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 BONING: OK, so on Tuesday, Hayden brought us up through analysis of variance, which had an implicit model of the process hidden underneath. It was mostly focused on the idea of, are there effects if you make different kinds of conditions? What we want to do is pick up on that, but get back to the model a little bit more, talk a little bit more explicitly about ways to express, from an empirical point of view, experimental data or other kinds of run data that we might have.

So I'm going to pick up a little bit on this idea of modeling effects when we've got multiple inputs. Much of the ANOVA analysis was described for a single input. And then what we will do is look at linear and quadratic models, and other kinds of models for fitting of that kind of data. We'll talk a little bit about calculation or fitting of model coefficients, both from a fairly powerful and general regression idea, but also, we'll start to build up the lingo for design of experiments.

One advantage of a design of experiments approach is that there are shorthand ways for rapidly assessing and, in fact, building models based on experimental data points in the selection-- careful selection of where you sample from your process or your conditions. So things like contrast will be perhaps a new terminology for you that summarize a little bit what it is we're trying to do with some of these design of experiments, or DOE.

So to pick up a little bit, here's, again, the injection molding data. We've looked at this from a number of different perspectives, and we'll look at it a little bit more today. So we've got some dimension that we're measuring with an injection molding, where we have a couple of different parameters-- one of which might be the velocity that we're using for the injection.

Another may be the whole time. Perhaps these are minutes-- one minute, two minute. And one would ask, is there an effect from a point of view of ANOVA? And then we can look and compare things like the spread within that particular condition, say, at one minute-- compare this spread and its mean to this spread and its mean, and make some assessment of whether there is a statistically significant effect.

But we might also want to build up a causal relationship, and actually a functional relationship between hold time and dimension. And then the question becomes, well, what do we use to fit such a set of data with? Would we fit it with a line? Would we fit it with a curve that looks something like that, a curve something like that? What is appropriate? What are appropriate kinds of functions to use for fitting of these things?

Now, hopefully you've got some intuition. And I'm not going to try to destroy any of that intuition today. I'm just going to try to extend it. One question for you-- hopefully you have an intuitive feel-- would you be justified in fitting this blue curve, in most cases, to this data? Well, I've now made it a red curve. Yeah?

AUDIENCE: Well, [INAUDIBLE].

DUANE BONING: Yeah.

AUDIENCE: So [INAUDIBLE] a physical model, or something, [INAUDIBLE]

DUANE BONING: Good-- because usually the answer is it depends, and you already got the depends in there. If all you had was the pure data, certainly, you've only got really two sample points, so all you can fit is a line. But your intuition is very good there. If you have a lot of physical knowledge that, in fact, the relationship would have to be parabolic, two data points might give you enough information to be able to justifiably fit some more complicated function.

We're going to be mostly focusing on empirical data and just sort of the limits of what you can infer from the data although. We're not going to be, at least today, trying to mix in both the physical functional dependencies and empirical. So one key question then is, what functional form might be justified by the data? In this case, if I've only got two data points, maybe it's only a very simple linear model.

If I had more data, well, I might have enough to fit parabolic or other, more complicated data. But in addition now, when we start to go to two inputs-- this is this is, in fact, the same data, where we have not only hold time, as we did on the previous screen, but we may also be playing with the velocity and then doing some number of replicates-- something like 14 or 15 runs at the low condition for velocity and low condition for hold time-- and then measure the data.

And what we have here is conditions where we've changed both of those conditions, and it may well be the case that these two inputs interact-- that the effective velocity gets amplified when I've got a longer hold time than a shorter hold time. So it's not necessarily just an additive model. So we also need to mix in this issue of interaction, when we start thinking about model form-- not just sort of the functional dependence terms of powers or for the shape of the dependence on a single input, but also when we have multiple inputs, if there are interactions between them.

So we can also turn this question around. I can say, depending on how much data I have, what kind of model might I fit? Can turn it around and say, I would like to be able to fit a model up to a certain kind of form-- you can turn it around and ask the question, what data points do I need? How much data do I need, and what's the location of those data points in order to be able to fit different functional forms?

And that becomes the design of the experiment in order to be able to meet your modeling or estimation needs. So what we'll be doing is dealing with some models of the process, where we have multiple inputs or multiple treatments-- not just different conditions or levels at these, but in fact, we may have multiple different kinds of treatments, such as the velocity and the hold time.

And so in general, we might talk about processes that have k input. So in our previous case, we had two inputs. And then we can look at different combinations of the data, depending on how many levels we pick for each of those inputs. So a level is simply where you're sampling along that axis of that particular factor or input.

And we'll be talking especially about full factorial models, where we're looking at every combination of the discrete selection of those levels for those factors that we're interested in. So if I've got a two level selection or choice for each of our factor levels, then, in general, we can have 2 to the k combinations of different discrete sample points, different discrete treatment sets.

You can extend this, of course. You might be looking at tree level experiments for each of our factor, and then I would have 3 to the k combinations. And you'll detect a trend here that's not especially encouraging. What happens as k gets big? We have exponential growth in the number of experimental points that we need.

