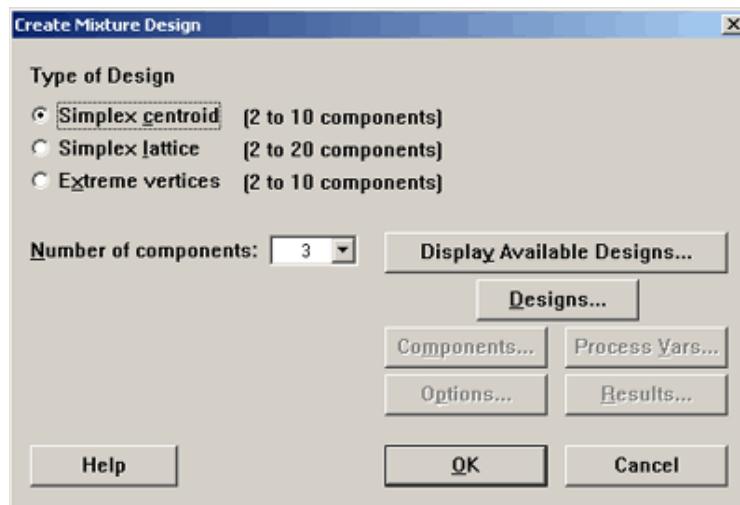
This design which has $2^p - 1$ design points consist of p permutations of $(1, 0, 0, \dots, 0)$, permutations of $(1, 0, 0, \dots, 0)$, $\binom{p}{2}$, permutations of $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$, $\binom{p}{3}$, and the overall centroid $(\frac{1}{p}, \frac{1}{p}, \dots, \frac{1}{p})$. Some simplex centroid designs for the case of $p = 3$ and $p = 4$ can be find in Figure 11.41.

Minitab handles mixture experiments which can be accessed through Stat > DOE > Mixture. It allows for building and analysis of Simplex Lattice and Simplex Centroid designs. Furthermore, it covers a third design which is named, Extreme Vertex Design. Application of Extreme Vertex designs are for cases where we have upper and lower constraints on some or all of the components making the design space smaller than the original region.

11.3.2 - Mixture Designs in Minitab

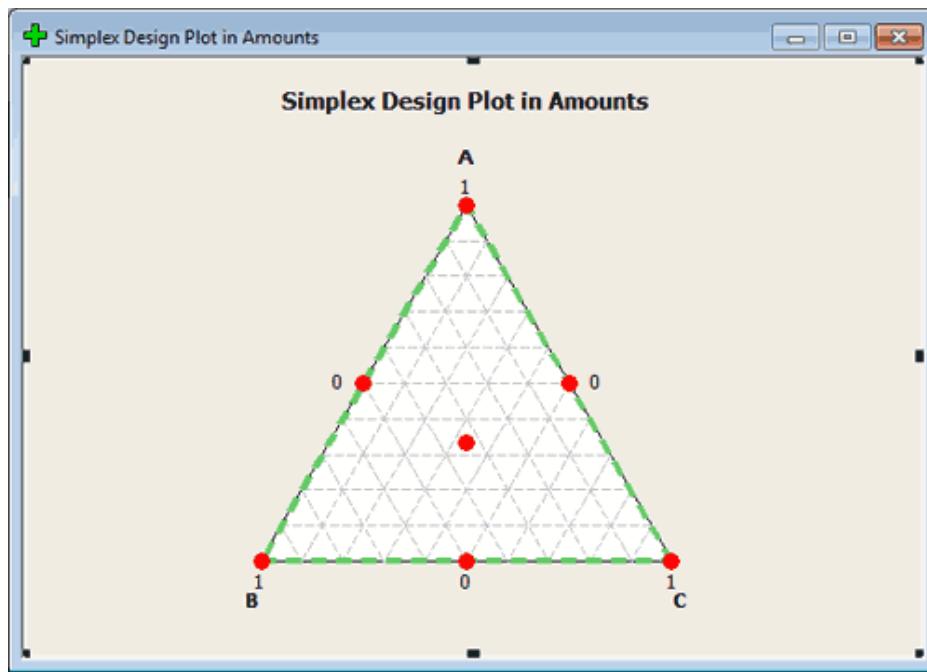
How does Minitab handle these types of experiments?

Mixture designs are a special case of response surface designs. Under the stat menu in many tab, select design of experiments, then mixture, create mixture design. Minitab then presents you with the following dialog box:



Simplex lattice option will look at the points that are extremes. Simplex lattice creates a design for p components of degree m . In this case, we want points that are made up of $0, \frac{1}{m}, \frac{2}{m}, \dots$ up to 1 . Classifying the points in this way tells us how we will space the points. For instance, if $m = 2$, then the only points we would have would be $0, 1/2$, and 1 to play with in all key dimensions. You can

create this design in Minitab, for 3 factors, using tab Stat > DOE > Mixture > Create Mixture Design and select Simplex Centroid. See the image here:

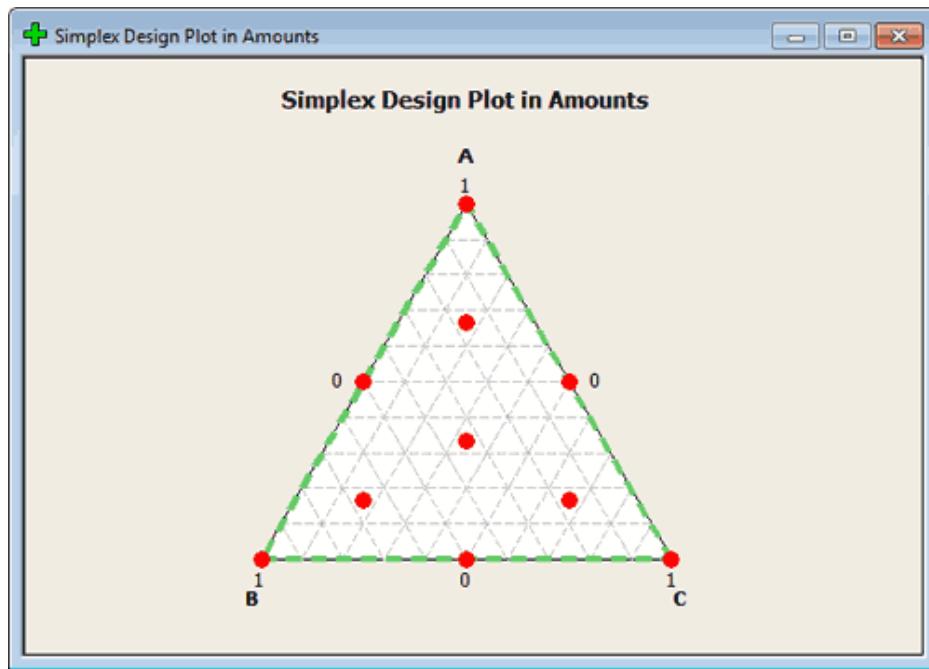


If we are in a design with the $m = 3$, then we would have 0, 1/3, 2/3, and 1. In this case we would have points a third of the way along each dimension. Any point on the boundary can be constructed in this way.

All of these points are on the boundary which means that they are made up of mixtures that omit one of the components. (This is not always desirable but in some settings it is fine.)

The centroid is the point in the middle. Axial points are points that lie along the lines that intersect the region of experimentation, points that are located interior and therefore include part of all of the components.

You can create this design in Minitab, for 3 factors, using tab Stat > DOE > Mixture > Create Mixture Design and select Simplex Centroid. See the image here:



This should give you the range of points that you think of when designing in a mixture. again, you want points in the middle but like regression in an unconstrained space you typically want to have your points farther out so you have good leverage. From this perspective, the points on the outside make a lot of sense. From an actual experimentation situation, you would have to be in a scientific setting also where those points make sense. If not, we would constrain this region to begin with. We will get in to this later.

How Rich of a Design?

Let's look at the set of possible designs that Minitab gives us.

Design	Components												
	2	3	4	5	6	7	8	9	10	11	12	13	
Centroid	3	7	15	31	63	127	255	511	1023				
Lattice 1	2	3	4	5	6	7	8	9	10	11	12	13	
Lattice 2	3	6	10	15	21	28	36	45	55	66	78	91	
Lattice 3	4	10	20	35	56	84	120	165	220	286	364	455	
Lattice 4	5	15	35	70	126	210	330	495	715	1001			
Lattice 5	6	21	56	126	252	462	792						
Lattice 6	7	28	84	210	462	924							
Lattice 7	8	36	120	330	792								

Where it is labeled on the left Lattice 1, Lattice 2, etc., here minitab is referring to degree 1, 2, etc. So, if you want a lattice of degree 1, this is not very interesting. This means that you just have a 0 and 1. If you go to a lattice of degree 2 then you need six points in three dimensions. This is pretty much what we looked at previously... (roll over the red mixture points, below).

Here is a design table for a lattice with degree 3:

Now let's go into Minitab and augment this design by including axial points. Here is what results:

This gives us three more points. Each of these points is $2/3$, $1/6$, $1/6$. These are interior points.

These are good designs if you can run your experiment in the whole region.

Let's take a look at four dimensions and see what the program will do here. Here is a design with four components, four dimensions, and a lattice of degree three. We have also selected to augment this design with axial and center points.

This gives us 25 points in the design and the plot shows us the four faces of the tetrahedron. It doesn't look like it is showing us a plot of the interior points.

11.3.3 - The Analysis of Mixture Designs

Example - Elongation of Yarn

Ex11.5.MTW [9] from the text.

This example has to do with the elongation of yarn based on its component fabrics. There are three components in this mixture and each component is a synthetic material. The mixture design was one that we had looked at previously. It is a simple lattice design of degree 2. This means that it has mixtures of 0, $1/2$, 100%. The components of this design are made up of these three possibilities.



