

MITOCW | Lec 14 | MIT 2.830J Control of Manufacturing Processes, S08

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

DUANE

So what I'm going to do is pick up a little bit where we left off from last time. Last time we were talking, primarily, about full factorial models. And then we dealt with a few important additional issues in experimental design.

BONING:

So we were talking about issues of blocking against nuisance factors, kind of a practical issue, but that also got us into a generic issue that's actually much more fundamental and important of confounding. And I want to pick up on this issue of confounding today because very often you will want to do fewer experiments than a full factorial. That 2^k grows very fast-- grows exponentially fast with a number of factors, for example.

And so very often you might ask the question, can I reduce the number of experiments and still get the key information that I want to? And so that's where we'll really pick up with fractional factorial designs today and understanding confounding and aliasing patterns that come with different subsets of a full factorial design. Then we'll touch on some implications for model construction that fall out pretty naturally from that. And then start talking a little bit, hopefully if we have time, on process optimization using some of these kinds of design of experiments techniques and the models that we're building.

So as I was saying, very often we want to run fewer than that exponentially growing number of experiments, even if it's just to level, building simple linear models. Again, we've got a 2^k exponential growth. And, as an example, imagine we said we wanted to run less than the full 2^k , say, for three inputs-- so for three inputs, if we run the full 2^k . And we wanted to form a full linear regression model with interactions-- so it's still on quadratic, but it does have all the interactions.

This is what the model looks like. And it's got, if you count them up-- it's got eight coefficients. So if we were to do less than the full 2^3 or 8 experiments, we obviously would not have enough data points to uniquely fit every coefficient.

So that's already giving us the biggest clue. Is if you do less than a full factorial model, then even for linear experiments, you won't be able to fit all of the coefficients. So what you end up having to do is make a decision, a priori, that some of these factor effects are going to be small. That is to say, we're going to treat the coefficient as essentially zero.

But then what's happening with the data? What impact does that have on other coefficients? So we can explore that slightly.

Here's an example. We'll call this a $2^3 - 1$ experiment. And I'll talk about these half fractions a little bit more. But instead of the full 2^3 or 8 experiments, we'll do a half fraction here. We'll just do 4 experiments instead of the 8.

And so if I were to pick my factor levels on x_1 and x_2 and then think about what the-- so I'm just doing-- in my mind, I might think if I were just doing 2 factors, I would have a full factorial on to 2 factors. And I also could calculate what an interaction term would be. That's fine, that works well. I've got 4 coefficients and a 2 factor experiment. And that I could uniquely fit all of those coefficients.

However, what if I were to simply relabel that. And instead of thinking of that as an interaction, what if instead I labeled that an x_3 column? Then I could fit the linear term for an x_3 model.

So imagine I was doing a 3 factor, but only worried about main effects. I wasn't looking for interaction. I didn't care about the x_1 , x_2 interaction.

If I did that, and just defined x_3 as if it were that column, then I've got a 3 factor experiment, but I can only see main effects. But you see what's going on here, I still have that interaction term. I still have an x_1 x_2 interaction. So the key idea of confounding, that we saw before, is lurking in here.

If I were to do truly a third factor and set on my control knobs for that experiment, my x_3 according to these high-low settings, those would also give me the same information, if you will, the same combination settings for x_1 and x_2 , that I would have used to help me detect an the interaction between x_1 and x_2 . That is to say, the x_1 , x_2 interaction is confounded with a third factor, if I were to do it.

Just expanding on that a little bit, if I were-- this should be a hat. There's another little weird font thing going on. If I were doing a 3 factor experiment, truly a 3 factor, I could build up to a linear model OK with those 4 experimental parameters, but I could not fit any interaction terms.

So I have a choice in my model building. Do I take my three factors and look for just main effects? Or do I stick with my two factors and look for an interaction effect and not model at all-- pretend the third factor doesn't matter or try to keep that constant and not let that enter into it? Question--

AUDIENCE: If you computed this [INAUDIBLE] you still can't say for sure that it was x_3 or [x_1 on x_2]. So you describe it [INAUDIBLE] you really have no way of knowing, right?

DUANE BONING: So if-- so the question-- by the way, could we put the back screen also to the slides? Great, thanks. So the question is, we can still distinguish x_1 , x_2 , and x_3 or not?

AUDIENCE: No-- x_3 -- the effect that we ascribe to x_3 -- x_1 , x_2 . So we cannot differentiate between the two. Or [INAUDIBLE] might tell us we have a huge effect because of x_3 -- a significant effect. So we're not really sure if it's x_3 that's having an effect or an interaction of x_1 , x_2 .

DUANE BONING: Yeah, you guys may not have heard that, but basically, just had the restatement of exactly the issue. That it's confounding between is it an x_3 factor or is it a β_1 , β_2 , x_1 , x_2 factor. You cannot distinguish between the two.

AUDIENCE: So should we do something different to find that out?

DUANE BONING: Yeah, so the question is what would you do then? There's a priori knowledge that may guide you to a belief that there is no interaction effect. That there's no possible physical way there could be an interaction effect, in which you'd should be safe.

Or you might be trying to say, I'm not even sure either of these effects exists. And so I'm quite happy to do just four experiments. Do an ANOVA that tells me is either of those, just one or two, significant. In which case, now I worry-- maybe then I add experimental points to differentiate. So we've talked a little bit more about that.