Similarly, as the number of levels go up, you can fit more powerful models, but again, you need a lot more data fairly rapidly. Today we'll be just pretty much focusing on relatively simple small numbers of k and small numbers of level. In fact, traditional 2 to the k design is what we'll be talking the most about.

When you get back from spring break, we'll also come back to clever approaches out of design of experiments for subsampling the space-- basically, not having to explore all of the combinations because, in fact, all of the combinations may give you more data and fit a more powerful model than you're even really interested in.

If I were doing, say, a 3 to the k, I might be able to actually estimate all kinds of very subtle three-factor interactions or four-factor interactions, but I might actually not believe they exist. I may had prior knowledge they don't exist. And so we'll look at reduced models next time.

There is kind of just a general question. Why would you even want to model or use more than one input? Why wouldn't you want to simplify your process, make it as simple as possible? Why might more than one input be especially important for manufacturing process control? These are fairly obvious, so don't overthink too much. Why would you need more than one input?

AUDIENCE: Because you have more than one input in the world--

DUANE BONING: You have more than one input in the real world. Why do you have more than one input in the real world?

AUDIENCE: [INAUDIBLE] complex. You can manipulate them in a number of ways, so [INAUDIBLE] process might have temperature and pressure, different chemical baths you could use, [INAUDIBLE] processing.

DUANE BONING: Good.

AUDIENCE: So they all create slightly different outputs, but [INAUDIBLE] very similar outputs at the same time.

DUANE BONING: Yes. OK, so I think you've touched on both of them there, as a freeform spoke. One is you often have more than one output. So it would actually be odd if you were trying to achieve just one parameter-- say, a length parameter-- with a dozen different inputs all at once. In fact, that would be a really challenging design problem.

Often, they do interact, but very often, what you're actually after are fairly good solid one-to-one correspondences where, if I play with this knob, I have good control over this output. So you often strive for almost univariate, very simple direct control. You can't always achieve that, and so often you've got a multiple input and multiple output problem with interactions.

But even if you only had one output, you might actually want two inputs, because one of them may actually give you control over the mean of the parameter and another input might let you deal with, or compensate for, or generate, or change the variance of the process. And so it's actually very nice to be able to deal with the process robustness, to be able to freely recenter the process without changing your variance, for example.

So just highlighting that the practical world-- you often do have more than one input, and there are good reasons where you can actually use those. So we're going to mostly be dealing with empirical models here, and here's an example of a fairly simple general linear model. Question? Is there a question from Singapore? Somebody maybe pushed their button? [INAUDIBLE] like the microphone was-- we got a close-up view of you. We know you're awake, though-- good.

OK, so here's just a quick overview of the kind of models that we'll be looking at. We've got some output. Here it's expressed in a univariate case. We'll extend this to multiple outputs in a minute. But I might have a single output, and the true value here may be eta. Again, I'm using a little bit the Greek to indicate truth, if you will.

And it's got some overall mean. We might model that with a single coefficient, some beta 0. And then it's got single factor direct dependencies on some x sub i input. So x sub i here is indicating which factor it is. So I might have x1 and x2 would be a two-factor-- two different kinds of input, sort of like our injection molding problem. I might have a direct linear dependence and coefficient on each of those.

And then we'll also be looking at interaction terms, ways of looking at a synergistic effect between the two of these terms-- still first-order. These are only direct x1, x2 kinds of interactions, but we'll see that actually gives you quite a bit of power for dealing with some interesting behavior. Then we'll also talk about are these higher order terms.

So in truth, the model may have a more complicated functional dependence. What we're hoping is, in many cases, that the linear version of the model captures most of the behavior, and that these higher order model terms are smaller effects. And so often, we'll actually be ignoring these higher order terms, absorbing whatever errors there are as a result of that, and referring to those as a model form or model structure error, and distinguish that from residual error or other kinds of experimental error, or just replication error.

So let's see-- couple other definitions just that we've got in here-- in here, you'll notice the k subscript indicates the number of factors, just like we were describing before-- two-factor, three-factor kinds of experiments. And then we often are forming sums over some particular-- either factor or input levels within those factors. So I'll try to be a little bit careful on what we've got then.

So with that simple two-order or two input model, where k is equal to 2 in this case, we've got something that we might try to use with our injection molding case. That was a two input model-- injection velocity and hold time. We've got four coefficients. Quick question for you-- how many data points do you think you need at least a minimum of in order to be able to fit those four coefficients?

If I did just one experiment and got one output, would that be enough data to estimate four coefficients? No. Four? Or was that just a wave of your hand?

AUDIENCE: No.

DUANE BONING: It was close.

AUDIENCE: Yeah.

DUANE BONING: So this is a simple extension of that linear case. If you had only one input and you wanted to fit a line, you would need at least two data points. Yeah?

AUDIENCE: [INAUDIBLE]

DUANE BONING: Oh, OK.

AUDIENCE: Because that's a zero--

DUANE BONING: Yes.

AUDIENCE: [INAUDIBLE] relative to your [INAUDIBLE]?

DUANE BONING: Yes. That's a good point. So if these are normalized in some way, such that there is a 0 in between, such that also the average of these are 0, then this would be the mean. So this would be more general-- some offset term. It may not necessarily be the mean-- good point, good point.