	StdOrder	RunOrder	PtType	Blocks	A	B	C	Y	
1	1	1	1	1	1.0	0.0	0.0	11.0	
2	2	2	2	1	0.5	0.5	0.0	15.0	
3	3	3	2	1	0.5	0.0	0.5	17.7	
4	4	4	1	1	0.0	1.0	0.0	8.8	
5	5	5	2	1	0.0	0.5	0.5	10.0	
6	6	6	1	1	0.0	0.0	1.0	16.8	
7	7	7	1	1	1.0	0.0	0.0	12.4	
8	8	8	2	1	0.5	0.5	0.0	14.8	
9	9	9	2	1	0.5	0.0	0.5	16.4	
10	10	10	1	1	0.0	1.0	0.0	10.0	
11	11	11	2	1	0.0	0.5	0.5	9.7	
12	12	12	1	1	0.0	0.0	1.0	16.0	
13	13	13	1	1	1.0	0.0	0.0	*	
14	14	14	2	1	0.5	0.5	0.0	16.1	
15	15	15	2	1	0.5	0.0	0.5	16.6	
16	16	16	1	1	0.0	1.0	0.0	*	
17	17	17	2	1	0.0	0.5	0.5	11.8	
18	18	18	1	1	0.0	0.0	1.0	*	

In the Minitab program, the first 6 runs show you the pure components, and in addition you have the 5 mixed components. All of this was replicated 3 times so that we have 15 runs. There were three that had missing data.

You can also specify in more detail which type of points that you want to include in the mixture design using the dialog boxes in Minitab if your experiment requires this.

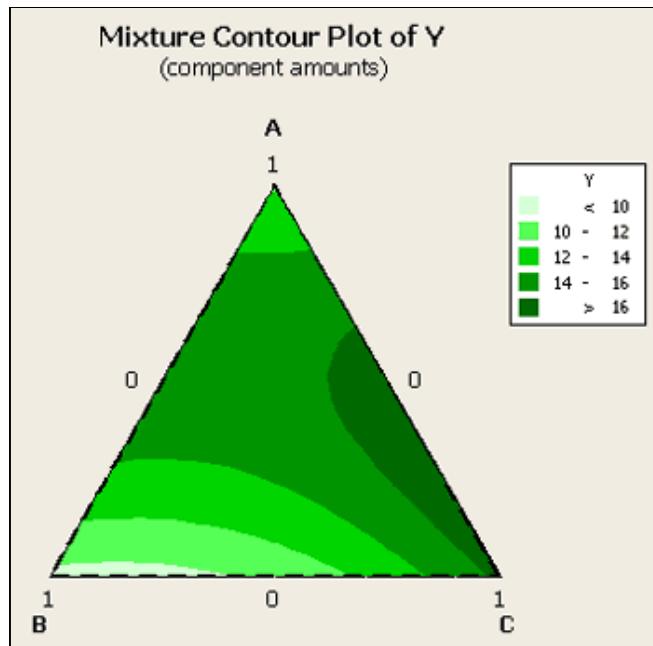
Analysis

In the analysis we fit the quadratic model (the linear + the interaction terms). Remember we only have 6 points in this design, the vertex, the half-lengths, so we are fitting a response surface to these 6 points. Let's take a look at the analysis:

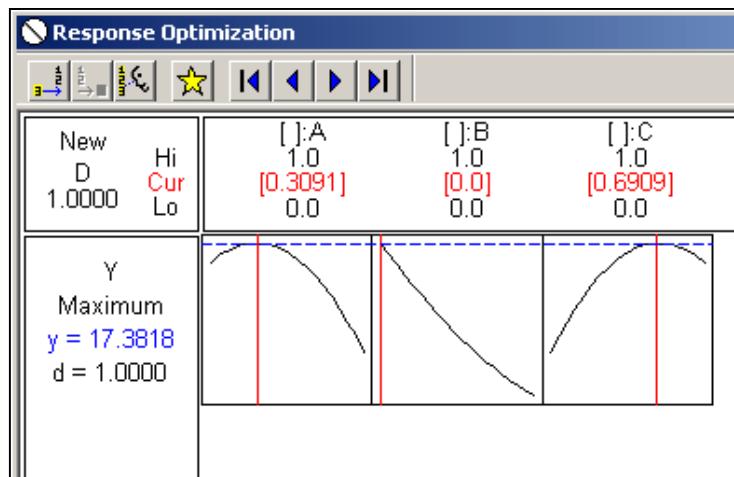
Regression for Mixtures: Y versus A, B, C						
Estimated Regression Coefficients for Y (component proportions)						
Term	Coef	SE Coef	T	P	VIF	
A	11.700	0.6037	*	*	1.750	
B	9.400	0.6037	*	*	1.750	
C	16.400	0.6037	*	*	1.750	
A*B	19.000	2.6082	7.28	0.000	1.750	
A*C	11.400	2.6082	4.37	0.002	1.750	
B*C	-9.600	2.6082	-3.68	0.005	1.750	
 S = 0.853750 PRESS = 18.295						
R-Sq = 95.14% R-Sq(pred) = 86.43% R-Sq(adj) = 92.43%						
 Analysis of Variance for Y (component proportions)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	5	128.296	128.2960	25.6592	35.20	0.000
Linear	2	57.629	50.9200	25.4600	34.93	0.000
Quadratic	3	70.667	70.6669	23.5556	32.32	0.000
Residual Error	9	6.560	6.5600	0.7289		
Total	14	134.856				

Here we get 2 *df* linear, 3 *df* quadratic, these are the five regression parameters. If you look at the individual coefficients, six of them because there is no intercept, three linear and three cross product terms... The 9 *df* for error are from the triple replicates and the double replicates. This is pure error and there is no additional *df* for lack of fit in this full model.

If we look at the contour service plot we get:



We have the optimum somewhere between a mixture of A and C, with B essentially not contributing very much at all. So, roughly 2/3rds C and 1/3 A is what we would like in our mixture. Let's look at the optimizer to find the optimum values.



It looks like A = about .3 and B = about .7, with B not contributing nothing to the mixture.

Unless I see the plot how can I use the analysis output? How else can I determine the appropriate levels?

Example - Gasoline Production

Pr11-31.MTW [10] from text

This example focuses on the production of an efficient gasoline mixture. The response variable is miles per gallon (mpg) as a function of the 3 components in the mixture. The data set contains these 14 points - which has



duplicates at the centroid, labeled (1/3, 1/3, 1/3), and the three vertices, labeled (1,0,0), (0,1,0), and (0,0,1).



+	C1	C2	C3	C4	C5	C6	C7	C8
	StdOrder	RunOrder	PtType	Blocks	A	B	C	Y-mpg
1	1	1	1	1	1.00000	0.00000	0.00000	24.5
2	2	2	2	1	0.50000	0.50000	0.00000	25.1
3	3	3	2	1	0.50000	0.00000	0.50000	24.3
4	4	4	1	1	0.00000	1.00000	0.00000	24.8
5	5	5	2	1	0.00000	0.50000	0.50000	23.5
6	6	6	1	1	0.00000	0.00000	1.00000	22.7
7	7	7	0	1	0.33333	0.33333	0.33333	24.8
8	8	8	-1	1	0.66667	0.16667	0.16667	24.2
9	9	9	-1	1	0.16667	0.66667	0.16667	23.9
10	10	10	-1	1	0.16667	0.16667	0.66667	23.7
11	11	11	1	1	1.00000	0.00000	0.00000	25.1
12	12	12	1	1	0.00000	1.00000	0.00000	23.9
13	13	13	1	1	0.00000	0.00000	1.00000	23.6
14	14	14	0	1	0.33333	0.33333	0.33333	24.1

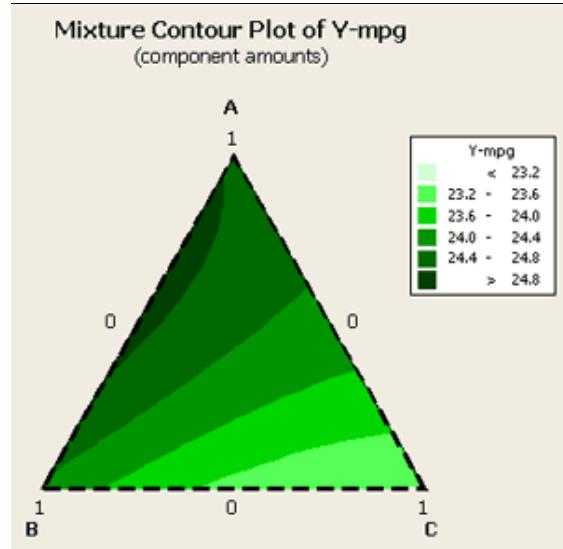
This is a degree 2 design that has points at the vertices, middle of the edges, the center and axial points, which are interior points, (2/3, 1/6, 1/6), (1/6, 2/3, 1/6) and (1/6, 1/6, 2/3). Also the design includes replication at the vertices and the centroid.

If you analyze this dataset without having first generated the design in Minitab, you need to tell Minitab some things about the data since you're importing it.

