**PROFESSOR:** Well, while people are filing in, we can chat a little bit about the quiz overall. I was very, very impressed, at least with the question 1 and question 2, which I graded. I haven't graded question 3. So here's all the statistics. You can see that this was a pretty difficult quiz, actually. We do agree. You have evidence to--

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** --to overcome that null hypothesis that it was a difficult quiz. So the average there about 78, 79, with a fairly wide spread-- what we will do is post both the quiz and the solutions-- which may be more interesting to you, since you already have the quiz. And if you do see anything, feel free to chat with us about the question or whatever.

If you're concerned about a point here or point there, I really wouldn't worry about it. Our grading is-- we attempt to be as accurate as possible, but we reserve the right to have a little bit of variation, given the nature of the course, in the grading. And we will, of course, be treating everything a little bit as a mix in at the end, from this quiz grade, the second quiz grade, class participation, homeworks-- all of that.

So I think the main thing to do is look at the solutions and make sure that you understand at least the approaches that we took in the solutions-- which is not to say that some of the approaches that people took on the quiz is wrong. In fact, on problem 2, I think the biggest thing that we observed is people answered the third part, if you remember, really the hard way.

And unfortunately, you may have burned a fair amount of time on that with very detailed t-tests, and f-tests, and complicated hypothesis tests-- whereas what you had was already the confidence intervals from parts A and B that you could use directly, in many cases, to make inferences.

So were there any questions on the quiz-- top level, not detailed? OK, well, take a look at those. And again, talk to Hayden or I if-- more if you have any confusions, I think, is the main thing. OK, so here's the plan for today.

We started on full factorial models design of experiments last time. I'm going to pick up where we left off on some of that, talk a little bit more about the extension of that to generalize to the k experiments. The key idea here is, again, picking out experimental design points so that you can fit particular kinds of models.

And we'll start, again, with ANOVA for linear models, and then move on to asking the question, what if a linear model is not sufficient? How do if a linear model is not sufficient? How do you check for the adequacy of the model form? Is there evidence that you need higher order terms, like a quadratic or other curvature? So we'll spend the first half really focused on that.

Then we'll start moving into some additional issues in experimental design-- these ideas of blocks and confounding, as well as-- those same issues come out very strongly in fractional factorial designs. So we'll get to that. Now, I posted on the website reading assignment.

The key is probably Montgomery chapter 12. I find that a little bit better description than May and Spanos. But there are some additional topics at the end of the May and Spanos chapter. I can't remember which chapter that is-- 7, 13-- can't remember. But it's posted on the website. And so there are some things you'll want to scan in May and Spanos as well.

So recall that last time we started talking a little bit about the kinds of models that are associated with different experimental designs. So for example, we start with the simple model-- it's not the simplest, because we already have two factors, but a very simple 2-by-2 or 2 to the 2 model where, again, what we've got are two levels in each of two factors.

So our factors here are x1 and x2, and we've got two different levels. So we're just picking the corner points or extremal points for that. And we talked last time about how this can generate a regression model that has both mean and first-order terms, as well as an interaction term. We also talked about the way that you can estimate these using contrast, and I'm going to pick up on that a little bit, where we talk about the effective varying x1-- the total effect being a movement in the output, or an average movement in the output by a magnitude A in these coded coefficients for x1-- as well as effects for B and the interaction effect.

Basically, we're seeing that there is this correspondence between a regression approach or a contrast approach. It's a very fast, quick way to estimate these coefficients using these kinds of design of experiments. OK, that's not good. That's very strange.

All right, so a little bit of the terminology that we talked about just to remind us, and then expand on that a little bit-- there is this notion of a contrast that's going to get us to being able to estimate the effect of, say, coefficient A, or beta 1, or beta-- each of our two main effects or the interaction effects.

And these condensed down tables summarizing the design of your experiment are really useful tools for being able to generate these contrasts, for being able to identify what the mix of level settings are for each of the trial runs that you do, and so on. So a little bit of terminology-- again, we said this column here, the trial column, is essentially our run.

And we can label them in a shorthand that tells us what our factor settings are, our level settings are for each of our two factors-- our x1 factor-- that's one-- and our x2 factor. So the first label here is just the low setting for both of our factors, and we'll label that with a 1. Our high setting for the a factor-- we label that one with a high setting, or a plus. Sometimes we'll label that as a plus 1-- and our low setting then, for everything else, for b.

Our b factor-- that's where we're set that one high. And when we set both of them high, that's our plus setting for the two of them. Now, we can also think about, what is the factor level for the interaction? And we do that with this very funky shorthand algebra, just saying, OK, AB is really-- think of that as a multiplication of the A and B factors. So a minus 1 and a minus 1 gives a plus 1 there-- and similarly for all of the other columns. That generates the level setting, if you will, for each of those combinations.