But that's, in a nutshell, the confounding and aliasing thought process. And we'll look at some rules of thumb that would lead one to believe, for example, that I may have, in many cases, more of an a priori belief that maybe the main effect will be stronger and more important. And I doubt that the interaction effect will be there. But those are kind of rule of thumb and assumptions that if you can do the experiments, you can check.

So the point out of this was simply-- again, these should be models-- y hat models. For exactly the same table of settings, if I looked at it one way, I could build a model where I've got some interaction term. Or I can look at the main effect. Or I can give up-- so this could have been the x^3 main effect.

Or similarly, I could say, get confused between some other main effect and an interaction effect. In other words, I've got four coefficients, four experiments. I've got some confounding going on, but it's not clear actually.

I've given you the example for one of them in the previous two slides. But the same thing holds true for any of the other interactions with one of the main effects. So if I really have three factors, I might really have eight kinds-- in a linear model-- I might actually have eight terms. And what I've done is folded four of them onto or confounded them with four of the other coefficients.

And so then the question becomes what's confounded with what? How do I structure and how do I pick which subset of experiments to run? So this is, essentially, just saying the same thing.

The point is we can carefully pick which subset, in many cases-- this kind of a very small experimental design. But in many cases, we can pick which subset of the rows we want to take, which have a fraction we want to do, based on our belief in which interactions are going to be least likely or least important. So what we want to do is get a little bit of a feel for that.

So here's a little bit larger picture, as well as now extending some of our shortcut techniques, our terminology, our little factor algebra that we've got for defining rows and columns in our experiments. So if I were doing a three factor experiment, a full 2 to the 3 array, this is our x matrix. The identity column-- the C column-- you can sort of see I've-- all low, all high.

And then within that, I've got the B columns, low and high, low and high, low and high. And then within each of those, low-high, low-high, low-high, low-high. So that's, again, all 16 of the possible-- excuse me, all eight of the possible combinations.

And then we can construct our interaction terms. These are two factor interactions. And then we've got a three factor interaction as well. So that would be all eight of our model coefficients.

The column would tell me how to form the contrasts for detecting or estimating each of those interaction terms in the model. So that's just our baseline.

And now let's consider what happens-- let's think for this full 2 to the 3 experiment. If I were to only do half of the experiments-- let's say I did the upper half-- the top four, the shaded four experiments only, what happens in terms of what coefficients can we estimate? What ones can we not?

Please come on in. So which coefficients can I estimate and which ones can I not? And what's confounded with what?

We just got some visitors filing in. I think there is, perhaps, some bench space over there, as well, if you need some. So everyone in Singapore wave. Wave to our guests. Great.

So if we do just this upper half fraction here, let's look at a couple of things that are immediately obvious. One that's really obvious here-- look at the C column. If we do those set of four experiments, can you estimate what the C effect is going to be? No, we haven't excited that variable at all. It's all four experiments, we're at the low setting of that.

So it's clear, if I picked those four, I'm basically making an upfront decision that the C effect-- I'm not going to be able to model. I'm not going to be able to fit it. I'm not going to know if it's significant.

I'm not exercising that variable in a way with the right combination of experiments to unambiguously say, yes, there was a C effect. So one way that we can describe that in our funky DOE algebra is to say that the C column is equal to minus the identity column. It's just minus 1 times the identity column.

And we can also look at some of the other columns. And here's a neat one. Let's say-- let's look at this. The AC column, right here, and the A column.

And again, in our funky algebra, you can see already this AC equals minus A. But if you look in one to one correspondence, the AC column is exactly the same, just with a minus sign. The same combinations of levels as the-- both the A and the AC have the same combinations. So that's confounding or aliasing between those two columns. So even if I run this experiment, there's no way for me to differentiate whether it was a main effect, an A effect, or an AC interaction.

Now the other point here is with just this selection, again, of the top four columns, the same thing happens to other columns. What's the B column confounded with? No, we've got some-- somebody leaning against the wall trying to stay awake, I think. The disembodied voice in the machine said to tell you that that's happened before.

So the B column is also confounded with-- let's see, what? Minus 1-- this one. And then finally, the last one is the AB column is confounded with-- let's see, that's just minus 1 of the ABC, it looks like. Do I have that right? Yeah.

So that was our point here. If I had all eight experiments, I could have fit eight coefficients in the model. But four experiments-- I can only fit four coefficients. And it's more important than just fitting four coefficients. It's really that I folded together the effects of two of those columns into one coefficient that I have to assign either to-- sort of nebulously-- either to one or the other of the main effect, or the interaction effect, or to some superposition of the two.

So this is just, basically, saying the same thing. I'm just pointing out-- looking at the columns, you can see that. And if you were to actually follow through and use our contrast map for picking out and detecting what the contrast for effect A or the contrast for effect AC would be, they're essentially the same contrast. The same sum of the output rows for that particular effect.

Now, we have a shorthand way of describing what this confounding pattern is in our funky algebra. Which is to say, we equate what columns are equal to what columns. And then do a little bit of our algebra of multiplying a column factor or level setting by each other until we get down to the identity column on one side or the other. And then that becomes a shorthand way of describing what the whole confounding pattern is.

So for example, I can pick, actually, almost any of these confounding patterns. Looking at the AC in the A column, where I detect that the AC and the A-- this was the A column is equal to minus of the AC column. Now if I multiply both sides by the A column-- A times minus AC-- then on a element by element basis, if I take the A column there and multiply it by itself, every minus 1 multiplies by minus 1. And I get the identity.

Over here I've got A times a minus A. And that also becomes the identity. And I end up with just the I equals minus C. Which I could have also looked right directly at the column and said, OK, what is aliased with the identity? But also so you get a little familiar with that funky column math.