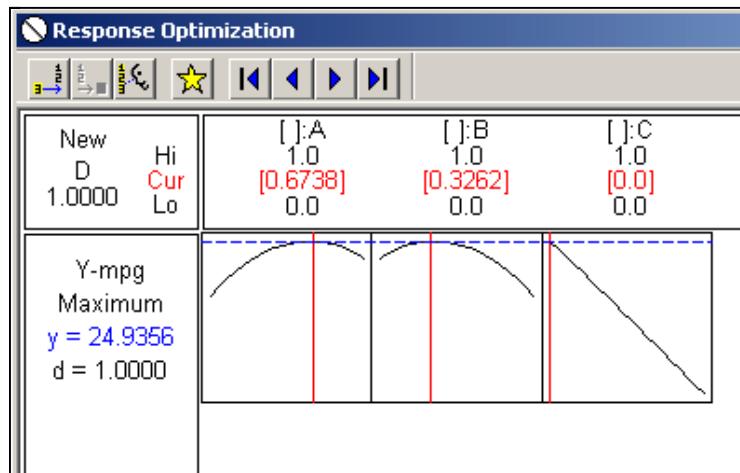
Analysis of Variance for Y-mpg (component proportions)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	5	4.2224	4.2224	0.84449	3.90	0.043
Linear	2	3.9247	2.7487	1.37433	6.35	0.022
Quadratic	3	0.2978	0.2978	0.09925	0.46	0.719
Residual Error	8	1.7319	1.7319	0.21648		
Lack-of-Fit	4	0.4969	0.4969	0.12421	0.40	0.800
Pure Error	4	1.2350	1.2350	0.30875		
Total	13	5.9543				

The model shows a linear term significant, the quadratic terms not significant, and the lack of fit, (a total of 10 points and we are fitting a model sex parameters - 4 df), it shows that there is no lack of fit from the model. It is not likely that it would make any difference.

If we look at the contour plot for this data:

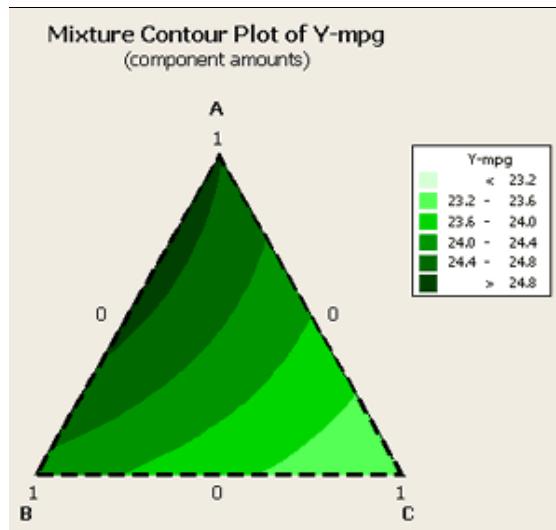


We can see that the optimum looks to be about 1/3, 2/3 between components A and B. Component C does not play hardly any role at all. Next, let's look at the optimizer for this data where we want to maximize a target of about 24.9.



And, again, we can see that component A at the optimal level is about 2/3rds and component B is at about 1/3rd. Component C plays no part, as a matter of fact if we were to add it to the gasoline mixture it would probably lower our miles per gallon average.

Let's go back to the model and take out the factors related to component C and see what happens. When this occurs we get the following contour plot...



... and the following analysis:

Analysis of Variance for Y-mpg (component proportions)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	3	4.0812	4.0812	1.3604	7.26	0.007
Linear	2	3.9247	3.1548	1.5774	8.42	0.007
Quadratic	1	0.1566	0.1566	0.1566	0.84	0.382
Residual Error	10	1.8731	1.8731	0.1873		
Lack-of-Fit	6	0.6381	0.6381	0.1063	0.34	0.882
Pure Error	4	1.2350	1.2350	0.3088		
Total	13	5.9543				

Our linear terms are still significant, our lack of fit is still not significant. the analysis is saying that linear is adequate for this situation and this set of data.

One says 1 ingredient and the other says a blend - which one should we use?

I would like look at the variance. ...

24.9 is the predicted value.

By having a smaller, more parimonious model you decrease the variance. This is what you would expect with a model with fewer parameters. The standard error of the fit is a function of the design, and for this reason, the fewer the parameters the smaller the variance. But is also a function of residual error which gets smaller as we throw out terms that were not significant.

11.4 - Experiments with Computer Models

In many cases, performing actual experiments can be much too costly and cumbersome. Instead, there might be a computer simulation of the system available which could be used as a means to generate the response values at each design point -- as an proxy for the real system output. Generally, there are two types of simulation models: Deterministic and Stochastic. Deterministic simulation models are usually complex mathematical models which provide deterministic outputs and not a random variable. The output from a stochastic simulation model is a random variable. Normally, and for optimization purposes, a program of the simulation model is built (which is

called Metamodel) and based on the assumption that the simulation model is a true representation of reality, the achieved optimum condition should be in compliance with the real system. Research into optimal designs for complex models and optimal interpolation of the model output have become hot areas of research in recent years. However, in this course we will not cover any details about "experiments with computer models." More information can be found in the text and of course the relative references.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/57>

Links:

- [1] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/Ex-11-1-output.doc>
- [2] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/Ex11-1.mpj>
- [3] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/Ex11-2.MTW>
- [4] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/EX11-2.MPJ>
- [5] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/BD9-1.MTW>
- [6] <https://screencast.com/t/yfyiPXdPq>
- [7] [https://onlinecourses.science.psu.edu/stat503/javascript:popup_window\('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/L11_response_viewlet_swf.html','l11_response', 718, 668\);](https://onlinecourses.science.psu.edu/stat503/javascript:popup_window('stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/L11_response_viewlet_swf.html','l11_response', 718, 668);)
- [8] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/ch11.pdf>
- [9] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/Ex11-3.MTW>
- [10] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson11/Pr11-29.MTW>



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 12: Robust Parameter Designs

Lesson 12: Robust Parameter Designs

Introduction

In what we have discussed so far in the context of optimization only the average location of the response variable has been taken into account. However, from another perspective the variation of the response variable could be of major importance as well. This variation could be due to either usual noise of the process or randomness in the nature of one or more controllable factors of the process.

The Robust Parameter Design (RPD) approach initially proposed by Japanese engineer, Genichi Taguchi, seeks a combination of controllable factors such that two main objectives are achieved:

- The mean or average location of the response is at the desired level, and
- The variation or dispersion of the response is as small as possible.

Taguchi proposed that only some of the variables cause the variability of the process, which he named noise variables or uncontrollable variables. Please note that noise variables may be controllable in the laboratory, while in general they are a noise factor, and uncontrollable. An important contribution of RPD efforts is to identify both the controllable variables and the noise variables and find settings for the controllable variable such that the variation of response due to noise factors is minimized.

The general ideas of Taguchi widely spread throughout the world; however, his philosophy and methodology to handle RPD problems caused lots of controversy among statisticians. With the emergence of Response Surface Methodology (RSM), many efficient approaches were proposed which could nicely handle RPD problems. In what follows, RSM approaches for Robust Parameter Design will be discussed.

Learning Objectives and Outcomes

- Understanding the general idea of Robust Parameter Design approaches
- Getting familiar with Taguchi's crossed array design and its relative weaknesses
- Understanding combined array design and response model approach to RPD

12.1 - Crossed Array Design

Loading [MathJax]/extensions/MathZoom.js

Crossed array design was originally proposed by Taguchi. These designs consist of an **inner array** and an **outer array**. The inner array consists of the controllable factors while the outer array consists of the noise factors. The main feature of this design is that these two arrays are “crossed”; that is, every treatment combination in the inner array is run in combination with every treatment combination in the outer array. Table 12.2 is an example of crossed array design, where the inner array consists of four controllable factors and outer array consists of three noise factors. Note the typo in the levels of the 6th column of data. It should be $\{+, -, +\}$.

Crossed array designs provide sufficient information about the interaction between controllable factors and noise factors existing in the model which is an integral part of RPD problems. However, it can be seen that crossed array design may result in a large number of runs even for a fairly small number of controllable and noise factors. An alternative for these designs are **combined array designs** which is discussed in the next section.

The dominant method used to analyze crossed array designs is to model the mean and variance of the response variable separately, where the sample mean and variance can be calculated for each treatment combination in the inner array across all combinations of outer array factors. Consequently, these two new response variables can be considered as a dual response problem where the response variance needs to be minimized while response mean could be maximized, minimized or set close to a specified target. The text book has an example about the leaf spring experiment in which the resulting dual response problem has been solved by the overlaid contour plots method (See Figure 12.6) for multiple response problems, discussed in section 11.3.4.

12.2 - Combined Array Design

The combined array design approach treats all the variables the same, no matter they are controllable or noise. These models are capable of modeling main effects of controllable and noise factors and also their interactions. To illustrate, consider a case with two controllable and one noise factor. Equation 12.1 in the textbook gives a first-order model as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \gamma_1 z_1 + \delta_{11} x_1 z_1 + \delta_{21} x_2 z_1 + \varepsilon$$

where the β_i are coefficients of controllable factors, β_{12} is the coefficient of interaction of controllable factors, γ_1 is the coefficient of the noise factor and δ_{ij} are the coefficients of interaction between controllable and noise factors. As can be seen, the response model approach puts all of the variables, no matter they are controllable or noise, in a single experimental design. There exist some assumptions which are mentioned as follows:

- ε is a random variable with mean zero and variance σ^2
- Noise factors are random variables (although controllable in experiment) with mean zero and variance σ_z^2
- If there exist several noise factors their covariance is zero

Under these general assumptions, we will find the mean and variance for the given

example as following:

Loading [MathJax]/extensions/MathZoom.js

and

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \sigma^2$$

$$Var(y) = \sigma_z^2 (\gamma_1 + \delta_{11} x_1^2 + \delta_{21} x_2^2)$$

Notice that although the variance model involves only controllable variables but it also considers the interaction regression coefficients between the controllable and noise factors.

Finally, as before, we perform the optimization using any dual response approach like overlaid contours, desirability functions or etc. (Example 12.1 from the text book is a good example of overlaid contour plots approach).