OK, so part of the question is I think there's-- it's fairly clear intuition that you're probably going to need at least four data points in order to be able to fit a four-coefficient model. We need at least that number, but there are some additional subtleties. In this case, we've got at least two factors, so we know we've got two factors at work here.

And I've got to at least sample across both of those input factors of both of those treatment things. But then there's a question, especially with this crazy interaction term-- how many levels do you really need in each case? Notice also here that, if I only have four data points, there's no way for me to really estimate higher order terms.

And we'll also chat a little bit, because if I have exactly four data points, it's also not easy for me to estimator terms, because if I have four data points, I can exactly fit-- I get a direct solution for the four coefficients. And I can perfectly fit those-- I may be wrong in terms of actually interpolating between those, but I have a direct solution for those four coefficients.

So part of the thing we'll get at is, how might you select levels and data samples so that you have more data to be able to do additional inferences on your model-- things like, what is the actual confidence level I have that an effect is real? What's my estimation not just on a point estimate for a model coefficient-- a beta 0 here or a beta 1-- but also, what is an integral estimate for those kinds of coefficients? So we'll get to many of those things.

OK, here's back to our injection molding case-- dimension versus old time. I might try to fit that with a very simple one input linear model, some offset beta 0 and some beta 1 linear dependence. And we've already said it looks like I'm going to need at least two levels. I'm going to label those with an x underscore minus and an x sub plus-- so a high level and a low level.

We'll actually talk about different labeling strategies for these things that make some of the calculations a little bit easier. So if I, again, do just one trial at each level, what I've got-- there's a little bit more notation for us in very simple vector notation-- what I've got is one factor, two levels, if I'm just doing a one-input model.

So what I will have is two measurements, an eta 1 and an eta 2. Those are my two actual sample points. And what I've done is run them at two different conditions. So in fact, these indicate the two different conditions that I ran at. And what I'm trying to do is fit the beta 0 and the beta 1 for this case.

Now, because I've got exactly the right amount of data, I'm going to have to estimate that the noise is 0, as we'll see in just a second. Now, I can express the two sets of samples-- so I've got, basically, an eta 1 is equal to some beta 0 plus beta 1x at the low condition. And I've, similarly, got eta 2. I'm estimating as beta 0 plus beta 1 at the high condition.

That's just a system of two equations I can write it in matrix form, as shown down here. So I've got this condition times my two coefficient. So I've got my beta 0 and my beta 1. That's my first equation, giving me my eta 1-- and similarly, for my eta 2.

Now, since we've got exactly the same number of model coefficients and model runs, my x is square, and I've got a direct solution to this equation-- where, again, eta is 0, because I don't have any replicates, I don't have any noise, and so I can simply directly invert that and solve directly for my coefficient. So my purpose here was both to remind us that there is some vector notation that just follows what you would think of naturally, and in the special case where you've only got exactly just barely enough data, there's a direct solution.

So that's pretty straightforward. But what we'd often like to do is remember that there's noise in our process-- that, if I were to, in fact, run at the same condition multiple times, I've probably got some variance in there. It may be measurement error. It may be underlying process variation.

So what do we do when we've got replicates? First off, we probably want to design the process or the experimental design for the process so that we do sample at multiple-- at the same condition multiple times so we can estimate these variances. I'm going to need that information for things like confidence intervals on the coefficients.

So how do we deal with data where there is some spread, even at the same factor level for hold time? Well, now we're into an over-specified set of linear equations. I've got multiple sets of data for the same input. I've got multiple y's. I got multiple outputs.

And what we'll generally be doing is trying to do our best fit in the same intuitive way you've probably been doing regression modeling or fitting of lines to data for a long time. We have a general approach of trying to get the best fit in a conceptual approach of minimizing some error metric. The most standard one is a sum of squared deviations of your data from your prediction.

So we often talk about least squares or minimum squared error data sets. So in those kinds of conditions, what we're trying to do is find the data not and the one that gives us the best line that, in aggregate across all of our data, minimizes the sum of all squared error terms, where each error term-- each e sub i is-- so this would be my e sub i term. When I ran run sub i for input at, let's say, level 1 for my single factor here.

This is my prediction. This is my data. So I've got a deviation between my best prediction at that input level and my actual measurement. I'm squaring that, and then what we're looking at is the sum of all of those across all of my data. And you can formulate it as sort of that optimization problem and do a search for the best beta 0, beta 1, conceptually.

Turns out, in the simple linear cases, the power of linear models is that one can still solve the matrix formulation of the problem using a pseudo-inverse in the over-specified case, where I've got multiple samples at the same condition. And we need to do that, because now, if I've got multiple rows where I've sampled the same conditions multiple times, my x matrix is no longer square, so I can't do a direct solution.

There's an aside here, which is simply the derivation of this relationship, this beta solution, based on the minimum squared error condition. Yes, question?

AUDIENCE: Yes. For the previous equation, isn't it beta 0 plus beta 1 time x based on beta 0 minus beta 1 time x?