So that, basically, is a shorthand way of either describing the interaction effect or more usefully, if I knew up front what interactions I was willing to confound with what effect-- If at the start I said, I really believe that the BC interaction is going to be minor compared to some other effect, then I could use that to actually help guide which half of the experiment to pick.

So for example, I could have picked this half or I could have picked this half. And each one of those would have been consistent half fractions. And the choice of what to pick depends on-- for example, do I think AC is an alias with A, maybe, because AC really shouldn't be interacting with each other? I'm happy to do that. But the AB interaction, I might actually be wanting to detect that I'm willing for that to confound with the B factor because I think that that may be less important.

AUDIENCE: Just seems to be asking that A is a processes pressure and that C is the temperature-- for sure temperature. So in that case, the interaction between pressure and temperature, AC, if temperature is high and pressure is low with the same, the interaction would be the same as A? Because that A is equal to AC.

**DUANE
BONING:** Yes.

AUDIENCE: It will just change one parameter, just temperature [INAUDIBLE]

**DUANE
BONING:** But what it's saying is that your model, then, for what the A affect was, wasn't purely an A affect. It had a little bit of the interaction lurking in it. So when you actually run that experiment, if there is an interaction, you can't tell whether it was due to the main effect of pressure or also because of some interaction with temperature. So it's not that you get both or you only get one or the other, it's that they're mixed together. They're confounded together.

So I alluded to some of the ideas of how you choose which design, based on what confounding you're willing to live with. But there's also a few additional guidelines that are at work. One is, there's this idea of balance and orthogonality-- that I'll talk about in just a minute-- that tells us you can't just willy nilly pick random rows out of your matrix and be able to use our design experiments and that analytic techniques. Things like the estimation of the effects and so on only apply if I've got balanced and orthogonal experiments.

And that leads, also, a bit, to this idea or very closely related to the idea of getting enough excitation of the inputs. So going back to our 2 to the 3 full table here-- this idea of balance, first off, says that in whatever subset of the design that you've got for the factors that you're interested in, you want all of the columns to have an equal number of plus and minus signs.

Now, we saw if I did the upper fraction for C, if I was trying to deal with C, that's not balanced because I've got four low settings and zero high settings. I have not at all excited that input. But if you think back to all of our algebra for dealing with contrasts and being able to take the average of this, and the average of that, and subtract the two with the contrast, that's basically the reason that we need this idea of balance between the high and the low settings for that particular factor.

So to use our shorthand approaches, you need an equal number of the plus and minus signs in each of the columns that you are trying to use in your model. Turns out you can relax that a little bit if you do some regression approaches. There's some other risky, nasty things that happen like you may not have the same amount of variance or residual error at different points in your experiment. But especially for the shorthand approaches, you have to have this notion of balance.

The second idea is orthogonality, which is basically saying that what I need is for the sum of the product, element wise product of two columns, to sum up to 0. So for example, here, the column A and B-- this is a product of one- - product minus 1, product minus 1, product of 1. Those sum up together to be 0. The A and B columns are orthogonal. Which is another way of saying those two columns are not confounded with each other.

So think in a linear algebraic sense. If the two vectors are orthogonal, they are not mixing together effects in any way. So if I have two columns that I want to be able to model separately and not have the coefficient trying to mix in randomly some amount of one or the other, but, in fact, to be identified with that particular effect, those two columns have to be orthogonal. So for example, what that's telling me is if I were to pick this upper half fraction, the A and the B columns are orthogonal. They are not confounded with each other. They are two main effects that, at least with respect to each other, are separable.

You could ask that same question now to our other confounding patterns. What did we say? A is confounded with BC-- was it? So A and BC-- are those two orthogonal? Certainly, we know by confounding, they're not supposed to be.

And if you do that product of sums, each one of those, I believe-- AC, sorry. Good-- I was about to say, that's weird. They look orthogonal. There we go.

Each one-- the product is minus 1 every time they all sum up to minus 4. They're not orthogonal. They are mixing in together. So those are a couple of these ideas of balance and orthogonality that are in some sense the same-- just another terminology for talking about the same thing that we've already built up some intuition about. So I think we've already just went through that.

So if I were, in this experiment, doing a half fraction and picking the upper column, you can see things like A and B are balanced, C is not. So I can't try to put that into my model. A, B, and C are orthogonal. Let's see, is that right? Yes-- so if I were-- except I'm already knowing I better not try model C because I don't have sufficient excitation in there.

But I could also then ask, OK, what else is confounded with each other? And if there are places where I don't have orthogonality, I might then be led to is there a different half fraction that I might want to pick. One might be this lower one. I've already looked at the upper half. But couldn't there be other meaningful combinations? And the answer is yes.

So for example, here's a better subset if I were looking for a particular property that was annoying with the previous one. In particular, this fact that I can't even model. I'm not even exciting a main effect. That feels weird.

So a better half fraction, then, at least lets me look at main effects. Might be this shaded blue half fraction. Now look what's going on. All three columns A, B, and C are balanced-- just the shaded blue parts. So I've got equal numbers of high and low, each of A, B, and C are mutually orthogonal.

I still have confounding going on. I'm not sure quite exactly what this means. I guess A-- let's see. I still have confounding because, for example--

So A is not orthogonal with what column here? BC in this one. Because 1, 1 minus-- so these are the same. So these two columns are confounded with each other. So those are not mutually orthogonal.