From the design point of view, using any resolution V (or higher) design for the two level factor designs is efficient. Because these designs allow any main effect or two factor interaction to be estimated separately, assuming that three and higher factor interactions are negligible.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/74>

Loading [MathJax]/extensions/MathZoom.js



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 13: Experiments with Random Factors

Lesson 13: Experiments with Random Factors

Introduction

Throughout most parts of this course we have discussed experiments with **fixed factors**. That is, the levels used for the factors are those of interest by the experimenter and the inference made is confined to those specific levels. However, when factor levels are chosen at random from a larger population of potential levels, the factor is called a **random factor**. In this case, the statistical inference applies to the whole population of levels. **Random factor** models have many industrial applications including measurement system studies.

Learning Objectives & Outcomes

- Understanding the concept of random effect
- Getting familiar with random effect models and components of variance in each model
- Learning how to deal with models containing two random factors
- Getting familiar with how to analyze experiments where one of the factors is fixed and the other one is random
- Finding the expected mean squares using a simple algorithm

13.1 - Random Effects Models

Imagine that we randomly select a of the possible levels of the factor of interest. In this case, we say that the factor is random. Typically random factors are categorical. While continuous covariates may be measured at random levels, we usually think of the effects as being systematic (such as linear, quadratic or even exponential) effects. Random effects are not systematic. The model helps make this clear.

As before, the usual single factor ANOVA applies which is

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

However, herein, both the error term and treatment effects are random variables, that is

$$\varepsilon_{ij} \text{ is } NID(0, \sigma^2) \text{ and } \tau_i \text{ is } NID(0, \sigma_\tau^2)$$

Also, τ_i and ε_{ij} are independent. The variances σ_τ^2 and σ^2 are called **variance components**.

There might be some confusion about the differences between noise factors and random factors. Noise factors may be fixed or random. In Robust Parameter Designs we treat them as random because, although we control them in our experiment, they are not controlled under the conditions under which our system will normally be run. Factors are random when we think of them as a random sample from a larger population and their effect is not systematic.

It is not always clear when the factor is random. For example, if a company is interested in the effects of implementing a management policy at its stores and the experiment includes all 5 of its existing stores, it might consider "store" to be a fixed factor, because the levels are not a random sample. But if the company has 100 stores and picks 5 for the experiment, or if the company is considering a rapid expansion and is planning to implement the selected policy at the new locations as well, then "store" would be considered a random factor. We seldom consider random factors in 2^k or 3^k designs because 2 or 3 levels are not sufficient for estimating variances.

Typesetting math: 100%

In the fixed effect models we test the equality of the treatment means. However, this is no longer appropriate because treatments are randomly selected and we are interested in the population of treatments rather than any individual one. The appropriate hypothesis test for a random effect is:

$$\begin{aligned} H_0 : \sigma_{\tau}^2 &= 0 \\ H_1 : \sigma_{\tau}^2 &> 0 \end{aligned}$$

The standard ANOVA partition of the total sum of squares still works; and leads to the usual ANOVA display. However, as before, the form of the appropriate test statistic depends on the Expected Mean Squares. In this case, the appropriate test statistic would be

$$F_0 = MS_{Treatments} / MS_E$$

which follows an F distribution with $a-1$ and $N-a$ degrees of freedom. Furthermore, we are also interested in estimating the variance components σ_{τ}^2 and σ^2 . To do so, we use the **analysis of variance method** which consists of equating the expected mean squares to their observed values.

$$\hat{\sigma}^2 = MS_E \text{ and } \hat{\sigma}_{\tau}^2 + n\hat{\sigma}^2 = MS_{Treatments}$$

$$\hat{\sigma}_{\tau}^2 = \frac{MS_{Treatment} - MS_E}{n}$$

$$\hat{\sigma}^2 = MS_E$$

Potential problem that may arise here is that the estimated treatment variance component may be negative. In such a case, it is proposed to either consider zero in case of a negative estimate or use another method which always results in a positive estimate. A negative estimate for the treatment variance component can also be viewed as a evidence that the linear model is not appropriate, which suggests looking for a better one.

Example 3.11 from the text discusses a single random factor case about the difference of looms in a textile weaving company. Four looms have been chosen randomly from a population of looms within a weaving shed and four observations of fabric strength were made on each loom. The data obtained from the experiment are below.

Loom	Obs 1	Obs 2	Obs 3	Obs 4	row sum
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

Here is the Minitab output for this example using **Stat > ANOVA > Balanced ANOVA** command.

Factor	Type	Levels	Values
Loom	random	4	1 2 3 4
Analysis of Variance for y			
Source	DF	SS	MS F P
Loom	3	89.188	29.729 15.68 0.000
Error	12	22.750	1.896
Total	15	111.938	
Source Variance Error Expected Mean Square for Each Term			
component term (using unrestricted model)			
1 Loom	6.958	2 (2) + 4(1)	
2 Error	1.896	(2)	

Typesetting method: 100%

The interpretation made from the ANOVA table is as before. With the p -value equal to 0.000 it is obvious that the looms in the plant are significantly different, or more accurately stated, the variance component among the looms is significantly larger than zero. And confidence intervals can be found for the variance components. The $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{(N-a)MS_E}{\chi_{\alpha/2, N-a}^2} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi_{1-\alpha/2, N-a}^2}$$

Confidence intervals for other variance components are provided in the textbook. It should be noted that a closed form expression for the confidence interval on some parameters may not be obtained.

13.2 - Two Factor Factorial with Random Factors

Imagine that we have two factors, say A and B , that both have a large number of levels which are of interest. We will choose a random levels of factor A and b random levels for factor B and n observations are made at each treatment combination. The corresponding linear model for this case and the respective variance components would be

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

$$V(\tau_i) = \sigma_\tau^2, \quad V(\beta_j) = \sigma_\beta^2, \quad V[(\tau\beta)_{ij}] = \sigma_{\tau\beta}^2, \quad V(\varepsilon_{ijk}) = \sigma^2$$

$$V(y_{ijk}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$$

Where τ_i , β_j , $(\tau\beta)_{ij}$ and ε_{ijk} are all NID random variables with mean zero and variance as shown above. The relevant hypotheses that we are interested in testing are:

$$\begin{array}{lll} H_0 : \sigma_\tau^2 = 0 & H_0 : \sigma_\beta^2 = 0 & H_0 : \sigma_{\tau\beta}^2 = 0 \\ H_1 : \sigma_\tau^2 > 0 & H_1 : \sigma_\beta^2 > 0 & H_1 : \sigma_{\tau\beta}^2 > 0 \end{array}$$

The numerical calculations for the analysis of variance are exactly like in the fixed effect case. However, we state once again, that to form the test statistics, the expected mean squares should be taken into account. We state the expected mean squares (EMS) here and assuming the hypothesis is true, we form the F test statistics, so that under that assumption, both the numerator and denominator of the F statistic have the same expectation. Note that the test for the main effects are no longer what they were in the fixed factor situation.

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2 \implies F_0 = \frac{MS_A}{MS_{AB}}$$

$$E(MS_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2 \implies F_0 = \frac{MS_B}{MS_{AB}}$$

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2 \implies F_0 = \frac{MS_{AB}}{MS_E}$$

$$E(MS_E) = \sigma^2$$

Furthermore, variance components can again be estimated using the analysis of variance method by equating the expected mean squares to their observed values.

$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_{AB}}{bn}$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - MS_{AB}}{an}$$

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n}$$

Typesetting math: 100%

Example 13.2 in the textbook discusses a two-factor factorial with random effects on a measurement system capability study. These studies are often called gauge ~~capability~~^{ANDE} studies or gauge repeatability and reproducibility (R&R) studies. In this example three randomly selected operators are selected to measure twenty randomly selected parts, each part twice. Data obtained from the experiment is shown in Table 13.3. The variance components are

2

$$\sigma_y^2 = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$$

Typically, σ^2 is called gauge repeatability because it shows the variation of the same part measured by the same operator and $\sigma_\beta^2 + \sigma_{\tau\beta}^2$

which reflects the variation resulting from operators is called gauge reproducibility. Table 13.4 shows the analysis using Minitab's Balanced ANOVA command.

Table 13-4 Analysis of Variance (Minitab Balanced ANOVA) for Example 13-2

Analysis of Variance (Balanced Designs)							
Factor	Type	Levels	Values				
part	random	20	1 8 15	2 9 16	3 10 17	4 11 18	5 12 19
				6 13	7 14		
operator	random	3	1 2	3			

Analysis of Variance for y					
Source	DF	SS	MS	F	P
part	19	1185.425	62.391	87.65	0.000
operator	2	2.617	1.308	1.84	0.173
part*operator	38	27.050	0.712	0.72	0.861
Error	60	59.500	0.992		
Total	119	1274.592			

Source	Variance component	Error term	Expected Mean Square for Each Term (using unrestricted model)
1 part	10.2798	3	(4) + 2(3) + 6(1)
2 operator	0.0149	3	(4) + 2(3) + 40(2)
3 part*operator	-0.1399	4	(4) + 2(3)
4 Error	0.9917		(4)

Table 13.4 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)

As it can be seen, the only significant effect is *part*. Estimates for components of variance and expected mean square for each term are given at the lower part of the table. Notice that the estimated variance for interaction term *part*Operator* is negative. The fact that the p-value for the interaction term is large along with the negative estimate of its variance is a good sign that the interaction term is actually zero. Therefore, we can proceed and fit a reduced model without *part*operator* term. The analysis of variance for the reduced model can be found in Table 13.5.