DUANE BONING: Good. Yeah. I guess I could change the sign in the middle, but I probably shouldn't do that. Yeah, that should probably be a plus there. Thank you. So we could, again, formulate our problem just to give you a quick sense of where that funky description of this-- using this pseudo-inverse idea comes from.

Again, if I have my linear model, I can express based on my measured data. These are my different conditions and these are my model terms. I've got a way to express the error for all of my runs, and now the squared error can be written in matrix form as the transpose of that error vector times the error vector.

And now I simply substitute in, and you get something like that. And now you can do the derivative. What we're looking for is the point at which this quadratic error is minimum. So I can do the derivative there, and at that point, I'm looking for the bottom of that squared error curve, that quadratic error-- which, if I differentiate this, I get this.

And now I can simply solve, and then, when you go ahead and plug that in, that drives for us-- this funky thing, where you take your data condition x transpose x, then you do the inverse of that times the x transpose with your measured outputs to solve for beta. So there is a matrix algebra formulation here that's basically giving us the best minimum error fit for all of our data. So that's a very nice, powerful general linear regression result.

Here's another picture of the same thing. This is the same idea-- not the derivation, but what I meant by when I've got replicates. Again, this is my measured vector. These are my conditions. Again, I'm either running at the low condition or at the high condition for each of my-- for my single input here. I've got two coefficients, so my x times my beta, again, gives me my beta 0 plus my either x high or x low times beta 1. And for each case, I get some measured output.

And the point is now, if I have replicates-- there's an x low, here's another x low, and if I continued on, there may be multiple places where my conditions were exactly the same-- oops-- my coefficients are exactly the same, but I get different measured outputs. So that's how my-- what used to be my square matrix becomes non-square.

I'm filling in additional data rows as I add more replicates. And so in that case, I still got the simple linear model, but now I have to use not a direct inverse, but I got to use this sort of pseudo-inverse in order to fit my beta coefficients. OK, so that's the very powerful, general linear regression kind of approach, and it's based on minimizing a squared error term.

But there's a lot more intuitive way of looking at this and some other observations that make estimation easier, and also connect up, I think, very nicely with our intuition. One important point is, when I've only got two levels-- so our one input, two levels-- we know a line has to fit that data. And if I've got replicates, the line actually has to go through-- provably, and we'll see a very quick little derivation-- the line has to go through the mean at each of the data points, my replicate points, at each of my two levels.

So in order to fit the data, one is you could put it all on a big matrix and send it off to Matlab, or Excel, or whatever to do this big matrix pseudo-inverse, or you can simply calculate, look at your data under your low condition, your data under your high condition, find the mean, and your line has to go through the mean of those two conditions.

So one of the two might be a little bit easier on a quiz. And this is simply working out and showing you, if, in this simple two-input condition, where I only had two data points, then you could see how this works. But with the mean, what's very nice is now this is the mean of my output, and what that gives you is a direct expression using that fact to be able to calculate your offset term and your beta 1 term.

Let me see. Am I going to do this yet? No. So here's a simple application of that for our real data, this injection molding case. Here now I'm doing something kind of sneaky. This is I'm treating all of that injection molding experimental design as if I'd only done an experimental design in hold time. So what's really hidden inside of this spread of data is it is all replicates at the low value x minus for hold time.

Similarly, these are replicates at x plus, but they're not really good pure replicates, because other things were changing. First off, time might often be changing, but what was going on in here is I was also playing with the injection velocity at the same time. So in this case, this is saying I'm going to treat these other things that might be changing, like injection velocity, as just noise factors. They're what contributes to the spread at the condition that I'm actually interested in.

And so you will often do that. I'm just trying to fit a very simple first-order linear model, just relating my one factor here of hold time to my output. And it's going to have some variance and error in my estimate because of all these other higher order terms that I'm not including, other factors that I'm not including, pure replication error, measurement error.

But what I am trying to do is understand as closely as possible the effective that single factor. So very often, I really do want to know just the hold time effect fact, even when it's contaminated with or spread out with these other effects. This is just showing, for that actual data, I can fit my two means for the spread, and then fit a very simple linear model.

Now, in this case, I'm treating hold time as kind of a continuous factor, so you might think, if these are hold time in minutes, now you can interpolate. And I could ask, what I think the hold time-- or the effective on my dimension might be at 1.5 minutes? So I'm interpolating within my data. I could also extrapolate and ask questions. Well, what do I think it'll be out here at 2.5 minutes? That might be a little bit risky, and we'll talk about that a little bit in a later lecture, when we talk about the confidence intervals.

But at another point I want to make is this same picture would apply even if these weren't continuous parameters. And I find it easier to think about the problem in terms of continuous parameters, because you're often changing a pressure, or a velocity, or a hold time, and looking at a continuous variation in some output parameter.

But a very important point that I do want to make here is, if I were to use exactly the same chart and relabel the axes and say something like, this is the dimension-- this is still dimension-- and I'm doing injection molding, but here I'm using it from a tool from different manufacturers. So here, I run the tool on-- I don't know-- Fritze Inc's injection molding tool, and over here it's on Xenon Corp's tool.