And I could ask, OK, what is a shorthand way of describing this? I could pick a couple of these columns and say A equals BC and do, again, my funky algebra. Multiply both sides by A and get I equals ABC. And there's my defining relationship for what's confounded with what. Could also look up here, again, and ABC is exactly-- for the blue-- for that subset-- for the blue element is confounded with the orthogonality matrix.

What might be better about this design or this subset than the previous one? Depending on what you're looking for in the-- what's better about this? Here we have, I guess, shorthand for us-- what the other aliases are. Again, we expect to have four aliases because I'm only doing four experiments.

What do you see about these aliases?

AUDIENCE: [INAUDIBLE] high and low values for all the main effects and one only third order, like ABC, is the one that you have not excited at all, it's the one that [INAUDIBLE]

DUANE BONING: Good-- so that's at least two of the three properties that I really like about this design. If everybody didn't hear it, he said, at least you're exciting A, B, and C. Second, the only column that you're not exciting at all is a third order interaction, which is fine.

How likely is a third order interaction? We'll chat about that a little bit more. There's another additional effect. There's another additional really nice characteristic about looking at these aliasing patterns here.

AUDIENCE: Like, A is [INAUDIBLE] BC so they're not measuring exactly the same amount.

DUANE BONING: Right, they don't-- I don't know quite what-- don't overlap is a nice way of describing it. I've got a main effect that is confounded with an interaction of somebody else's second order term. So it kind of gets around that discomfort with A being compounded with AC that you were describing.

Now, I still have to worry about it because if any second order effects are out there, if AC is still active, now I'm not confusing it with A main effect. I am confusing though with the B main effect. But the nice thing is now I can-- you'll see it a little bit later-- I can also appeal to some physical causality to also talk about the likelihood of it being a main effect or an interaction effect, when I finally analyze my overall experiment.

What am I saying? Here's the peek forward at that. If I were to find that this-- either the B main effect or the AC interaction is at work, but I also did an ANOVA and I found out already that this was not significant, in other words, they're not an A main effect, it's highly unlikely that there's an interaction with A.

If it doesn't have an overall effect, why would it have a subtle interaction. It's kind of unlikely. So by ordering this, you actually can start to decode what's going on in the process with a little bit better visibility.

So what we're actually talking about here there is some additional terminology for which is design resolution. It's basically a characteristic of the aliasing patterns and how decoupled you are able to get between main effects and interaction effects or second order interactions with other interactions, and so on. And you will hear, sometimes, description of particular half fraction or other experimental designs, DOEs, as being resolution three, or resolution four, resolution five.

And a resolution three is a weaker ability to discern than a higher resolution in your experiment. But a resolution three is nice in that no main effect-- an A factor is alias with another main effect. So A is not alias with B. But you will have main factor and other interaction aliases.

Let's say you didn't want that. You might need to build a more powerful-- do more experimental combinations. And you could go to a resolution four, which is essentially designed as an experiment where no main effect is an alias with another main effect.

And I'm not even aliasing any main effect with its second order interaction. I might be so worried that there are second order interactions that I have to make sure that I do enough experimental points to be able to detect them separately. Do an ANOVA on significance of each of them separately.

The example that we did just a second ago, what resolution do you think that is? It's only resolution three. We were clearly seeing interaction with the second order effect.

If I wanted to get to resolution four, what would I need to do? It's the full factorial, in this case. I'd have to go to the full 2^3 . I couldn't pick a half fraction.

So if there were four factors, it turns out you get 16 columns. And now I can pick 16 rows. I can pick up 16 columns and rows. But I can pick, now, some subset if I did a half fraction and still achieve a resolution four if I wanted to.

So this is just looking back, again, at the column. Again, the main effects are aliased with interactions only in this defined experiment. And that, you will sometimes see referred to as a 2^{3-1} sub resolution three experimental design.

So now, if you see the shorthand notation there, you look and go, oh, that's a half fraction because it's a 2^{3-1} . And it's resolution three, telling you something about the aliasing.

AUDIENCE: In the previous chapter, is the difference between resolution 4 and 5 both have no [INAUDIBLE]

DUANE BONING: So there was a subtle point in here, which is no main effect interacts with a second order interaction. But I might have a second order interaction aliased with another second order interaction. The resolution five says, no second order interactions are aliased with any other second order interactions.

So here we've been talking about half fractions-- 2^{3-1} , especially as the number of factors gets large. Let's say I have a design-- I'm designing a process and it's got eight different control knobs. That's a 2^8 experiment. That's an awful lot of different combinations.

I might not want to just go down a 2 to the 8 minus 1. I might want to do even lower half fractions. For example, if I cut it in half, and then cut it in half again, that's a quarter fraction. I'm only picking a quarter of the full factorial. And again, there would be a resolution associated with that.

You could look and see what is the aliasing pattern with that. And as you get to the higher order models, you will often see increased half fractions in early parts of experimental design where you're just trying to see are the main factors important. And might there be some second order effects, some second order interactions, at least detecting whether they exist. And then if they do, you'll often then go back in and start filling in the other parts of the experiment.

Here's another half fraction. Trying to think-- this is, again, for our 2 to the 3 minus 1. This is a different set of four. It's just got a different defining relationship. And you could, again, ask, in this case, is that-- number one, is that a legitimate half fraction? Could I legitimately pull these four columns out-- four rows out?

What kind of balance do I have? What kind of orthogonality? Or what's aliased with what? And you can either do it by looking at-- you can detect or decide what are the aliases by looking at the columns or simply doing your little math down here from the defining relationship.