Now, in the actual experiment, of course, we're only setting x1 and x2 for each of our trials. These experimental interaction columns are essentially giving us information that we need in order to form the contrast in the shorthand way of estimating effects, as well as-- and we'll see this a little bit later-- as well as telling us what the interaction term should be, if we're going to form regression-- including do the regression on the interaction factor. So it's pretty funky, all intertwined here.

Something strange in our slides-- I don't know what that is. That's supposed to be a dot, just a multiplication-- because now we can say, OK, how do we form the contrast between our different combinations of settings in order to estimate the effect of A? And essentially what we're going to be doing is forming the average between the high for A-- this column and this column-- or the difference between the average of those two columns and these two settings.

So I'm basically forming the contrast in each case for each factor as the difference between the high and low setting. And so that's actually why we need the interaction factor, so we can look at the high settings for the interaction, the AB and the 1, and then the minus A and the minus B.

Where are these coming from? A much better way to picture this is if we look at this graphically. So what I'm going to do here first is generalize. This is a 2 to the 3 design. In more general terms, we can do a full factorial with any number of factors and form this sort of hypercube, again, with corner points at all combinations at the high and low values-- or levels for each of our setting.

And on that hypercube-- so now I've got A factor, B factor, and C factor. This is a nice picture for forming what that contrast is doing and how we're estimating the average effect of changing one factor at a time on this picture. So just conceptually, or intuitively, if you were trying to form an estimate of the factor effect A associated with some variable x1, notice that-- what we've got.

At all of these settings on the left, shaded with the blue, those are all low settings for our a factor. I'm varying B and C also, but for all of those, I'm holding the a factor at the low value. And then similarly, on the red side, I'm holding the a factor at the high value in all of those cases-- maybe varying other things too. But if I think that those other effects are basically balancing out, now what I've got is a difference between the A high and the A low as a net overall estimate of the total effective changing variable x1, changing the variable associated with A.

And that's all that the contrasts are doing. They're simply saying, OK, the high contrast is all of the places where A is set high minus all of the places where A is set low. Form that difference. Recall that we've got-- I need to take the average. So in each of these cases, I've got four points at high and four points at low, so I'm taking the average and taking the difference. That's my estimate of the total net effect. Question--

**AUDIENCE:** What happens if I locate AC and I have a diagram of four points of the same [INAUDIBLE]?

**PROFESSOR:** OK, say that question again. Here's AC.

**AUDIENCE:** Yeah. It's like then I think if I move from 0.1 to AC, and then I have that points AC, BC, and AB are in between. [INAUDIBLE]

**PROFESSOR:** So far what we've done is only estimated one of the contrasts. It looks like you're actually getting close to estimating a different term in a model. So what we did here was just to estimate the a effect. I would also need to do the same thing for the B effect. So that's telling me, if I vary x2-- and I can similarly do that for the C effect.

But then there's the question, well, what if there's an interaction between two of these terms? And one can also form a contrast for an interaction effect. And maybe this is not exactly what you are describing, but it's getting close to other ways of combining some of these effects. So it turns out what this does is makes a balanced combination of our data points to be able to take advantage of that AB column.

This is just basically taking the high and low combinations and forming that estimate of the AB interaction effect using the contrast between these points. Now, there would also be another contrast for the BC interaction, if you rotated these around and formed the plane where you took the high-low combinations of B and C, instead of A and B. OK?

Does that get at your question, or was there-- you can always form different combinations of things. The question is, what combination do you need in order to be able to estimate the coefficient in the model?

**AUDIENCE:** Yeah. I mean, [INAUDIBLE]. I don't know. It does not like derivatives of A, B, and C.

**PROFESSOR:** Right.

**AUDIENCE:** You just take [INAUDIBLE] and cross them.

**PROFESSOR:** Yes. That cross thing is intuitive in the sense-- only in the sense that it corresponds nicely-- you can generate it once you have this column. So now I can connect up the high points, and that would be one plane. Connect up the low points-- that's my other plane. And then I take the difference. But it's not entirely intuitive why that particular combination gives you the estimate of that coefficient. I agree with that.

So this is the shorthand that gives the answer. In fact, I don't have a good intuitive explanation, just looking purely at this picture, why that estimates the interaction. The closest I can come is almost in two dimensions, instead of three, with these interaction plots that we've drawn before.

So the basic interaction plot would look and say, if I just have x1 and x2 values, I just draw them as here's my output as I vary x1 from high to low, holding x2 low, that would be one effect. If now I let x2 be high, I don't have just an offset, but I've got a different delta here than here.

That means that, in combination, x1 and x2 work together to do something different than purely x1 or x3. There is not just purely an additive A effect and a B effect, but there is an AB effect. So in two dimensions, essentially what you're doing is looking at forming the delta from this combination and this combination, and saying there is a difference-- an extra delta between the two. In 3D, it's a little harder to see, but I think it's doing conceptually that same thing. Yeah?