I could have essentially the same model, the same formalism if these were discrete levels, discrete choices in some factor. Here my factor is, which choice of tool would I want to be making? I want to know, is there an effect? I maybe even want to build a model with model coefficients for the effect, depending on what factor level I pick.

Now, if these were discrete choices, it no longer makes sense to interpolate or extrapolate, but I do want to just highlight that the same formalism-- especially for factorial designs-- applies to both continuous parameters and discrete parameters. I'm going to come back to that a little bit when we talk about what might-- slightly different ways of looking at the model coefficients be in those cases. But I just want to remind you or highlight that you can use these for both continuous and discrete kinds of designs and factors.

OK, so continuing with our plan here-- we've talked about multiple inputs and effects-- what I want to do next is say, if I do have multiple inputs, how does our ANOVA formalism apply to being able to decide whether both factors matter, or not, or three factors, which factors actually matter. And based on my observations, do I think that that factor actually has an effect or not?

So I want to derive first real quickly, or show for us the different ANOVA formalisms in these cases. And then we'll come back a little bit more to the estimation of these model coefficients. Oh, what happened? There we go. OK. So back to our injection molding problem, now with two factors, where we have the velocity and we have the hold time-- and this is where I said I was being a little bit sneaky when I was showing you the one factor data.

In fact, the setup here was to do four different combinations of these two factors-- two-factor, two-level experimental design-- with replicates at each factor-- something like 14 or 15 replicates. And here nothing was intentionally changing, so these are more like pure replicates within that set of experimental conditions.

And what we can do is look at, in that case, where I've got pure replicates, what is the average at that test condition-- so test condition 1, where it was low and low for the two factors, and so on for test condition 2, 3, and 4. And now we can ask the question-- not just in the univariate case, but in the multivariate case, what kind of model works?

And this is actually a description of the model that might be more appropriate if they were discrete choices and you weren't trying to do interpolation. But it's also implicitly the model that we're using for ANOVA analysis, because the ANOVA doesn't care whether it's a continuous parameter or a discrete parameter that we're using.

So the basic idea for a two-- this is a two-factor kind of model. This is very close to what we saw with the ANOVA on Tuesday. One of the factor might be-- or treatment effect that we might be looking for might be the velocity effect. The other might be the hold time effect, and then we may be looking for interactions between the two.

And the picture you should have in your mind is that what we're going to try to do is model this with an overall mean-- some mu-- and then we're asking the question-- think back to the ANOVA question-- is there a deviation from the grand mean, from the overall mean, if I have velocity at the high condition? Is there a statistically significant shift up for a high setting? And similarly, is there a statistically significant low effect?

So is there an up or down shift, depending on whether I pick the high or low condition for velocity? And similarly, I would ask the question, is there an effect-- a beta 1 effect or a beta 2 effect-- depending on whether I selected the low-- I guess this is the high-- the high condition or the low condition for that actual variable?

So these are discrete offsets, depending on whether I was at the low or the high condition. So i indicates, in our earlier notation, either the plus or minus. Here, we're switching rotation a little bit to just be a numerical index-- either 1 or 2 in this case, but if I had multiple levels of that factor, I could extend that. I can have little deltas with a coefficient that would go for each delta, depending on the discrete value that I picked for that level. Yeah?

AUDIENCE: What if it's a shift of variance, and not a shift in the mean altering one of them?

DUANE BONING: It's a good question. The question was, what if it's a shift in variance? Here we are looking for just-- we're just modeling the mean value of the output with this. You could similarly start looking at building up models of variance. You would probably not be thinking of it as additive variance terms in those cases, so you might also end up doing some other kinds of transformations on your data to get to additive models of variance.

But basically, if I had enough data, you could start to try to model variance. If you think back to the ANOVA formulation, the simple ANOVA is all assuming your variance is exactly the same across all of your conditions. There are wonderful books on more powerful modeling and how ANOVA extends to those.

In fact, one of my favorites-- I lent it to Hayden-- is a book by George Milliken, the title of which is *Analysis of Messy Data.* And it deals with all kinds of wonderful conditions-- not only modeling of variance, but also when you've got very complicated design situations with partially replicated, data unbalanced design. You run the whole experimental design and you're missing a data point, because the experiment crashed-- what do you do now?

Looks at every possible thing-- so here, we're really looking just at mean modeling. Just to finish this up, I should also point out that there would be discrete offsets, depending on every combination of our interaction factors as well. And so I might have discrete model terms for those as well.

Now, an important point just to make here is that, in this kind of a model-- and it'll maybe become clearer a little bit later-- each of these model coefficients-- this tau 1 and a tau 2 effect, for example, for the velocity-- they're not really independent. And this gets back to, I think, your question about, in this model, this-- if I were modeling that as a beta 0 term, this really is the mean, and that imposes an additional linear constraint on my other coefficients. And in particular, that tau 1 and tau 2 would have to be equal and opposite here.

This is a great model for being able to conceptualize and plug in, depending on what factor levels you have. You just have to be careful in fitting the model, and recognize that, in this kind of a formulation, there's some implied dependencies within this-- in particular, that each of the factor levels-- factor effects have to be 0 mean overall, and the interaction terms, similarly, have to be, in aggregate, 0 mean.