So if I start with I equals AC and say, OK, what's aliased with C? So I do C times I equals CA C equals A because the C times C is the identity, and so on. And similarly, I could say, OK, if I start with I equals AC multiplied by B, I get B equals ABC. There's that.

So there's a main effect aliased with the third order. That's not too bad. I don't really like that A aliasing with C. That's not even a resolution three experimental pattern. So I'd be very careful in using this unless I really believed, strangely, that C was not going to have a main effect but I might be worried about it or interested in it for the purposes of looking for interactions.

There is one other subtle place where you might actually-- nope, never mind. I've already alluded to a couple of ways of deciding what aliasing pattern to choose. The most important one is your knowledge of the process-- your experience with the process. What factors are likely to actually, based on physical causality, interact.

But there are a few very important rules of thumb that are worth mentioning here. They are this idea of sparsity effects, hierarchy of effects, and inheritance. And we've already talked qualitatively about them. But just to nail them down-- if I have eight factors, the experimenter, you, likely has a certain amount of a priori knowledge that there's an ordering and likelihood of those effects.

But it's also the case, I have different eight factors on my experiment, it's highly likely that they don't all equally influence the process. Most processes have a top few factors that have most of the influence or most of the effective. It's like a Pareto type of a rule.

So early on in screening, in fact, you might be very happy doing one of these half fraction, quarter fraction, and so on. The purpose of which is really just to narrow down a really large number of candidate effects down to a smaller number that you're then going to model more accurately and look for interactions among those large effects. So this sparsity of effects is one of the things that's at work in early screening experiments and is a good rule of thumb a rule or a generic effect to be able to take advantage of.

The second one is this notion of hierarchy, which is basically-- again, these main effects are usually more dominant than second order effects, which are more dominant than third order interactions. Furthermore, usually you have to-- not have to-- but usually, you will see that the main effect or the lower order interaction has to be at work before you have substantial interactions with some other factor.

So it's going to be rare for you to have a big AB effect and no main effect with A and no main effect with B. So that's another rule that you can often use. And the likelihood of having these very high order interactions-- the idea that you will have an extra delta in an eight factor experiment attributable to exactly the combination of all of those settings being important or large enough to be important is pretty small.

And then, I guess, I've actually mixed in the inheritance here already. This idea that I really need both of these factors. So the hierarchy is really just more saying the size of the effect, schematically pictured by the size of these boxes. The main effect tends to be larger. And then the inheritance is I kind of need the lower order interactions be at work.

So here's just some additional examples in the context of half fractions. We've already almost exhaustively explored the 2 to the 3 minus 1 resolution three kind of picture. Here's an example defining relationship for resolution four. And here, you really have to get up to four factors before having enough leeway to be able to have that kind of an interaction pattern where I would only have-- so, for example, now if I did the algebra, A would be confounded with BCD. Or if I multiplied AB on one side, AB is equal to CD.

So that was the earlier question. There you can see, a two-factor effect or two way interaction is confounded with another two way interaction. But all main effects are only compounded with three way interactions and none with each other. And here would be an example you can play with to see what the aliasing pattern would be with a resolution five kind of experiment.

AUDIENCE: [INAUDIBLE] the 4 minus 2 design-- is that-- would that have-- what's the identity?

DUANE BONING: I think that-- the identity-- I don't remember offhand. I think we've got, actually, an example right here that shows what the aliasing pattern would be. You asked exactly the question. And so we can work it out.

So here's an example. First off, we already know it's probably not going to be a defining relationship like this one because I've only got 2 to the 4 minus 2, which is 2 square. I've only got four experiments out of the 16 that I'm picking.

So I've got an awful lot of confounding going on. And the question would be, I can actually build it and look and see what the confounding pattern would be. And in particular, I might want to pick it such a way that I want to at least be able to detect the four main effects and lots of other things could be aliased in with that.

So I could build it a priori and say, OK, I'm willing for A to be aliased with that and with that. So in this case, when you've got a double half fraction I'm going to have aliasing be between more than two columns. Now I actually have multi-way aliasing between three columns.

So in essence, what I've got is something like A equals BC, D equals ABC. And then you can work that out as I is equal to ABCD. In this case, is equal to BC.

So it still has this defining relationship, but I'm now only picking four of the rows. So I don't know if you guys can see that. It's kind of tiny. But this is all 16 of these columns. And, again, this is the same combination that I just went through before.

If you look, the A column here-- these four-- minus 1, minus 1, 1, 1 is now aliased with both the BCD-- where'd that go-- the BCD minus 1, minus 1, 1, 1, and the ABC column, minus 1, minus 1, 1, 1. So if I do that, I have folded two other columns onto it in order to get down to the full half fraction.

And you can also explode out in the other direction and see what are all of the interactions and aliases that go on, again, either looking at the columns-- and here, I've shaded the columns that are alias with each other. So in each case, we've got all four. Or you can do the column math. And this has another nasty effect, which is those four columns are not even resolution three.

So going back to your earlier question, I don't know what the defining relationship is for a 2 to the 4 minus 2 resolution three. I'm not sure what the defining relationship is. Sounds like a good problem set problem. Remember that one.

So I think we've explored aliasing. Do you understand aliasing? Any questions on the aliasing? Confounding?

There is one more aspect to it that I want to explore a little bit, which is, what are the implications for model construction? We've already alluded to this. So let's just sort of work through it.

But also, folding in and remembering interaction terms, but also, potential higher order terms and some implications that arise if-- some of my factors, I think, might have quadratic elements to it. You get, actually, then complicated aliasing patterns in those cases.