**AUDIENCE:** Why do you take the average divide it by 4 but not do the same in the two factor?

**PROFESSOR:** Oh, I think we did in the two factor.

**AUDIENCE:** No, no, in the slide previous to that--

**PROFESSOR:** So the 4 is basically just because I'm taking the average from each of the-- this plane has four points, so I'm just taking the average from that the difference from that, so the 4 is coming from the number of points I had in each average. That's an estimate of the total effect. Then there's one additional subtle point, which is, how do I get from that factor-- total of net effect, main effect A, or interaction effect AB-- to an estimate of the model coefficient? And that's where some scaling comes into play.

What we've done in these pictures is scale our x1 and x2 to a low of minus 1 and a high of 1, and intermediate point of 0. Let's say this was the overall a effect. What I'd like is something that goes like y equals beta 0 plus beta 1 x1. And so what I've got is a total range in x1 of 2. So I have to basically scale this so that I've got A/2 x1. As x goes from 0 to 1, I've got half of the effect, and as x goes from 0 to minus 1, I've got the other half of the effective.

So that's where beta 1 is equal to the A/2. Again, a lot of what I'm saying is only true if we're using the scaled coefficients for x1, from minus 1 to plus 1. And then all of this magical shorthand of these column tables, the multiplication column tables, the formation of these contrasts works out really nicely. And so that's where you'll look in the literature, and there's so much machinery around this that makes it relatively quick and easy.

So we can extend all of this, again, to-- beyond one factor, or two factors, to three factors, or any k factors. So we can consider, for example, a 2 to the third experiment, where now we have three factors-- an x1, x2, x3-- or going with an A effect, B effect, and C effect. And then we can just extend this to the different treatment combinations.

And essentially, what we're doing here is formulating every possible corner point, every possible combination of the high and low for each of the settings. And the same idea applies. Here I've indicated them with a plus 1, minus 1, rather than just plus minus, but it's exactly the same idea. And you can see C already lurking here. Which runs would you use to estimate the C effect?

That one sort of leaps out at us, as we happen to have ordered this table. I would average these runs-- runs 5, 6, 7, and 8-- average runs 1, 2, 3, and 4, take the difference, and that's my estimate of the C effect. That makes good intuitive sense.

Now, that's the contrast that I would do, again, for C. If I also want to form contrasts for the other coefficients, it's exactly the same thing, but the table is not quite oriented to make it leap out at me. But essentially, for the A contrast, I'm just forming those for my A, AB. That's my plus 1's, all my pluses, and all my minuses, forming that as my contrast.

I can very quickly then estimate that main effect using this big column. I can also estimate my interaction terms. We already talked about an AB interaction. If I have three model-- or three factors, I can have three-way interactions where, depending on all of the settings of A, B, and C, I might get a little boost up or down, different than I would just from A, B, C, and even different from the boost I got by having A and B in combination.

I can have these subtle three-way interactions, and you can form the contrast and estimate that effect in a similar fashion using the ABC interaction column, and form the contrast using that. Two steps to get to here-- one step is, how do I generate this column? And then the second step is then using that column to form an estimate of the interaction effect. So how do we get to ABC column? By extension of what I told you about the AB column-- how do you generate that?

**AUDIENCE:** Just multiply.

**PROFESSOR:** Just multiply. It's kind of this repeated multiplication. I just take the A, B, and C columns and pointwise multiply out those factors. So minus 1 minus 1 minus 1 gives me-- I overwrote it-- there we go-- gives me a minus 1. And I similarly do that for all of the columns. Once I have that, then doing the second step is pretty simple to estimate that contrast.

Once we've got the contrast, just the differences in the columns, then I do have to do a little bit of that normalization. Actually, I probably am not completely consistent here. The contrast is just the differences between the points. I then have to do the average over the number of points. Oops-- well, two things-- to form the estimate of the overall effect, first off, I have to divide by the number of points I had in each of those contrasts, which is basically just 2 to the k minus 1.

When we had that cube with eight points, it was 2 to the third 8 divided by 2, because I'm forming half of the points in one contrast, half of the points in the other. And then this n, which we haven't talked about very much, is just the replicates. I can do multiple runs at each experimental point, so I can also average those in.

So it's also possible that we've got more than one run at every experimental combination point. So that's how we can estimate those effects. And then, once we have that, for example, here, that's how we get the-- in the 2 to the third, we get a 1/4 factor that we saw earlier for the three factor cube in three-dimensional space as an estimate for the overall effect.

OK, now, another point is here I'm talking about estimating our model using these contrasts and factor effects. Another approach, and probably the approach that you would end up using more often if you actually had computers handy, is to do a regression. Take your data, put it in, solve the regression equations, and regress to get your estimate of your coefficients.

And essentially, these big tables that we talked about, these combinations are still critical and very, very useful, because that's basically how you form the x matrix that describes the combinations of all of your factor levels that you feed into the regression. In fact, your coded values from minus 1 to plus 1, this is your x matrix right there, because what you're trying to do is estimate a beta 0, a beta 1, and a beta 2, and a beta 3-- oops-- beta 3, and then also trying to estimate these interaction terms-- a beta 1, 3; a beta 1, 3; beta 2, 3; and even a beta 1, 2, 3.