That kind of makes sense, because-- think of the case where I've got two inputs, two outputs-- or one output, got my two factors-- velocity and hold time-- and I'm just doing two levels for each. I've got four different conditions. I may have replication, but I've only got four different conditions. How could I possibly fit a mean, two taus for factor 1, two coefficients-- taus for-- or betas here for factor 2, and four interaction terms?

That's four, five, six, seven, eight-- nine model coefficients, four data points. So clearly, there's got to be some additional constraints on the model coefficients hidden under that. OK, so that's one way of looking at the experimental design and the data. Here's another way, simply pointing out-- what we'll often be doing is building tables to show the high and low condition for our different factors and the levels.

We might label those, and then often, we'll illustrate our design space or experimental space with a design matrix or a design plot. So here I'm showing simply-- the four different conditions for velocity and hold time either, low or high. and I'm also doing something a little bit subtle, which is labeling them with not just a plus or a minus, but in fact, giving them a normalized value of plus 1 or minus 1.

There's some subtle reasons for doing that. One is, by definition, when I now fit for a two-level experiment-- when I fit a model coefficient to that, it will always be 0 mean, and so now that beta 0 term really does turn into the mean estimate. But-- might see that a little bit later.

But the point is here now this is basically expressing all four of my conditions, my test conditions from the previous design. And I can also plot that, each one of these test conditions, on the x1 axis. We said this is velocity, basically. And this is my x2 axis, which is hold time. So it's either at the plus 1 or the minus 1 level for that factor.

And so I've got an experimental design, where I've got four sample points at the plus minus every combination of those particular factors. And so we'll often talk about this as a corner point DOE or a full factorial DOE, where I've got every combination of those discrete factor levels being mapped out.

And then I go and I run my tests. And here I've just written down the results of the average of the output over all replicates for that point. So this was over the 14 or 15 replicates. Notice here, this is just mapping out the design space, my x1 versus x2. It's not really showing you on a third axis what the output is.

We could also do that and have a 3D plot, and we'll see that in a second. But the point for doing this formulation, this full effects model with these effects offsets and just reminding us what the model exploration here is, is that what we would also like to do, even when I've got more than one factor-- I've got these x1 and x2 factors-- is ask the ANOVA question, which would be-- from an ANOVA point of view, I would like to ask, does velocity have an effect?

If I change from the low to the high, do I get an appreciable and significant change in the output? So from an ANOVA point of view, if I'm just looking at velocity, I would like to ask a question-- just looking together across all of the data that I've got at the low condition for x1 and all of the data from the condition at x2, I might ask that question and say, is there a shift from that perspective or not?

So really, we're back to a hypothesis test-- oh, what happened there? This should be a dot, dot, dot. [INAUDIBLE]-- which is basically asking the question, in the null hypothesis, there is no effect-- all of my taus are equal and 0, and all my interactions are 0. And in the alternative hypothesis, something has an effect. So that's the key question.

Now, what you can do is start formulating, exactly like we did with ANOVA, sums of squared deviations across, for example, all of the A levels averaged over the B factors and any replicates. Remember, this is replicates. This is the B factor. So for example, this sum right here would be-- let's see-- forgot which one-- I'm looking at the responses for the first level, so call it the A level.

So that would be basically the big red thing that I had drawn here. I basically call this first one factor A, and what I'm simply doing is averaging over all of the n replicates at each of these points and over the B levels at that point. So I can form a sum of squared set of deviations from the mean as I change factor A. I can similarly do that as I change factor B.

Then I can also look at these interaction sum of squared terms, and once I have these, I can now start to break down the total sum of squared deviations in my data into components due to factor A, sum of squares from-- deviations from the grand mean for factor B, and terms here from interactions. And then this is residuals, sum of squared error.

So basically, it's just like we did with the single factor ANOVA, but now I can do that, well, looking and seeing, did factor A have an effect? What's the sum of squared deviations that I observed in my data from the mean as I change that factor-- and similarly, for factor B. And I could extend this on to more factors, if I wanted-- if I had three factors or four.

This is also just reminding us or laying out that there are different degrees of freedom in each case, which we're going to need when we want to estimate the mean square, because what we'll do is we'll take the sum of squared deviations, divide it by its degrees of freedom-- which is the number of factor levels I had minus 1 in each of those cases-- and put these together to get a mean square estimate for that factor, which is essentially an estimate of the variance due to changing factor A, just like in the ANOVA formulation.

And you can plot that into an ANOVA table for multiple factors, where, again, we have the sum of squared deviations due to each of those factors. This is just plugging in those formulas that we saw earlier. And the interaction-- if I've got pure replicates, that's a great case where I can separate out exactly just the replication error, perhaps coming from measurement or pure process replication error.

And then, with the right appropriate degree of freedom, I can now form the mean square estimates. Remember, these are variance estimates. And then I can form the right F factor to decide, does this mean square, with respect to my within group mean square-- is that large enough effect to be statistically significant using an F-test? So it's exactly the same formulation for the ANOVA, but now where we can break it out across multiple factors or multiple interactions.