Typesetting math: 100%

Table 13-5 Analysis of Variance for the Reduced Model, Example 13-2

Analysis of Variance (Balanced Designs)								
Factor	Type	Levels	Values	1	2	3	4	5
part	random	20	8	9	10	11	12	13
			15	16	17	18	19	20
operator	random	3	1	2	3			

Analysis of Variance for y						
Source	DF	SS	MS	F	P	
part	19	1185.425	62.391	70.64	0.000	
operator	2	2.617	1.308	1.48	0.232	
Error	98	86.550	0.883			
Total	119	1274.592				

Source	Variance component	Error term	Expected Mean Square for Each Term (using unrestricted model)
1 part	10.2513	3	(3) + 6(1)
2 operator	0.0106	3	(3) + 40(2)
3 Error	0.8832		(3)

Table 13.5 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)

Since the interaction term is zero, both of the effects is tested against the error term. Estimates of the variance component are given below at lower part of the table. Furthermore, as mentioned before, estimate of the variance of the gauge can be achieved as

$$\hat{\sigma}_{gauge}^2 = \hat{\sigma}^2 + \hat{\sigma}_\beta^2 = 0.88 + 0.01 = 0.89$$

which is relatively small comparing to the variability of the product.

13.3 - The Two Factor Mixed Models

Next, consider the case that one of the factors is fixed, say A, and the other one (B) is a random factor. This case is called the two-factor mixed model and the linear statistical model and respective components of variance is

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

$$V(\beta_j) = \sigma_\beta^2, \quad V[(\tau\beta)_{ij}] = [(a-1)/a]\sigma_{\tau\beta}^2, \quad V(\varepsilon_{ijk}) = \sigma^2$$

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

Here τ_i is a fixed effect but β_j and $(\tau\beta)_{ij}$ are assumed to be random effects and ε_{ijk} is a random error. Furthermore, β_j and ε_{ijk} are NID. The interaction effect is also normal but **not** independent. There often is a restriction imposed on the interaction which is

$$\sum_{i=1}^a (\tau\beta)_{ij} = (\tau\beta)_{.j} = 0 \quad j = 1, 2, \dots, b$$

Because of the sum of interaction effects over the levels of the fixed factor equals zero, this version of the mixed model is called the **restricted model**. There exists another model which does not include such a restriction and is discussed later. Neither of these models is "correct" or "wrong" - they are both theoretical models for how the data behave. They have different implications for the meanings of the variance components. The restricted model is

Typesetting math: 100%

often used in the ANOVA setting. The unrestricted model is often used for more general designs that include continuous covariates and repeated or spatially correlated measurements.

Once again the tests of hypotheses for the mixed-model are:

$$\begin{array}{lll} H_0 : \tau_i = 0 & H_0 : \sigma_{\beta}^2 = 0 & H_0 : \sigma_{\tau\beta}^2 = 0 \\ H_1 : \tau_i \neq 0 & H_1 : \sigma_{\beta}^2 > 0 & H_1 : \sigma_{\tau\beta}^2 > 0 \end{array}$$

Furthermore, test statistics which are based on the expected mean squares are summarized as follows

$$\begin{aligned} E(MS_A) &= \sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1} \implies F_0 = \frac{MS_A}{MS_{AB}} \\ E(MS_B) &= \sigma^2 + an\sigma_{\beta}^2 \implies F_0 = \frac{MS_B}{MS_E} \\ E(MS_{AB}) &= \sigma^2 + n\sigma_{\tau\beta}^2 \implies F_0 = \frac{MS_{AB}}{MS_E} \\ E(MS_E) &= \sigma^2 \end{aligned}$$

In the mixed model, it is possible to estimate the fixed factor effects as before which are shown here:

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} \\ \hat{\tau}_i &= \bar{y}_{i..} - \bar{y}_{..} \end{aligned}$$

The variance components can be estimated using the analysis of variance method by equating the expected mean squares to their observed values:

$$\begin{aligned} \hat{\sigma}_{\beta}^2 &= \frac{MS_B - MS_E}{an} \\ \hat{\sigma}_{\tau\beta}^2 &= \frac{MS_{AB} - MS_E}{n} \\ \hat{\sigma}^2 &= MS_E \end{aligned}$$

Example 13.3 is the measurement system capability experiment where here we assume the *operator* has become a fixed factor while *part* is left as a random factor. Assuming the restricted version of the mixed effect model, Minitab's balanced ANOVA routine output is given as follows.

Typesetting math: 100%

Table 13-6 Analysis of Variance (Minitab) for the Mixed Model in Example 13-3.
The Restricted Model is Assumed.

Analysis of Variance (Balanced Designs)								
Factor	Type	Levels	Values	1	2	3	4	5
part	random	20	1	2	3	4	5	6
			8	9	10	11	12	13
			15	16	17	18	19	14
operator	fixed	3	1	2	3			

Analysis of Variance for y					
Source	DF	SS	MS	F	P
part	19	1185.425	62.391	62.92	0.000
operator	2	2.617	1.308	1.84	0.173
part*operator	38	27.050	0.712	0.72	0.861
Error	60	59.500	0.992		
Total	119	1274.592			

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 part	10.2332	4	(4) + 6(1)
2 operator		3	(4) + 2(3) + 40Q[2]
3 part*operator	-0.1399	4	(4) + 2(3)
4 Error	0.9917		(4)

Table 13.6 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)

Like before, there exists a large effect of *parts*, small *operator* effect and no *part*operator* interaction. Notice that again the variance component estimate for the *part*operator* interaction is negative, which considering its insignificant effect, leads us to assume it is zero and to delete this term from the model.

As mentioned before, there exist alternative analyses for the mixed effect models which are called the **unrestricted mixed models**. The linear statistical model and components of variance for the unrestricted mixed model are given as:

$$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

$$V(\gamma_j) = \sigma_\beta^2, \quad V[(\alpha\gamma)_{ij}] = \sigma_{\alpha\gamma}^2, \quad V(\varepsilon_{ijk}) = \sigma^2$$

$$\sum_{i=1}^a \alpha_i = 0$$

In the unrestricted mixed model, all of the random terms are assumed to be Normally and independently distributed (*NID*) and there is not a restriction on the interaction term which was previously imposed. As before, the relevant tests of hypotheses are given by:

$$\begin{array}{lll} H_0 : \alpha_i = 0 & H_0 : \sigma_\beta^2 = 0 & H_0 : \sigma_{\alpha\gamma}^2 = 0 \\ H_1 : \alpha_i \neq 0 & H_1 : \sigma_\beta^2 > 0 & H_1 : \sigma_{\alpha\gamma}^2 > 0 \end{array}$$

And the expected mean squares which determine the test statistics are

$$E(MS_A) = \sigma^2 + n\sigma_{\alpha\gamma}^2 + \frac{bn \sum_{i=1}^a \alpha_i^2}{a-1} \implies F_0 = \frac{MS_A}{MS_{AB}}$$

Typesetting math: 100%

$$\begin{aligned} E(MS_{AB}) &= \sigma^2 + n\sigma_{\alpha\gamma}^2 + an\sigma_\gamma^2 \implies F_0 \equiv \frac{MS_{AB}}{\frac{MS_B - MS_{AB}}{an}} \\ E(MS_E) &= \sigma^2 \end{aligned}$$

Again, to estimate the variance components, the analysis of variance method is used and the expected mean squares are equated to their observed values which result in:

$$\hat{\sigma}_\gamma^2 = \frac{MS_B - MS_{AB}}{an}$$

$$\hat{\sigma}_{\alpha\gamma}^2 = \frac{MS_{AB} - MS_E}{n}$$

$$\hat{\sigma}^2 = MS_E$$

Example 13.4 uses the unrestricted mixed model to analyze the measurement systems capability experiment. The Minitab solution for this unrestricted model is given here:

Table 13-7 Analysis of the Experiment in Example 13-3 Using the Unrestricted Model

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values	20	1	2	3	4	5	6	7
part	random			8	9	10	11	12	13	14	
				15	16	17	18	19	20		
operator	fixed	3	1	2	3						

Analysis of Variance for y

Source	DF	SS	MS	F	P
part	19	1185.425	62.391	87.65	0.000
operator	2	2.617	1.308	1.84	0.173
part*operator	38	27.050	0.712	0.72	0.861
Error	60	59.500	0.992		
Total	119	1274.592			

Source	Variance component	Error term	Expected Mean Square for Each Term (using unrestricted model)
1 part	10.2798	3	(4) + 2(3) + 6(1)
2 operator		3	(4) + 2(3) + 0[2]
3 part*operator	-0.1399	4	(4) + 2(3)
4 Error	0.9917		(4)

Table 13.7 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)

It is difficult to provide guidelines for when the restricted or unrestricted mixed model should be used, because statisticians do not fully agree on this. Fortunately, the inference for the fixed effects does not differ for the 2 factor mixed model which is most often seen, and is usually the same in more complicated models as well.

13.4 - Finding Expected Mean Squares

As we have demonstrated, determining the appropriate test statistics in the analysis of variance method depends on finding the expected mean squares. In complex design situations and in the presence of random or mixed models it is tedious to apply the expectation operator. Therefore, it would be helpful to have a formal procedure by which we could derive the expected mean squares for the different terms in the model. Page 523 has listed a set of rules which works for any set of balanced models to derive the expected mean squares. These rules are consistent with the **restricted** mixed model and can be modified to incorporate the **unrestricted** model assumptions, as well.

Typesetting math: 100%

It is worth mentioning that the test statistic is a ratio of two mean squares where the expected value of the numerator mean square differs from the expected value of the denominator mean square by the variance component or the fixed factor in which we are interested. Therefore, under the assumption of the null hypothesis, both the numerator and the denominator of the F ratio have the same EMS.

13.5 - Approximate F Tests

Sometimes in factorial experiments with three or more factors involving a random or mixed model, we determine that there is no exact test statistic for certain effects in the model. Satterthwaite (1946) proposed a test procedure which uses the **linear combinations** of the original mean squares to form the F-ratio. These linear combinations of the original mean squares are sometimes called "synthetic" mean squares. Details on how to build the test statistic and adjustments made to degrees of freedom based on Satterthwaite procedure can be found in Section 13.6.