And so you're basically fictitiously creating these additional variables for the product of x1, x2, x3 in your regression. So you need this table, even if you're doing regression, in order to be able to set up the regression if you want to also include these interaction terms in your regression model. You have to create fictitious columns of data based on those interactions, the product of your three input settings.

And this is basically telling you exactly what those were. So this is essentially showing that same thing I just described in words. What we're talking about here is basically setting up to do the pseudo-inverse when I've got my data, with my x matrix being my setup of my input conditions, and then y being my output, and then beta being my vector of coefficients with both the mean, the main effects, and the interaction effects.

So I would have to create the column of the x1, x2 product in order to be able to do that. And then you can see again here the relationship between the contrasts idea and the regression idea that comes out. OK, relatively clear? Questions on that? OK.

So now we've estimated coefficients. We've estimated effects, either in the regression or the contrast approach. A key question is-- you do that. You get a value for A-- the A effect that's non-zero. Do you think it's significant or not? Is it real or not, or would you have observed that amount of A by chance alone, given random noise in your experiment?

This is exactly back to the ANOVA question that we were asking earlier. I would like to know the significance of these effects. Is it large enough to give me 90% confidence or 95% confidence that it's real, that it wouldn't have occurred by chance alone? And so essentially, what we want to do is run the ANOVA for each of our effects.

If I have three factors, I'd like to know, is the main effect associated with factor two-- is that significant? I can also do that for interaction effects. And very often, you might see large main effects, but you want to ask the question, is there really an interaction between these two? Can I think of them as separable coefficients or separable knobs on my equipment, or do I have to think of them as highly coupled?

I'd like to know the significance of the interaction effect. So we essentially just use the ANOVA approach fairly directly, where what we're doing is forming an estimate of the total deviations associated with that effect, the sum of squared deviations, which turns out to be the contrast squared, and then divided by the n over 2k factor.

There's a homework problem set where you get to help derive why that it's true, or convince yourself why that is true. You're looking forward to it, I can see. But what that lets us do is look again at the total amount of magnitude of sum of squared deviations around a mean associated with that effect, and then look at that in comparison to some estimate of the random noise, and do an F-test, just like we normally do with ANOVA.

So you can start to, again, do that. If you've got two effects or two factors, you would look for the question, is there something significant in the A effect, in the B effect, and in the interaction? And then the remaining sum of squared deviations or the residual gives you your error estimate that you use in the denominator, and that's how you decompose the total observed variations.

So then you can form that into an ANOVA table, again, looking at the main effects A and B, the interaction effect, and an estimate of the error. Again, we get this little funky sine. That's supposed to just be a multiplication. That's just a little dot-- where, again, you're forming estimates of the mean squares, and then ratios of those. And then you do your F-test to see if the F-is higher than that.

So this is one way of looking at it, in terms of contrast. If, instead of forming contrast, you had used the regression and come up with regression coefficients associated with each of the terms, you can similarly form the sum of squares and the mean squares from that perspective. And in the regression form, by the way, that turns out nicely, because it's-- also gives you the chance to not subtract off the mean from everything, but actually estimate whether the mean may be significantly different than 0. But it's essentially the same thing.

OK, so what I've tried to convey to you here is that you can use ANOVA to estimate-- that's a tool we already know-- to estimate whether the effects are significant or not, and whether you should include that term in your model. If you don't have enough evidence that that value-- the coefficient associated with x1-- might be non-zero, you're probably better off not including it in your model-- in which case, those-- any little squiggles that might be there, but you're not confident really are there contribute to your error term. And you treat them basically as noise, rather than trying to fit them.

OK, so let's try this out for an example. This is one of the sets of data that we looked at very early in the term. It's the break forming. Remember, the press-- we have either aluminum or stainless steel, and we've got an angled press, and we're pressing it down into some depth and trying to bend that piece of sheet metal.

And that was actually a designed experiment. In fact, the two factors that we were looking at here is how deep the punch goes in and what the material type was-- either aluminum or steel. And this is just identifying the factor levels that I'm going to associate with these terms.

Now, this is actually an interesting one, because this is a discrete factor level. Well, this is a more of a continuous factor level. And you'll catch a little bit later-- there's something a little bit odd, where we talk about the zero factor sitting between aluminum and steel. That actually doesn't make any sense.

But I can talk about perhaps the zero factor setting of 0.45 inches halfway between the low and the high of a continuous parameter. So I like this because it's mixing or showing that the same tools and technology mostly work for both discrete decisions and design of experiments, and continuous ones. And we'll see where it doesn't a little bit later.

So let's try out a very simple designed experiment, a 2 to the 2 factor where the final angle after removing the punch is our output. This is the actual data. What we've got is, for each of the level settings, we've got replicates. So I'm doing 10 replicates at each test setting of the high and low combination for x1 and x2. So I've got four different corners, four different combinations, 10 replicates.