So a simple case, when I just got one input and I've got one output, but I think there might be a quadratic effect, what we're seeing here is that I cannot do just a two level full factorial to exercise that. As we talked about last time, I have to add some kind of a center point or some other-- I really need all three data points in order to be able to fit a quadratic term. And we talked last time about being able to do ANOVA residual analysis to differentiate whether that deviation compared to spread within a replication error is significant. So I could decide whether that coefficient is significant or not.

If I were generalizing this now, to more than one factor-- last time we talked about a two factor example. And we had up to second order terms. But if we expand this out to a full quadratic with all of the interactions, what you see is a very rapid explosion in the number of coefficients. Because I've got my main effects-- so this is still just two factors, but now three level-- a full factorial in three levels.

I've got my average. I've got my main effect. I've got my interaction between x_1 and x_2 . But then I've got, also, x_1 squared and my x_2 squared. And then I've got the interaction between x_1 squared and x_2 , x_1 and x_2 squared.

And if both of the terms were squared, the question is, do I have-- this is what the full quadratic model with all of the interactions would be. Do I have enough data to be able to fit this if I did a three squared problem? Why not? How many coefficients do we have?

AUDIENCE: You have one extra.

**DUANE
BONING:**

Do I have one extra? One, two, three-- I have nine coefficients. How many experiments is three squared? Nine experiments-- so it's exactly like what we saw with the two experiments. I have exactly, just barely, the number to perfectly fit.

So actually, if I do the regression formulation, which I think is coming-- ew, nasty-- think these are supposed to be vertical dot, dot, dots. And these are supposed to be horizontal dot, dot, dots. Weird font problem here.

If I were to actually formulate this for the three factor, I would have exactly the same matrix relationship. And I have exactly the same number of rows and columns. And I would be able to fit that-- I guess that doesn't come till later. I could solve that directly and get beta is equal to x minus 1-- my output.

But I don't have any replicate data. If I had replicate data, then I would also need to do the pseudo inverse exactly as before. So the point the point here is if you want to build every possible model in the full quadratic, I have to have a full factorial in three levels, as well. So we can meaningfully talk-- although you don't see it that much in the literature and we'll see why in a moment, you can talk about full factorial more than two levels. I'm talking about full factorial 3 to the k , where you've got three levels per test.

But what you will see and I want to touch on a little bit, are slightly different designs that have a couple of additional properties that might be a better way to go than starting a priori with the full tree level factorial, doing every possible combination of all three levels. But rather, start with the two level experiment. And then if you start to see some important interactions or some indication that additional effects are needed, adding experimental design points incrementally when it's easy to do that. It's not always easy to do that in your experimental setting.

AUDIENCE: How would you detect that? Would you say other-- observe other effects?

**DUANE
BONING:** Yes, so if I do purely the corner model, is it possible for me to detect if there might be curvature? No. So the first thing I would do, if I wanted to detect, is at least add some center points.

And certainly, for continuous parameters we talked about last time, that makes sense. It doesn't make sense for discrete parameters. Really, curvature is only a term that makes sense with a continuous parameters. So that's kind of the domain I'm talking in here.

So in fact, my rule of thumb is I always have at least the center points in my original design. I never do just a pure two level factorial. I always add at least the center points because they tell me-- and I try to replicate at least the center points so I can distinguish between curvature and I can also, then, not fit exactly perfectly every interaction, but I can also start to ask questions about the significance of these effects.

Then I've got some indication that there might be curvature. Now I might go in and start to say, OK, there's curvature but I don't really know the nature of it. Now I want to add these three level points on more than one of the experimental factors. That's where I might then add even more experimental points.

But I'm always shocked because a few of the books-- you read Montgomery, [INAUDIBLE], you read most of the experimental design books. They rarely talk and emphasize the value of the center point. It's just absolutely crucial in my mind. I always want to add at least that off corner-- one off corner point. And preferably some replicates of that because it gives you so much more power.

And think about it. If I'm doing four factors, I'm only adding one more experimental combination. I'm not exploding out the whole design. It's a very cheap way to learn an awful lot more about your experiment.

So this is simply making the point. We just looked at the 3 to the 2 case-- how many coefficients there are in the full model. In particular, if we do the full model quadratic 3 to the 2, we already got nine coefficients. But if I add just one more factor and I'm still worried about the full model with all of the interactions, the number of experiments that I need to do explodes rapidly. Up here with only five factors, I got 243 model terms.

Do you really think each of those model terms is going to be nonzero significant? Probably not-- so this is also a wonderful opportunity for saying, sparsity of effects, hierarchy, all of those-- I'm going to alias some of those in and discount certain interactions. And in fact, if I just do main effects and the third order term, but only on the single effect. So an x_1 squared-- second order term, quadratic model, x_1 squared, and x_2 squared, and so on-- then it only grows linearly with the number of factors. If I know a priori, I can neglect those higher order interactions.

So this is just working out and giving an example, now using some of our earlier terminology. Again, here I can refer to the different combinations. Again, I have my A and B. And I can label the AB interaction. That's that AB interaction or the AB effect goes with the Beta 1, 2.

This A^2 is just an A squared, B squared, A squared B , B squared A , A square B squared. And you can, again, see, now, the factor levels where I've added 0 to indicate the center level setting for each of those. But we can also borrow, and import, and use all of the same aliasing terminology that we had earlier.

So for example, if I only did-- I don't know a-- well, if I did a 3 to the 2 minus 1, I guess that's a half fraction, but kind of an odd one because that means I get to pick three rows. I could pick-- is that right? I think that is.