OK. You actually have to try it and use this to get it, but the key point here is, when we are building these multiple factor models, it's still an ANOVA formalism to decide whether or not that factor is statistically significant or not. So let's look at this in the case of a linear model. And I'm going to give you a little bit more terminology here and start to get to this idea of contrasts, which are easy ways of estimating some of these models.

So this is still our injection molding data. Factor x1 here was velocity. This was hold time. And I've got every combination. What's new on this page that I want to draw your attention to is some additional terminology that we often use that's shorthand for indicating the different combinations of those factor levels, that tell us exactly what point, what condition, combination we're running for each of our tests.

So I already showed you this table which was saying, OK, for my first factor-- which we've also labeled as factor A-- factor x2 is our second factor. So note here, we've got three different kinds of-- ways of referring to the same factor. The first factor, factor a x1 value-- it's the velocity term. What I can also do is indicate in shorthand which condition is actually at work.

And so it takes a little bit of getting used to, but what you'll see here is-- let's look at this condition first. What I've selected here for test number two is the high condition for factor A and the low condition for factor B. So what we're doing in shorthand is indicating with a lowercase version of that factor level all of the high selections for that factor.

So the high here meant I picked the high condition for factor A, and I denote that with a little a. And the low factors we just leave off, or think of those as a 1 setting. So my high condition for factor B-- I'll indicate that test setup as-- with a little low b. And this interaction, where I vary and pick the high condition for both of those, gives me the AB combination. And the overall low condition we often refer to as the 1 test condition.

Now, the reason for doing this will maybe not be completely clear, but-- give you some intuition. When we extend this to larger numbers of factors, where you want a very shorthand way of describing which combination-- which test condition, which combination of your factor levels you want to do, and as ways of forming sums on the outputs at those conditions in order to be able to do our estimates. So that's a little bit of additional notation.

Now, the model that we might want to fit is, in fact, perhaps a regression model. It's a little bit different than the factor effects model, which becomes a very nice way to do it when we've normalized our inputs. This works best with normalized input levels-- in other words, my x sub i taking on values of plus 1 and minus 1.

And we'll see how those relate to the offset or effects model in a second. So this is now the same plot, my input x1 and my input x2, but now I'm also adding the output. So I'm getting to my 3D view of the results. This is the same graph as before-- my factor A and my factor B.

But then the height here-- each of the values is the y bar at my n replicates averaged together to give me a sense of what my actual process output was at each of my four input conditions. So we can talk about here the mean values of each of those different treatment conditions, again, where each of those is an output across those. And now we can start to look and say, OK, are there trends as a function of factor A, are there trends as a function of factor B, and fit model coefficients that go or capture with those trends?

So a couple more pieces of terminology-- we can be asking, is there an effect in the ANOVA sense-- and we don't want to estimate what the effective is from a model coefficient effect-- is there an affect or a main effect-- which is the direct linear term-- and are there these interaction effects, or these AB interaction or multiple factor interactions?

And the perspective that we take is very close now to that sneaky trick that I showed-- described with the injection molding data. If I'm trying to estimate the main effective my first factor, I basically am treating my second factor as just noise conditions. And I form my aggregate perspective for, is there a velocity effect, averaging together all of the other data, just considering holding the velocity at its low condition or the velocity at its high condition.

And then I basically do a linear fit through the mean of those two cases. So it's just like we were doing that the linear case, but ignoring or lumping together all of the other data to estimate the main effect in each case. So there's a way to look at this then in our data here for estimating the main effect.

We would label the main effect for A as basically looking and saying, what I'd like to do is take both of my high settings for A-- this is the plus setting for A-- average the outputs together for those two cases-- similarly, take the low condition for A-- let me use a slightly different color here, make this blue-- look at the low condition, and average those together, and then look at in aggregate as I went from the low to the high for condition A.

Is there an average jump up? Is there an effect of factor A in aggregate? And so what you can actually do is very simply form-- or model that overall delta coming from factor A as just the overall average between these two points and the overall average between these two points, and simply fit the line through the mean of those two cases.

So essentially, I project all of my data just down onto my A axis and ask the univariate question, does my y output change as a function of A-- lumping together all of the other data? And so that's simply what we're doing with this equation. And now that terminology that I introduced in the earlier picture comes into play. You can talk about now the outputs at y sub A high as being those cases where I had the plus value for A and the low value for B.

And the other one would be the plus value for A and the high condition for B. So these two points I'm simply referring to-- it's getting kind of messy here, isn't it? I can refer to this average right here as the A and AB outputs averaged. I've got two of those levels, and then n replicates at each of those points. So that's just a way of expressing that average, using our test condition terminology.

OK. Everybody get mean effects? We can also do interaction effects. And they're a little bit more subtle to form the contrast, but it's essentially asking a different set of averages, and trying to say, as I change both A and B together, do I get something more than the A effect alone or the B effect alone? And what we actually do in this case is take the AB condition-- this one right here-- take the low-- so I'm taking the high-high and the low-low, forming its average.

Then I take the other corner, the just A high and the just B high-- take its average, and then basically take the difference between those two. So I'm averaging the corners together, in contrast to the main effect, where I just averaged along the two factor levels. And that forms an estimate of the AB offset.