And here's the data for that. By the way-- we'll come back to this a little bit later-- but if this is run number and if those are done in time, there's actually something a little risky in this experimental design. This is foreshadowing to coming back to additional issues of nasty things you've got to keep in mind when you do design experiments, but what if there was some time degradation, some time where in the tool?

The problem is you wouldn't really be able to distinguish a time effect from a change in the factor setting. So one approach that you might-- if you can cheaply do it-- here you might actually have had to make maybe some kind of machine setup change, and you wouldn't want to do this. But you might randomize the order of your runs in time to block against the effective time.

And we'll come back to that. But here, it's a simple-- I do all of my low-low combinations first, then I did my low-high, then my high-low, than my high-high. So that's our data. Now, what we're doing here, first off, is I'm just going to form the mean-- I take that the mean of those 10 points, and that's going to be my y bar at each of my level settings.

So I'm reducing it down from the 10 replicates down to four means. So this is my mean at each of those four different conditions. And now I have my output, my y vector-- setting up for a regression-- as being the observed means. What's my x matrix setting up for my regression? Oops-- it's based on the level settings.

But notice, we also add in that one factor. That one factor is basically giving us-- when I multiply out my matrix, that's corresponding to my beta 0, my overall mean, so that what that's going to do is be able to add all of those up, divide by 4, and give me an estimate of the overall grand mean of my output.

And we also create the x1, x2 cross factor in order to be able to estimate and interaction term. So overall then, this here is my x matrix. I can then solve it. Now, notice we're using the direct inverse, and not the pseudo-inverse. Why?

Remember, ultimately, we've got y equals beta x. That's the model we're trying to form-- beta 0 plus beta 1, beta 2 plus an interaction term. In this approach, how many data points do I have fitting into this regression? My y vector was four data points, because it was just the means.

How many coefficients? Four-- I can exactly fit my four coefficients, and that's why I can do the direct inverse. And when I do that, it directly pops out these values from my data coefficients, which feed or give me my regression model. What's epsilon in this model? 0-- it's 0.

I exactly fit my four data points with my four coefficients. I don't even have any extra data to estimate epsilon, if I just use the mean values. However, if I go back and expand out-- realize I had 10 replicates at each of those mean values, that's one form-- or one source of information about pure replication error-- that is to say, noise factors in the data.

So in this approach, where I just use the mean and the regression, I don't have an estimate for epsilon. I would have to look back and form a separate pool of estimate across my four experimental data points for the noise variance in the system. OK, let's see. What else do I want to say here?

I thought I had another example. OK, we'll come back to that. So once I've got the model, this is essentially the model that I have of the true underlying process-- again, first-order, linear, plus an interaction-- plus maybe I have lingering in my head, there might still be some higher order quadratic dependence, for example, on punch depth. I haven't really sampled enough to be able to estimate that. And I also have some noise.

The model that I have fit right here, my y hat estimate, includes estimates-- I guess I should say estimates here-- for these different terms. And what we're basically saying is we're lumping in together all of the errors that might be lurking as two effects together. One is random noise and the other is neglected higher order terms. That's our acronym for higher order terms. Question--

**AUDIENCE:** What if you put the order of [INAUDIBLE]? Then you can actually get epsilon 0.

**PROFESSOR:** Right. So the question was, what if you add additional higher order terms to drive epsilon to 0? This is a very good point, because if I don't have any replicates, I could always add model terms until I finally get to where I have a perfect fit-- just like we saw with the four data points and four-- in which case, my residual then is 0. But I've added model terms.

But I also have no way, in that case, of knowing whether the model terms are significant or not. If I have replicates, on the other hand, then we will almost never-- if there's any noise in the replicates, where the replicates do not lie exactly on top of each other for every rerun at the same condition, then there's no way to drive the residuals to 0, because I will have pure process error at least, even if I've added all the model terms I need to drive model error, the HOT, to 0.

So we've got actually these two things going on. We've got two key sources of residuals, lack of perfection in our model that we want to talk about how we distinguish and detect-- because one of those is whether we have an adequate model form. Do I have the right terms and enough of the terms in there? Or is there just replication variance, noise-- which I can't fit? All I can do is estimate the stigma associated with that replicate noise.

That's the key question that we want to get to. Before we get to breaking down the residuals, one thing that you should always do after you fit your model is actually look at the residuals to see if there are trends qualitatively that may be missing, because that will often give you a lot of insight into whether it's a model structure problem or not.

And so for this data, what we've actually done here is look at the residuals associated with the 10 replicates that we've got. And what you'll see is, first off, the mean of your residuals had better be 0. Otherwise, you've actually got a screw up somewhere in your fit, just mechanically.

But what you would like to see is, essentially, the residuals ought to not depend on the level setting or the experimental data point. Especially with continuous parameters, if what you see is the residuals for small values are very tiny and the residuals for big values are really, really large, that actually violates some of the assumptions we've made in the model fitting.