You'd have a lot of aliasing going on. So actually, it's not as easy to talk about a half fraction. And in fact, we'll see-- I can do these aliasing, but I want to leap ahead. Here we go.

A good way to try to visualize these, which works for lower number of factors because it's not too high dimensional a space, is actually plot out in your x_1 versus x_2 experimental space what design points you're actually exercising with some half fraction. So for example, in this case, I'm just doing say, these $2/3$ and leaving off this $1/3$ of the experiment. In which case, what I'm essentially doing is giving up and saying, I'm not going to do the high level setting on my x_2 factor.

And now you can start to get a good feel, nice intuitive feel that goes together with the mathematics, of what you're giving up when you do that. If I were not doing those experiments, I just picked this subset, what model coefficients am I detecting or am I going to be able to fit, and what am I not in that experiment? Should be pretty intuitive. What model coefficient would I not be able to fit? Next to what?

AUDIENCE: [INAUDIBLE]

DUANE BONING: Yeah, for this one. No worries-- I said, we are going to pick these six points. These are the six columns going with these rows. But I'm not doing those three experiments in my x_1 and x_2 factor [INAUDIBLE].

Can I do a quadratic model in x_1 ? Sure, I got three data points projected along in that. Can I do a quadratic model in x_2 ? No, I'm only exercising two different levels. So I can only go up to linear in that term.

And that's what I mean by intuitive. I think you can start to see what, at least main order effects as well as second order terms, are at work. It's a little more subtle to see, can I do an x_1 and x_2 or an x_1 squared and an x_2 ? I think you can see, I've got combinations to be able to do some of those, but there may be some confounding going on. And then you can look back at the columns to see what kind of confounding may be occurring.

So I haven't actually done that on this to figure out which is confounded with what. But let's see-- anybody-- it's probably so tiny you don't have a hope in the world of seeing it. There is one.

B is equal to minus B squared, in this case, where B was my x_2 . So this is basically confounding and saying-- I was telling you, I could not fit the B squared-- the quadratic term in x_2 . If there actually is curvature, where did it go? It's hidden inside of the B^2 linear term. It's confounded with the B linear term. So it's the same kind of terminology.

Here's a different pattern that looks almost the same as the other one. I can still fit quadratic in x_1 , linear in x_2 -- in fact, it's not even clear to me right up front what I can't fit with one or the other. I kind of like having the center point in the other design.

But now, what's the difference between these two? Actually, I think a lot of the aliasing is relatively similar. I can think of one reason why I might pick the lower design over the upper design.

AUDIENCE: Wouldn't it make the effect more pronounced?

DUANE BONING: Yeah, I think that in combination with what I was thinking, but I think you're right. The statement was, this would make the effective more pronounced. I'm also thinking it explores a larger space of x_2 . This one is just zeroing in or zooming in on the 0 to low setting. The other one goes all the way and spans the low to the high setting.

So I'm exploring or fitting around the larger portion of the space. So I might prefer, for that reason, the lower one. It doesn't have a pure center point, but it's not clear that the other one does either.

AUDIENCE: Is there [INAUDIBLE]

DUANE BONING: Yeah, so the same idea is a balance and orthogonality are at work. But again, it's mostly with respect to-- it's more subtle now because it's with respect to either the linear term or the quadratic term. So these are both balanced and orthogonal with respect to things like the main order effects.

But not orthogonal B to B squared in these two cases. So you can use the same reasoning. But you've got three levels, you've got to expand it out and think it's not just the whole variable or the interaction. It's also the second order term, possibly, being nonorthogonal with a lower order term.

Here's another one. What do you think about that one compared to the previous two? This one looks interesting to me.

AUDIENCE: It's going to be bigger because we can do quadratic [INAUDIBLE] and since the combinations are less-- they have less effect [INAUDIBLE].

**DUANE
BONING:**

Yep-- I'm not sure everybody heard that. But at least here, we're exercising all three levels of x_1 , all three levels of x_2 . So we can fit quadratic terms in each of the cases. Where we might need to be a little careful is we're actually making an interesting trade off here-- going back, also, to your earlier point.

We're giving up a little bit of balance in this design. In that for the low setting of x_1 , I've actually got more data than for the high setting of x_1 . And so if there's noise effects, I'm actually fitting, in effect-- if you do the regression, you'll have a narrower confidence interval over here at the low setting than you would at the high setting.

So you actually have to be a little careful. You can do the regression math. You have to be careful in forming your contrast.

Normally, you get to these high r models, you're probably throwing it into a regression anyway. But you also then have to be careful in the interpolation and use of the model in different parts of the space because its accuracy is a little bit different in different parts of this space.

But I like this-- I kind of like this one is as well, even with those caveats. But you can still use this and the same aliasing terminology to figure out which coefficients I'm giving up. What's aliasing with what. And I'm not going to go through that, but you can do that.

AUDIENCE:

Would you get a higher [INAUDIBLE]?

**DUANE
BONING:**

Well, it's not clear you would get an overall higher or lower r squared. You almost always-- and we'll do a little bit more regression later. But you almost always, if you add more terms, you get a better r squared. But then the question is, is it a fair model? So we can also talk about an adjusted r squared where you penalize for the additional model-- the additional model in terms.

And then this is just pointing out that you can still use the linear algebraic-- either direct solution or quasi inverse solution if you've got replicates to be able to fit the model. So these are the sorts of things that come out. This is for an x_1 , x_2 . I don't know which is which-- x_1 , x_2 .