Minitab will analyze these experiments and derive "synthetic" mean squares, although their "synthetic" mean squares are not always the best choice. Approximate tests based on large samples (which use modified versions of the Central Limit Theorem) are also available. Unfortunately, this is another case in which it is not clear that there is a best method.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/82>

Typesetting math: 100%



Published on STAT 503 (<https://onlinecourses.science.psu.edu/stat503>)

[Home](#) > Lesson 14: Nested and Split Plot Designs

Lesson 14: Nested and Split Plot Designs

Introduction

Nested and Split Plot experiments are *multifactor* experiments that have some important industrial applications although historically these come out of agricultural contexts. "Split plot" designs -- here we are originally talking about fields which are divided into whole and split plots, and then individual plots get assigned different treatments. For instance, one whole plot might have different irrigation techniques or fertilization strategies applied, or the soil might be prepared in a different way. The whole plot serves as the experimental unit for this particular treatment. Then we could divide each whole plot into sub plots, and each subplot is the experimental unit for another treatment factor.

Whenever we talk about split plot designs we focus on the experimental unit for a particular treatment factor.

Nested and split-plot designs frequently involve one or more *random* factors, so the methodology of Chapter 13 of our text (expected mean squares, variance components) is important.

There are *many* variations of these designs – here we will only consider some of the more basic situations.

Learning Objectives & Outcomes

- Understanding the concept of nesting factors inside another factor.
- Getting familiar with the two-stage nested designs where either or both of the factors could be fixed or random.
- Getting familiar with split-plot designs and their applications where changing the level of some of the factors is hard, relative to other factors.
- Understanding the two main approaches to analyze the split- plot designs and their derivatives and the basis for each approach.
- Getting familiar with split-split-plot designs as an extension of split-plot designs.
- Getting familiar with strip- plot designs (or split-block designs) and their difference from the split-plot designs.

14.1 - The Two-Stage Nested Design

When factor B is nested in levels of factor A, the levels of the nested factor don't have exactly the same meaning under each level of the main factor, in this case factor A. In a nested design, the levels of factor (B) are not identical to each other at different levels of factor (A), although they might have the same labels. For example, if A is school and B is teacher, teacher 1 will differ between the schools. This has to be kept in mind when trying to determine if the design is crossed or nested. To be crossed, the same teacher needs to teach at all the schools.

As another example, consider a company that purchases material from three suppliers and the material comes in batches. In this case, we might have 4 batches from each supplier, but the batches don't have the same characteristics of quality when purchased from different suppliers. Therefore, the batches would be nested. When we have a nested factor and you want to represent this in the model the identity of the batch always requires an index of the factor in which it is nested. The linear statistical model for the two-stage nested design is:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{k(j)} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

The subscript $j(i)$ indicates that j^{th} level of factor B is nested under the i^{th} level of factor A. Furthermore, it is useful to think of replicates as being nested under the treatment combinations; thus, $k(j)$ is used for the error term. Because not every level of B appears with every level of A, there is no interaction between A and B. (In most of our designs, the error is nested in the treatments, but we only use this notation for error when there are other nested factors in the design).

When B is a random factor nested in A, we think of it as the replicates for A. So whether factor A is a fixed or random factor the error term for testing the hypothesis about A is based on the mean squares due to B(A) which is read "B nested in A". Table 14.1 displays the expected mean squares in the two-stage nested design for different combinations of factor A and B being fixed or random.

$E(MS)$	A Fixed B Fixed	A Fixed B Random	A Random B Random
$E(MS_A)$	$\sigma^2 + \frac{bn \sum \tau_i^2}{a - 1}$	$\sigma^2 + n\sigma_B^2 + \frac{bn \sum \tau_i^2}{a - 1}$	$\sigma^2 + n\sigma_B^2 + bn\sigma_\tau^2$
$E(MS_{B(A)})$	$\sigma^2 + \frac{n \sum \sum \beta_{Ai}^2}{a(b - 1)}$	$\sigma^2 + n\sigma_B^2$	$\sigma^2 + n\sigma_B^2$
$E(MS_E)$	σ^2	σ^2	σ^2

Table 14.1 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition)

The analysis of variance table is shown in table 14.2.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
A	$bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$	MS_A
B within A	$n \sum \sum (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b - 1)$	$MS_{B(A)}$
Error	$\sum \sum \sum (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n - 1)$	MS_E
Total	$\sum \sum \sum (y_{ijk} - \bar{y}_{...})^2$	$abn - 1$	

Table 14.2 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition)

Another way to think about this is to note that batch is the experimental unit for the factor 'supplier'. Does it matter how many measurements you make on each batch? (Yes, this will improve your measurement precision on the batch.) However, the variability among the batches from the supplier is the appropriate measure of the variability of factor A, the suppliers.

Essentially the question that we want to answer is, "Is the purity of the material the same across suppliers?"

In this example the model assumes that the batches are random samples from each supplier, i.e. suppliers are fixed, the batches are random, and the observations are random.

Experimental design: Select four batches at random from each of three suppliers. Make three purity determinations from each batch. See the schematic representation of this design in Fig. 14-1.

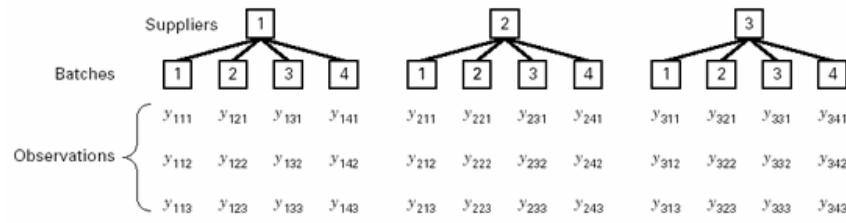


Figure 14-1 A two-stage nested design.

Figure 14.1 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition)

It is the average of the batches and the variability across the batches that are most important. When analyzing these data, we want to decide which supplier should they use? This will depends on both the supplier mean and the variability among batches?

Here is the design question: How many batches should you take and how many measurements should you make on each batch? This will depend on the cost of performing a measurement versus the cost of getting another batch. If measurements are expensive one could get many batches and just take a few measurements on each batch, or if it is costly to get a new batch then you may want to spend more money taking many multiple measurements per batch.

At a minimum you need at least two measurements ($n = 2$) so that you can estimate the variability among your measurements, σ^2 , and at least two batches per supplier ($b = 2$) so you can estimate the variability among batches, σ^2_{β} . Some would say that you need at least three in order to be sure!

To repeat the design question: how large should b and n be, or, how many batches versus how many samples per batch? This will be a function of the cost of taking a measurement and the cost of getting another batch. In order to answer these questions you need to know these cost functions. It will also depend on the variance among batches versus the variance of the measurements within batches.

Minitab can provide the estimates of these variance components.

Minitab General Linear Model (unlike SAS GLM), bases its F tests on what the expected mean squares determine is the appropriate error. The program will tell us that when we test the hypothesis of no supplier effect, we should use the variation among batches (since Batch is random) as the error for the test.

Run the example given in Minitab [Example14-1.MPJ](#) [1] to see the test statistic, which is distributed as an F -distribution with 2 and 9 degrees of freedom.

Practical Interpretation – Example 14.1

There is no significant difference (p -value = 0.416) in purity among suppliers, but significant variation exists (p -value = 0.017) in purity among batches (within suppliers)

What are the practical implications of this conclusion?

Examine the residual plots. The plot of residuals versus supplier is very important (why?)

An assumption in the Analysis of Variance is that the variances are all equal. The measurement error should not depend on the batch means, i.e., the variation in measurement error is probably the same for a high-quality batch as it is for low-quality batch. We also assume the variability among batches, σ^2_B , is the same for all suppliers. This is an assumption that you will want to check! Because the whole reason one supplier might be better than another is because they have lower variation among their batches. We always need to know what assumptions we are making and whether they are true or not. It is often the most important thing to learn - when you learn there is a failed assumption!

What if we had incorrectly analyzed this experiment as a crossed factorial rather than a nested design? The analysis would be:

The inappropriate Analysis of variance for crossed effects is shown in Table 14.5.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Suppliers (S)	15.06	2	7.53	1.02	0.42
Batches (B)	25.64	3	8.55	3.24	0.04
$S \times B$ interaction	44.28	6	7.38	2.80	0.03
Error	63.33	24	2.64		
Total	148.31	35			

Table 14.5 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition)

This analysis indicates that batches differ significantly and that there is significant interaction between batch and supplier. However, neither the main effect of Batch nor the interaction is meaningful, since batches are not the same across suppliers. Note that the sum of the Batch and the $S \times B$ Sum of Squares and Degree of Freedom is the Batch (Supplier) line in the correct Table.

For the model with the A factor also a random effect, analysis of variance method can be used to estimate all three components of variance.

$$\hat{\sigma}^2 = MS_E$$

$$\hat{\sigma}_{\beta}^2 = \frac{MS_{B(A)} - MS_E}{n}$$

And

$$\hat{\sigma}_{\tau}^2 = \frac{MS_A - MS_{B(A)}}{bn}$$

14.2 - The General m-Stage Nested Design

The results from the previous section can easily be generalized to the case of m completely nested factors. The text book gives an example of a 3-stage nested design in which the effect of two formulations on the alloy harness is of interest. To perform the experiment, three heats of each alloy formulation are prepared, two ingots are selected at random from each heat, and two harness measurements are made on each ingot. Figure 14.5 shows the situation.