And you might need to go and do a transformation of your data-- perhaps something like a logarithmic transformation or some other transformation of your y output data-- because your residuals are not all n 0 sigma squared, as we assume-- the same variance for each of your data points. Question--

**AUDIENCE:** [INAUDIBLE] mentioned if we have four thumbs and four data points to error, thumb moves to 0. But isn't that supposing the x1 and x2, that we've chosen our only two factors which [INAUDIBLE]. We could have [INAUDIBLE] and x3 [INAUDIBLE].

**PROFESSOR:** Yes.

**AUDIENCE:** In which case you're still [INAUDIBLE] and it should still be pretty significant.

**PROFESSOR:** Yes. So the question, for folks in Singapore who may not have heard that, was, aren't there other possible sources of residual or error-- including, for example other factors, like an x3 factor that we didn't even consider is lurking underneath? And that's absolutely true. Those are all potential additional sources of-- I guess you could look at it as either model error or noise.

What we're doing is treating it all as noise. And we'll talk a little bit more about nuisance factors and how we might walk against that. But essentially, we're lumping all of those things into just the sigma squared, unless we're explicitly breaking them out, treating them in our model.

**AUDIENCE:** Hold on. There's a question.

**PROFESSOR:** Yes-- question--

**AUDIENCE:** What if I don't have enough data-- for example, I want to have my replicate, but I still want to estimate the residual? Can I just drop the higher orders and use the data points to estimate the residual?

**PROFESSOR:** Yes. So you did, essentially, the special case where, if you believe really the higher order term, or even the interaction term in our 2-by-2-- let's say I really believe that really is not any reason for there to be an AB interaction-- you can use that as essentially a replicate, and not include that in the model term-- in which case, then I've got four data points, three model terms-- beta 0, beta 1, beta 2.

And I've got the barest bones of additional data to be able to form an error estimate, with basically treating the interaction as noise, because I don't think that it's real. And then you can do-- you've got just enough to be able to at least ask minimal questions about significance.

One last thing that we always want to do with our residuals is do a quick test or check to see, is it reasonable to think that these things are normally distributed? So you can, again, do a qq plot and see if there's strong evidence of something strange going on. Now, I said here we had to do-- if I used my y average as my output, I can do the regression and solve directly.

You could also formulate the regression problem using all of the data, and not pre-calculating the average for each of my four settings, in which case we set up our data little bit larger. I still have my-- I don't know why it became an eta. I guess there's my eta. That's still a y. This is my y data.

I still have my x, which has my main effects and my interaction. But now each of the replicates-- for example-- I don't know if you can see it, but this is a 1, 1, 1 setting, and that's one of those outputs. Here's another 1, 1, 1, and the output for that particular run. In other words, I have multiple rows in my data. I have 10 rows with the same factor settings, and each one of which has a different output, y.

And now I can set up and do my regression. Now I have to use the pseudo-inverse, because I've got more data than I've got coefficients. So I'm doing my least squares fitting using my pseudo-inverse. And what pops out, if you actually do that with this data, is exactly the same data. Thank God, right? So it's the same estimates of beta as before.

Now, you might also have directly a residual based on the difference between your predicted y-- your y hat-- and your actual measured. I've got 40 of those values, and I can use that as my overall estimate for my overall residual in that-- in this case. So it sort of pops right out as well, and you can form the same residual.

So if you actually plot this for the response surface for our very simple linear model with interaction, this is what you see. As you change the material axis from your high to low in your depth axis, you've got this-- looks like mostly a slanted plane kind of effect. Oh, let me ask you. Just looking at the plot, do you think there's a significant effect from factor x1? Significant effect from factor x2?

I see some heads shaking. Oh, looks like it's a nice trend-- there's lots of movement in the output values. Sure-- why not? Just looking at this plot, I argue, you can't tell. This is just your best fit based on your betas, but you don't actually know whether they're significant or not, just plotting the output of your model.

Here's why. For example, if I were to tell you, OK, at this combination of depth and material, at the plus 1, plus 1, I actually plotted the 10 different values you saw, and they all looked like this. And if I plotted the low-low and they all look like this, now you'd start to say, yeah, my model output-- there's a big separation between those two. I think that's significant.

On the other hand, if I plotted and showed you a scatter around here that was huge and a scatter of your actual data points around this condition that was also huge, and there was big overlap between the two, now you'd be a little more leery, right? You'd say, well, that was my best bet, but I'm not sure that they're actually significant.

So this highlights, again, you've got to use the ANOVA as well to be able to decide, is there actually significance to the model coefficient that we were looking at? It's just another picture. This is looking a little bit at the side. This is pointing out, if I believe that the terms are significant, you could ask the question, is there an interaction effect?

It looks like there's not quite parallelism between these two lines. There may be an AB interaction, but I'd also need to ask the question about significance on that. So that's the key question. The effects of these terms-- are they significant or not? We would do the ANOVA analysis.