If, in fact, there are true quadratic terms, these are a little different kinds of surfaces than the ruled surface, which looked like it had kind of a funky kind of curvature because of an x_1 , x_2 interaction. But if I projected down on any one variable, it's always linear. Here, if I were to do a slice holding x_1 constant, I do, in fact, get a true quadratic.

Now, the nice thing about quadratic surfaces is you can start to think about an optimal point much more easily within these. Certainly, if the space is large enough to cover a minimum or a maximum, there's a natural motion if you're trying to minimize or maximize your output y of finding the optimum space. Now, it's also possible that if I had a smaller space in the true minimum or maximum of the full equation where outside, then I might have-- so for example, let's say my space was constrained to only right here, then I might run into the minimum or the maximum at one of my boundaries.

But now we've got this extra notion that my min or maximum might occur somewhere in the interior of the space. As opposed to with linear models, it's always at one or the other of the boundaries. So this starts to get to the desire for using this kind of a model now to find the optimum point, which is one of the main reasons we do experimental designs. Not just to find out which factors matter, but build the model, and then use it either in, maybe, feedback control or more often, just to set-- find the process settings or find the design optimal point in order to achieve some criteria.

Now, I alluded already to adding additional points. If we did a full factorial three level-- that's all nine combinations of high-low. And in the x_1 , x_2 space, that's all nine points. There is an alternative approach and probably one of the most important experimental designs after two level full factorial. It's referred to as a central composite design.

And you often will design up front to do this, but also very often it will be what you extend a full factorial corner point design with. If I did my first experiment with 2 to the 2, just pure full factorial two levels, and I got four tests, first off, this is probably wrong. The model is shown to have a poor fit.

If I actually did this for all four points in the interaction, I don't have enough to detect that. So somewhere in here would be ad center points. The power of the center point. If you remember nothing else from today, the power of the center point. The power of the center point.

Then I can start to detect whether it actually has a lack of fit in a formal sense. And let's say then I decide I want to go quadratic, but I'm not quite sure of the shape where my min or maximum might be. What we often do is add our additional points.

So there was our original four corner points. Certainly, we wanted to add the one at the center. Should have done that already.

But now I can also decide where to add my interaction points different-- so let's say I added this. The full typical 3 to the 3 would say, OK, you add them at exactly these 3 by 3 grid array. I would add them at the center points of each of these.

We can do something else that's a little bit clever. Which is instead, add these interaction points off the grid, but at the location of an outer circle equidistant or circumscribed around the original points of my cube-- my hypercube. See I think-- there we go.

So what we would do in that case here, is these would be my first four points. I'd add my center point. And now I add these points off axis.

There's something really clever, really valuable in doing this-- some things that are subtle, and kind of mathematical, and not all that important, but some other parts that are really nice and intuitive. I'll give you the subtle mathematical part. This actually maintains a little bit of fitting-- if you go and you do the regression, the fact that all of your off center points are now equidistant from the center point means that your model variance, no matter which direction you go, is the same.

In other words, there's a confidence interval on your coefficients and an interpolation or extrapolation error that grows as you go further from the center of your design. And by doing this, that maintains asymmetry. It doesn't matter which direction you look. You've got the same model accuracy purely as a function of distance from the center. So that's a nice mathematical property.

But there's another more intuitive value picking these corner points. And the way I think of it is if I just picked those 3 by 3 on a regular grid, I could then project down to x_1 or x_2 . And then in terms of exercising the x_1 variable, I've only got three levels.

Here, if I were to project down my points-- so let me get rid of some of this scribble here. If I were to project down onto just my x_1 axis-- let's say I did the experiment and I found that x_2 actually was not that important and I project down all of these points. That means for my x_1 , I've actually exercised at five different levels of x_1 .

I've got more than just-- I've got extra redundancy built in. If I'm fitting just a quadratic model, I've got ways to actually check whether the quadratic model is sufficient. I can do a lack of fit test on even the quadratic model.

Whereas if I only had exactly three samples of x_1 and I add the quadratic in, there's no way for me to ask the question, might there be an even higher order, a logarithmic dependence, or something else that's subtle that's going on. Because I fit exactly all of the data that I have. So the central composite design, I think, has this nice intuitive feel of I'm actually kind of exercising my space a little more thoroughly to be able to build a model that can apply a little bit more broadly. Everybody see that?

So here's what the central composite would look like. What you end up with is the original corner points. You add the center point. And now you can see, if you were to do the geometry, the distance there is square root of 2.

So I'm going a full one on both the x_1 and the x_2 . And you can extrapolate that to a third order hypercube. I think it ends up being, instead of square root of 2, that distance is a square root of 3 and so on. So you can find what the distance is and pick those different corner points.

So what we'll do next time is build on a little bit this idea of using the model. We built it. We can assess it. We can pick our points to build different orders of model.

What we'd like to do is talk a little bit about picking one of these surfaces and how I actually-- a couple of related, but slightly different ways of looking at optimizing to find an optimum point either after you've built the whole model or-- the clever thing here when you're on a process line-- doing it interactively or incrementally. Actually, picking your design points almost one at a time based on your current view of the model driving towards the optimum. So instead of doing all your experiments up front, you might actually want to do them in an evolutionary way to try to find the optimum. So we'll talk about both of those approaches next time.

So again, the problem set is due tomorrow. I think we'll hold off and we'll give ourselves, Hayden and me, an extension on issuing the new problem set so you're not doubled up, and we can think about it a little bit more as well. And issue that tomorrow, as well. We'll see you then on Tuesday.