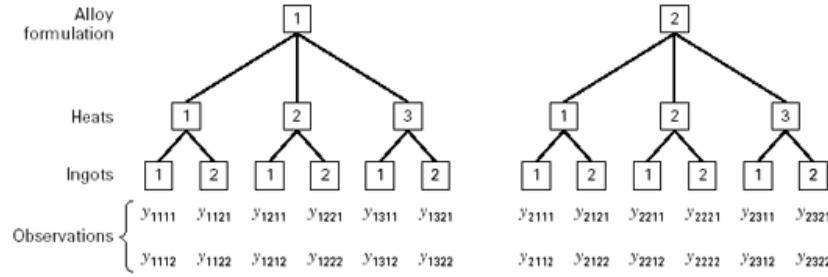


Figure 14-5 A three-stage nested design.

Figure 14.5 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition)

The linear statistical model for the 3-stage nested design would be

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(j)} + \varepsilon_{l(ijk)} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, n \end{cases}$$

Where τ_i is the effect of the i^{th} alloy formulation, $\beta_{j(i)}$ is the effect of the j^{th} heat within the i^{th} alloy, and $\gamma_{k(j)}$ is the effect of the k^{th} ingot within the j^{th} heat and i^{th} alloy and $\varepsilon_{l(ijk)}$ is the usual NID error term. The calculation of the sum of squares for the analysis of variance is shown in Table 14.8.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
A	$bcn \sum_i (\bar{y}_{i...} - \bar{y}_{...})^2$	$a - 1$	MS_A
B (within A)	$cn \sum_j \sum_i (\bar{y}_{ij..} - \bar{y}_{i...})^2$	$a(b - 1)$	$MS_{B(A)}$
C (within B)	$n \sum_i \sum_j \sum_k (\bar{y}_{ijk.} - \bar{y}_{ik.})^2$	$ab(c - 1)$	$MS_{C(B)}$
Error	$\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \bar{y}_{ijk.})^2$	$abc(n - 1)$	MS_E
Total	$\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \bar{y}_{...})^2$	$abcn - 1$	

Table 14.8 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition) (Please note: the Sum of Squares formulas for B(A) and C(B) have an error - they should have the A means and B means subtracted, respectively, not the overall mean.)

To test the hypotheses and to form the test statistics once again we use the expected mean squares. Table 14.9 illustrates the calculated expected mean squares for a three-stage nested design with A and B fixed and C random.

Factor	F_a i	F_b j	R_c k	R_n l	Expected Mean Square
τ_i	0	b	c	n	$\sigma^2 + n\sigma^2_\gamma + \frac{bcn \sum \tau_i^2}{a-1}$
$\beta_{j(i)}$	1	0	c	n	$\sigma^2 + n\sigma^2_\gamma + \frac{cn \sum \sum \beta_{j(i)}^2}{a(b-1)}$
$\gamma_{k(ij)}$	1	1	1	n	$\sigma^2 + n\sigma^2_\gamma$
$\epsilon_{l(ijk)}$	1	1	1	1	σ^2

Table 14.9 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition)

14.3 - The Split-Plot Designs

Note: It is worth mentioning the fact that the notation used in this section (especially the use of Greek rather than Latin letters for the random error terms in the linear models) is not our preference. But, we decided to keep the text book's notation to avoid any possible confusion.

There exist some situations in multifactor factorial experiments where the experimenter may not be able to randomize the runs completely. Three good examples of split-plot designs can be found in the article: "[How to Recognize a Split Plot Experiment](#)" [2] by Scott M. Kowalski and Kevin J. Potcner, *Quality Progress*, November 2003.

Another good example of such a case is in the text book in Section 14.4. The example is about a paper manufacturer who wants to analyze the effect of three pulp preparation methods and four cooking temperatures on the tensile strength of the paper. The experimenter wants to perform three replicates of this experiment on three different days each consisting of 12 runs (3×4). The important issue here is the fact that making the pulp by any of the methods is cumbersome. Thus method is a "hard to change" factor. It would be economical to randomly select any of the preparation methods, make the blend and divide it into four samples and cook each of them with one of the four cooking temperatures. Then the second method is used to prepare the pulp and so on. As we can see, in order to achieve this economy in the process, there is a restriction on the randomization of the experimental runs.

In this example, each replicate or block is divided into three parts called **whole plots** (Each preparation method is assigned to a whole plot). Next, each whole plot is divided into four samples which are **split-plots** and one temperature level is assigned to each of these split-plots. It is important to note that since the whole-plot treatment in the split-plot design is confounded with whole plots and the split-plot treatment is not confounded, if possible, it is better to assign the factor we are most interested in to split plots.

Analysis of Split-Plot designs

In the statistical analysis of split-plot designs, we must take into account the presence of two different sizes of experimental units used to test the effect of whole plot treatment and split-plot treatment. Factor A effects are estimated using the whole plots and factor B and the A*B interaction effects are estimated using the split plots. Since the size of whole plot and split plots are different, they have different precisions. Generally, there exist two main approaches to analyze the split- plot designs and their derivatives.

1. First approach uses the Expected Mean Squares of the terms in the model to build the test statistics and is the one discussed by the book. The major disadvantage to this approach is the fact that it does not consider the randomization restrictions which may exist in any experiment.
2. Second approach which might be of more interest to statisticians and the one which considers any restriction in randomization of the runs is considered as the tradition approach to the analysis of split-plot designs.

Both of the approaches will be discussed but there will be more emphasis on the second approach, as it is more widely accepted for analysis of split-plot designs. It should be noted that the results from the two approaches may not be much different.

The linear statistical model given in the text for the split-plot design is:

Where, τ_i , β_j and $(\tau\beta)_{ij}$ represent the whole plot and γ_k , $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$ and $(\tau\beta\gamma)_{ijk}$ represent the split-plot. Here τ_i , β_j and γ_k are block effect, factor A effect and factor B effect, respectively. The sums of squares for the factors are computed as in the three-way analysis of variance without replication.

To analyze the treatment effects we first follow the approach discussed in the book. Table 14.17 shows the expected mean squares used to construct test statistics for the case where replicates or blocks are random and whole plot treatments and split-plot treatments are fixed factors.

		r	a	b	1	
	Factor	R	F	F	R	
		i	j	k	h	Expected Mean Square
Whole plot	τ_i	1	a	b	1	$\sigma^2 + ab\sigma_\tau^2$
	β_j	r	0	b	1	$\sigma^2 + b\sigma_{\tau\beta}^2 + \frac{rb \sum \beta_j^2}{a-1}$
	$(\tau\beta)_{ij}$	1	0	b	1	$\sigma^2 + b\sigma_{\tau\beta}^2$
Subplot	γ_k	r	a	0	1	$\sigma^2 + a\sigma_{\tau\gamma}^2 + \frac{ra \sum \gamma_k^2}{(b-1)}$
	$(\tau\gamma)_{ik}$	1	a	0	1	$\sigma^2 + a\sigma_{\tau\gamma}^2$
	$(\beta\gamma)_{jk}$	r	0	0	1	$\sigma^2 + \sigma_{\tau\beta\gamma}^2 + \frac{r \sum (\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
	$(\tau\beta\gamma)_{ijk}$	1	0	0	1	$\sigma^2 + \sigma_{\tau\beta\gamma}^2$
	$\epsilon_{(ijk)h}$	1	1	1	1	σ^2 (not estimable)

Table 14.17 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition)

The analysis of variance for the tensile strength is shown in the Table 14.16.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Replicates (or Blocks)	77.55	2	38.78		
Preparation method (A)	128.39	2	64.20	7.08	0.05
Whole Plot Error (replicates (or Blocks) \times A)	36.28	4	9.07		
Temperature (B)	434.08	3	144.69	41.94	<0.01
Replicates (or Blocks) \times B	20.67	6	3.45		
AB	75.17	6	12.53	2.96	0.05
Subplot Error (replicates (or Blocks) \times AB)	50.83	12	4.24		
Total	822.97	35			

Table 14.18 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition)

As mentioned earlier analysis of split-plot designs using the second approach is based mainly on the randomization restrictions. Here, the whole plot section of the analysis of variance could be considered as a Randomized Complete Block Design or RCBD with Method as our single factor (If we didn't have the blocks, it could be considered as a Complete Randomized Design or CRD). Remember how we dealt with these designs (Step back to Chapter 4). The error term which we used to construct our test statistic (The sum of square of which was achieved by subtraction) is just the interaction between our single factor and the Blocks. (If you recall, we mentioned that any interaction between the Blocks and the treatment factor is considered part of the experimental error). Similarly, in the split-plot section of the analysis of variance, all the interactions which include the Block term are pooled to form the error term of the split-plot section. If we ignore method, we would have an RCBD where the blocks are the individual preparations. However, there is a systematic effect due to method, which is taken out of the Block effect. Similarly, the block by temperature has a systematic effect due to method*temperature, so a SS for this effect is removed from the block*temperature interaction. SO, one way to think of the SP Error is that it is Block*Temp+Block*Method*Temp with $2*3+2*2*3=18$ d.f.

The Mean Square error terms derived in this fashion are then be used to build the F test statistics of each section of ANOVA table, repectively. Below, we have implemented this second approach for data. To do so, we have first

produced the ANOVA table using the GLM command in Minitab, assuming a full factorial design. Next, we have pooled the sum of squares and their respective degrees of freedom to create the SP Error term as described.

	DF	SS	MS	F	P-value
Blocks	2	77.556	38.78	4.28	0.1014
Method	2	128.389	64.19	7.08	0.0485
WP Error (Blocks*Method)	4	36.278	9.07		
Temp	3	434.083	144.69	36.43	0.0000
Method*Temp	6	75.167	12.53	3.15	0.0272
SP Error	18	71.5	3.97		
	35	822.973			

As you can see, there is a little difference between the output of analysis of variance performed in this manner and the one using the Expected Mean Squares because we have pooled Block*Temp and Blocks*Method*Temp to form the subplot error.