By the way, the ANOVA and the t-test-- almost the same. In fact, I think another problem on the problem set is understanding a little bit the relationship between these two ways of looking at the problem. If I were to just look at this data and look at the effect of depth only on the aluminum-- so I'm just holding-- just looking at the aluminum parts-- for the depth of 0.6 and the depth of 0.3, there's the spread in my data. I can look and say, what's the variation within the test compared to the variation between tests?

I could then think of that as my F-test. Or I could do a pairwise t-test, or do a t-test looking at the difference in the means, given the spread between the two-- turns out to be the same test-- gives the same significance level whether you do it with an F or a t. But essentially, this is the same question I was just describing qualitatively on the 3D plot, looking at one factor at a time.

I can ask the question, do I think the effect of depth is significant? This is looking at it qualitatively. I can go and chug, and do the ANOVA or the t-test machinery, but it gives the same idea. I can do the same thing for the material effect. Say, if I look at aluminum versus the stainless steel, is that effect significant?

By the way, you can go, and chug through, and do all of that. What do you think? Are these factors significant? I have high confidence. I'm pretty confident that those are real effects. There's a huge spread between these, so the intuition is very clear. The interaction effect-- that's a little tricky.

So we can go in-- I would actually want to run the ANOVA test in that case. Here's my raw data. There's the model I fit, again, using the ANOVA table that we saw a minute ago. I could by hand form that, or I could put it in JMP and come out with the output. But now here, this is very nice, because I can look and say, OK, associated with the mean x1, x2, and the interaction x1, x2, what are the mean squares and the F compared to our estimate of the noise?

Remember, we had 40 data points. I'm estimating four model coefficients. That's leaving 36 replicate degrees of freedom, so I have a very good estimate of noise in the system. That's the denominator in all of my F ratios. And I'm seeing F's from 10,000 down to 77, and my F critical for 95% confidence is 4.1.

So it looks like all of my F's are much bigger than that. I have high confidence that the mean is significant, the main effects are significant, and the interaction is real. It is bigger than I would expect, given the noise in the process. So I would accept all of those model terms.

So so far, we used the idea of replicates to ask the significant question. We said our residual could have two sources of error in it. One is pure replication noise, but the other thing that we haven't really assessed yet is, is the model form adequate? You might still be nervous that maybe there's a quadratic relationship between output and the punch depth.

With those two data points, the best I could do is get a linear model. How would you possibly check to see if there might be curvature in the model, or a quadratic effect? Do you have enough data to do that?

**AUDIENCE:** There could be a center point.

**PROFESSOR:** What?

**AUDIENCE:** There could be a center point.

**PROFESSOR:** There could be a center point. Let's add a center point. Let's consider additional experimental center points. OK, so what you would do then is run some additional things at the center level, the zero setting of those two data points, and then look at the deviation between what the linear model suggests and what the actual data comes up with.

Now, you've got to ask the significance of the delta. Again, your model may not be perfect in predicting what that center point is, even if there is no curvature in the model, because there's noise also in the model. So we have to do two things. One is we have to estimate the deviation from our simpler model-- the linear model-- versus the estimation with the quadratic term that would go through, say, the center point-- go through that additional center point, and then compare the size of that to the natural spread in our data, and do essentially an F-test or an ANOVA test on that quadratic model term-- same thing.

Anything bother you about this picture? We can set up and do this center point, and run it. So I'm down on the two-all. I'm setting the depth factor. That's my x to factor, I guess. And I put it at 0.45 inches. I pick my part. I got my aluminum. I've got my stainless steel. I need my 0-th level. There is no center point on that, right?

So you'd have to actually be very careful. If you've got only discrete values that don't have some ordering or some continuous degree to them, you can't do a center point. So what would you do in this case?

**AUDIENCE:** Physical model--

**PROFESSOR:** A physical model-- it's so complicated, it's hopeless. We don't have a physical model. You might try a physical model. There's something simpler you could do, still using design of experiments. Yeah?

**AUDIENCE:** You just use the easiest one.

**PROFESSOR:** What's that?

**AUDIENCE:** Just use the simplest one.

**PROFESSOR:** Just use the simplest point?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Maybe-- I'm thinking, what additional experimental point should I run to be able to decide if there is additional curvature? Remember, we were asking the question, is there curvature or a higher order term associated with what variable? Just punch depth, right? So why don't I still pick this, but now I can pick data points right here, and I might also pick data points right here?

I don't have to pick central point. I could actually pick a couple. In fact, I could even just do the runs right here and say, this is what the linear model would predict. But if there's curvature in my surface, then I'm going to get a deviation, just at the stainless steel setting. Now, if I were careful and trying to keep things balanced, I might do the stainless steel and the aluminum, and get an estimate across both of those at that deviation.

So essentially, what we're trying to do here is add a data point and then look at a lack of fit test in particular, where I want to say, is the model error-- the deviation because of this extra higher order term-- is it explainable by pure replicate variance, or is it larger than that, and therefore not explainable by my noise factor?