Basically, this average of the A high and the A low we'll refer to as the contrast, the right combination of the output values that give me an estimate for the A effect. And similarly, we'll talk about the contrast here for the B factor. Which combinations of output values do you use to estimate that effect and the contrast for the interaction terms?

And what's cool, if I go back to that early slide where I said, if I'm just fitting a linear output, y as a function of one factor, and I got two data points, my line goes through the average of those two conditions. If I form the contrast now in a multiple output case, again, the best estimate of the linear line goes through the averages-- these are the averages at each of those conditions.

And what I've simply got is, directly from the contrast, a direct estimate of the model coefficient as a function in the normalized axis space. If our x sub i's are in this normalized plus 1 to minus 1 range, directly calculating the contrast of my outputs gives me directly what an estimate of the effects are-- the main effects as well as the interaction effects.

So I didn't actually have to go and do the optimization problem, do the linear regression. I'm using the fact that the linear regression forces the data to go through the averages. And I can simply look at the data conditions that I've got, the averages of the output at each of my four combinations, and then look at it either from the A factor perspective and estimate the A affect, with a B factor perspective and estimate the B effect, and then also look and see if there's an interaction between the two because of cross-coupling between those two factors.

So what you've got is the regression perspective and this contrast perspective are exactly the same. Now, I said this was a linear model, and what's cool is it actually is from either the A perspective-- I guess that's the B factor. That's the slope, if you will, versus my B factor, my x2 variable. I could similarly have the slope from my A factor. This was my A axes, or x1 axes.

So I've got a simple linear dependence of output on the input in each of those two cases, and the interaction basically gives me linear dependencies that add additional offsets, but that are locally linear in each case. So what we get is this so-called 3D ruled surface that is linear every time I project it on either the A axis or the B axis.

But it's got this kind of funky ability to shift because of the interaction, so it's kind of a subtle non-linearity at work. It's not a curvature non-linearity in the sense of quadratic curvature. It is a ruled surface projection down to linear dependencies on A and B for each of my combinations.

So here's an example of the surface that results if I don't have the interaction. And a very simple model example is x1 and x2 with some linear coefficient, and then what I've simply got for my ruled surface is a plane. It's a slanted plane as a function of x and x2 factors. Sometimes we also draw these interaction plots, and all of that this is just looking at one of the factors-- say, along x2-- and picking some value for the other factor-- say, x1 is-- let's see-- x1 is low-- that would be this line right there-- or x1 is high, which would be this line right there-- and then simply plotting y just as the univariate function of x2, where I picked or set some other value for my x1.

And in this case without interaction, you'll notice those lines have to be parallel-- that what I've got as a simple linear additive model without interaction is I just get off sets because of whatever value of x when I pick or offsets because of whatever value of x2 I pick. So this is simply saying the x1 effect is shown right here as an additional shift in the output dependence on x2.

So this is just another slice through the 3D picture, but what's important here is you would see this kind of interaction pattern on your output as a function of your input, if there's no cross term between x1 and x2. And in contrast, here's the response surface of output as a function of my two inputs. This is, again, a ruled surface. Each one of these is just a line, but now it's got this funky non-linearity because of this interaction term.

And in this case, if I take a slice at different low and high of x1, you can see that there's something at work more than just an additive effect from x1, depending on my value of x2. There is synergy between whatever x1 value I pick and x2 value I picked that together-- I need both to explain the output and how they work together. You can also have negative interactions, where the cross term causes a negative synergistic effect between the two.

OK, so what we'll pick up on next time is extending this a little bit to a more general way of talking about these contrasts. I think hopefully now you've got the picture from a univariate kind of modeling and regression picture up to a two-factor kind of picture with these interactions. Next time, what we'll do is extend that-- oops-- extend that or generalize that to when I might have three factors-- an A, B, and a C factor, for example.

And how do you visualize your design space in those cases and how do you form the contrast to very quickly estimate model coefficients in those cases? And then we'll also get to-- oops-- why'd that happen? We'll also get next time to looking a little bit more at, how do I check the adequacy of these model terms, are they significant, as well as, what are good estimates for confidence intervals on these things-- and then come to these other subtle points, like what happens if I don't want to do 2 to the k, when k is 5?

I don't necessarily need every combination. And we'll talk about fractional factorial designs. So one thing you should do in preparation for next time is start reading the experimental design chapter. And I can't remember which chapter it is in Montgomery. I'll post that on the-- as an announcement on the website.

But there's a lot of lingo to get used to, but it all comes back to nice intuitive relationships, I think. But you do need to start at least scanning that chapter in Montgomery for the experimental design stuff. I think Montgomery has a more thorough description than we can find [INAUDIBLE]. Actually, [INAUDIBLE] might be a good place to get the quick read on it first, and then Montgomery in more detail.

So with that, I hope you guys have a great spring break. And enjoy the MIT spring break, for our MIT students in Singapore. You can rub it in to all your classmates who are still meeting for other classes in Singapore. You get a break to work on your projects out in Singapore. So we'll see you in two weeks.