Advantages and Disadvantages of Split-Plot Experiments

In summary, when one of the treatment factors needs more replication or experimental units (material) than another or when it is hard to change the level of one of the factors, these designs become important. The primary disadvantage of these designs is the loss in precision in the whole plot treatment comparison and the statistical complexity.

14.4 - The Split-Split-Plot Design

The restriction on randomization mentioned in the split-plot designs can be extended to more than one factor. For the case where the restriction is on two factors the resulting design is called a **split-split-plot design**. These designs usually have three different sizes or types of experimental units.

Example 14.4 of the text book (Design and Analysis of Experiments, Douglas C. Montgomery, 7th and 8th Edition) discusses an experiment in which a researcher is interested in studying the effect of technicians, dosage strength and wall thickness of the capsule on absorption time of a particular type of antibiotic. There are three technicians, three dosage strengths and four capsule wall thicknesses resulting in 36 observations per replicate and the experimenter wants to perform four replicates on different days. To do so, first, technicians are randomly assigned to units of antibiotics which are the whole plots. Next, the three dosage strengths are randomly assigned to split-plots. Finally, for each dosage strength, the capsules are created with different wall thicknesses, which is the split-split factor and then tested in random order.

First notice the restrictions that exist on randomization. Here, we can not simply randomize the 36 runs in a single block (or replicate) because we have our first hard to change factor, named Technician. Furthermore, even after selecting a level for this hard to change factor (say technician 2) we can not randomize the 12 runs under this technician because we have another hard to change factor, named dosage strength. After we select a random level for this second factor, say dosage strength of level 3, we can then randomize the four runs under these two combinations of two factors and randomly run the experiments for different wall thicknesses as our third factor.

The linear statistical model for the split-split-plot design would be:

$$y_{ijkh} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \delta_h + (\tau\delta)_{ih} + (\beta\delta)_{jh} + (\tau\beta\delta)_{ijh} + (\gamma\delta)_{kh} + (\tau\gamma\delta)_{ikh} + (\beta\gamma\delta)_{jkh} + (\tau\beta\gamma\delta)_{ijkh} + \varepsilon_{ijkh}$$

$$\left\{ \begin{array}{l} i = 1, 2, \dots, r \\ j = 1, 2, \dots, a \\ k = 1, 2, \dots, b \\ h = 1, 2, \dots, c \end{array} \right.$$

Using the Expected Mean Square approach mentioned earlier for split-plot designs, we can proceed and analyze the split-split-plot designs, as well. Based on Expected Mean Squares given in Table 14.25 to build test statistics (assuming the block factor to be random and the other factors to be fixed), , and are whole plot, split-plot and split-split-plot errors, respectively. Minitab handles this model exactly in this way by GLM. (This was Table 14.22 in the 7th edition. The 8th edition has only the factors and EMS without the list of subscripts.)

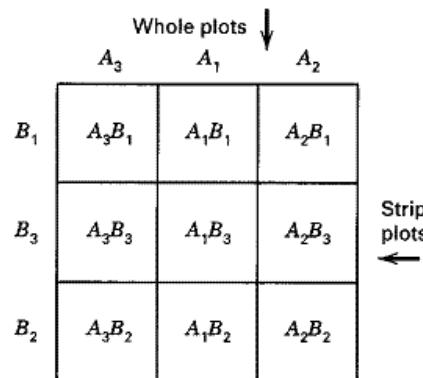
		<i>r</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>l</i>	Expected Mean Square
	Factor	<i>R</i> <i>i</i>	<i>F</i> <i>j</i>	<i>F</i> <i>k</i>	<i>F</i> <i>h</i>	<i>R</i> <i>l</i>	
Whole plot	τ_i	1	<i>a</i>	<i>b</i>	<i>c</i>	1	$\sigma^2 + abcd\sigma_\tau^2$
	β_j	<i>r</i>	0	<i>b</i>	<i>c</i>	1	$\sigma^2 + \sigma_{\beta\beta}^2 + \frac{rbc \sum \beta_j^2}{(a-1)}$
	$(\tau\beta)_{ij}$	1	0	<i>b</i>	<i>c</i>	1	$\sigma^2 + b c \sigma_{\tau\beta}^2$
Subplot	γ_k	<i>r</i>	<i>a</i>	0	<i>c</i>	1	$\sigma^2 + ac\sigma_{\gamma\gamma}^2 + \frac{rac \sum \gamma_k^2}{(b-1)}$
	$(\tau\gamma)_{ijk}$	1	<i>a</i>	0	<i>c</i>	1	$\sigma^2 + ac\sigma_{\tau\gamma}^2$
	$(\beta\gamma)_{ijk}$	<i>r</i>	0	0	<i>c</i>	1	$\sigma^2 + cc\sigma_{\beta\gamma}^2 + \frac{rc \sum \sum (\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
	$(\tau\beta\gamma)_{ijkl}$	1	0	0	<i>c</i>	1	$\sigma^2 + c\sigma_{\tau\beta\gamma}^2$
Sub-subplot	δ_h	<i>r</i>	<i>a</i>	<i>b</i>	0	1	$\sigma^2 + ab\sigma_{\tau\delta}^2 + \frac{rab \sum \delta_k^2}{(c-1)}$
	$(\tau\delta)_{ih}$	1	<i>a</i>	<i>b</i>	0	1	$\sigma^2 + ab\sigma_{\tau\delta}^2$
	$(\beta\delta)_{ih}$	<i>r</i>	0	<i>b</i>	0	1	$\sigma^2 + b\sigma_{\beta\delta}^2 + \frac{rb \sum \sum (\beta\delta)_{jk}^2}{(a-1)(c-1)}$
	$(\tau\beta\delta)_{ijkh}$	1	0	<i>b</i>	0	1	$\sigma^2 + b\sigma_{\tau\beta\delta}^2$
	$(\gamma\delta)_{kh}$	<i>r</i>	<i>a</i>	0	0	1	$\sigma^2 + aa\sigma_{\gamma\delta}^2 + \frac{ra \sum \sum (\gamma\delta)_{kh}^2}{(b-1)(c-1)}$
	$(\tau\gamma\delta)_{ijkh}$	1	<i>a</i>	0	0	1	$\sigma^2 + aa\sigma_{\tau\gamma\delta}^2$
	$(\beta\gamma\delta)_{ijkh}$	<i>r</i>	0	0	0	1	$\sigma^2 + \sigma_{\beta\gamma\delta}^2 + \frac{r \sum \sum \sum (\beta\gamma\delta)_{ijkh}^2}{(a-1)(b-1)(c-1)}$
	ϵ_{ijklh}	1	1	1	1	1	σ^2 (not estimable)

Table 14.25 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition)

However, we can use the traditional split-plot approach and extend it to the case of split-split-plot designs as well. Keep in mind, as mentioned earlier, we should pool all the interaction terms with the block factor into the error term used to test for significance of the effects, in each section of the design, **separately**.

14.5 - The Strip-Plot Designs

These designs are also called Split-Block Designs. In the case where there are only two factors, Factor *A* is applied to whole plots like the usual split-plot designs but factor *B* is also applied to **strips** which are actually a new set of whole plots orthogonal to the original plots used for factor *A*. Figure 14.11 from the 7th edition of the text is an example of strip-plot design where both of the factors have three levels.

**Figure 14.11 (Design and Analysis of Experiments, Douglas C. Montgomery, 7th Edition)**

The linear statistical model for this two factor design is:

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

Where, $(\tau\beta)_{ij}$, $(\tau\gamma)_{ik}$ and ε_{ijk} are the errors used to test Factor A, Factor B and interaction AB, respectively. Furthermore, Table 14.26 shows the analysis of variance assuming A and B to be fixed and blocks or replicates to be random.

Source of Variation	Sum of Squares	Degrees of Freedom	Expected Mean Square
Replicates (or Blocks)	$SS_{\text{Replicates}}$	$r - 1$	$\sigma_e^2 + ab\sigma_\tau^2$
A	SS_A	$a - 1$	$\sigma_e^2 + b\sigma_{\tau\beta}^2 + \frac{rb \sum \beta_j^2}{a - 1}$
Whole Plot Error _A	SS_{WP_A}	$(r - 1)(a - 1)$	$\sigma_e^2 + b\sigma_{\tau\beta}^2$
B	SS_B	$b - 1$	$\sigma_e^2 + a\sigma_{\tau\gamma}^2 + \frac{ra \sum \gamma_k^2}{b - 1}$
Whole Plot Error _B	SS_{WP_B}	$(r - 1)(b - 1)$	$\sigma_e^2 + a\sigma_{\tau\gamma}^2$
AB	SS_{AB}	$(a - 1)(b - 1)$	$\sigma_e^2 + \frac{r \sum (\tau\beta)_{jk}^2}{(a - 1)(b - 1)}$
Subplot Error	SS_{SP}	$(r - 1)(a - 1)(b - 1)$	σ_e^2
Total	SS_T	$rab - 1$	

Table 14.26 (Design and Analysis of Experiments, Douglas C. Montgomery, 8th Edition)

It is important to note that the split-block design has three sizes of experimental units where the units for effects of factor A and B are equal to whole plot of each factor and the experimental unit for interaction AB is a subplot which is the intersection of the two whole plots. This results into three different experimental errors which we discussed earlier.

Source URL: <https://onlinecourses.science.psu.edu/stat503/node/68>

Links:

[1] <https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson14/Example14-1.MPJ>

[2]

https://onlinecourses.science.psu.edu/stat503/sites/onlinecourses.science.psu.edu.stat503/files/lesson14/recognize_split_plot_experiment.pdf