If we've got more deviation because of the quadratic term-- say, if I'm adding an x squared-- in this model, an x2 squared term-- then I have to reject the null hypothesis that says, I've just got random variance [INAUDIBLE]. I really need an extra model term. So we can still use the basic ANOVA setup.

So this is our model with higher order terms. You might think about trying to add two higher order terms. I would argue, if this is the material type, it makes no sense to add that term. I wouldn't do that. That would be bad. But if this is associated with the punch depth, then I think it does make sense to add the hypothesis of that. And then you check for deviation at one or more of those points.

And the basic question is asking whether beta 2, 2 is equal to 0 or not. Is the effective due to that deviation plausibly the same as noise or not? So we can go through and do that. Here's a simple test, simple example. And this is aggregating what we saw going back, in fact, to some of the earlier things we talked about of, if I don't have replicate data, what can I do?

If I just do my full factorial four corner point test, no replicates, and I use the full linear model with interactions, I cannot test for significance and I cannot test for lack of fit. That's the most bare bones situation you've got. I've got a perfect fit. But you better be awfully sure that there's not noise going on in your model or in your system.

Even without replicating my corner points, I can still step back and ask the question about lack of fit by adding center points. But if I just add one center point-- one replicate at the center point, can I really ask the question about the significance of the quadratic term? No. I would be able to perfectly fit the quadratic term with that one extra data point, one extra model coefficient.

So you have to have replicates somewhere in the system. And one typical approach is maybe I don't want to have 10 replicates at each of my corner points. I'm just going to add replicates at my center point. So now, with the addition of center points to potentially check for curvature and replicates, I've got an estimate of random-- a replicate noise and deviation from the linear. So together, by adding replicates at the center point, you get a lot of power.

So then you can use the ANOVA approach. And this is just working through and saying, I estimate the quadratic term or the deviation and sum of squared deviations from that, form an estimate of the variance because of that, and compare that to the replication error. Here's just working that out for our example, where I've got the average of my output from the linear model, the average at my replicate-- if my overall grand average is a little different than the average coming from my center point replicates, that's the deviation due to the hypothesized quadratic term. I can form the variance of that, and then compare that to my underlying estimate.

OK, and then this example I'll let you work through. This is just doing that for a simple run, where, again, I have just one run at each of the four corner points. This is a different data set, by the way. This is not the break forming anymore. And then I've got five replicates here at the center point. Question--

**AUDIENCE:** Are we assuming that the application error and the center point is going to be the same as the--

**PROFESSOR:** Yes. So we are making the assumption here-- and it's the same one we've made pretty much in most ANOVA analyses-- that the variance is the same at all run points. In other words, we have the n 0 sigma squared assumption, where sigma squared is operated at the center point as well as all the corner points.

So this is a neat example where, again, we're trying to estimate main effects, perhaps interaction effects. We've got an AB interaction column here. So just using the corner points, we can fit a model. We can ask the ANOVA question on significance of that. But I can also form the sum of squares due to this quadratic term and use the replicated information there to be able to form an estimate of the overall process variance. And what that lets me do is form the F-test line in here for the quadratic term, in addition to the AB and the interaction term.

And in this case, when I look at that F ratio, the sum of squared deviations from the quadratic estimate versus the linear estimate compared to my estimate of noise in the system, I get a very small F, well below the critical F. And therefore, I conclude there is no evidence of lack of fit. I do not include the quadratic term, and I stick with the linear model.

Now, would you keep the interaction term? Same thing-- it's also not significant. Would you keep the two main effect terms? Yeah, sure-- those are above the F0, so I'm happy with those. So this is a nice, simple data set-- additive effects, no interaction. This is the best for ultimately using the model in a control kind of setting, because I can presumably change A and B independently, and try to get the output that I want.

**AUDIENCE:** Would the f-test be quite close to the boundary? Or do we have to do sensitivity analysis to see-- [INAUDIBLE] to 96% or 97% [INAUDIBLE].

**PROFESSOR:** Sure. You always got to be a little careful and use exactly that knowledge. This was a 95% confidence, and if I was anywhere near the boundary, I might say, well, my-- I want to include that if I wanted to just be 90% confidence, because that's probably suggesting that that model term really does have an effect. In this case, that cut-off-- those are very obvious. But in other subtle cases, you actually have to use some intuition and knowledge.

One part of that does look at-- maybe you're getting at this a little bit with sensitivity. There's another sensitivity analysis you can do, which has to do with not tweaking the alpha level, the significance level-- you can tweak that a little bit, but you also might ask the question slightly differently about whether the model coefficient is real or not.

That's a significance question. You can also ask the question, how big is the effective? Is it so big that I would want to include it in the model? Or it's real, but it's so tiny compared to my other effects that I'm going to neglect it. So there is this difference between significant, real, and important or not. And so that's where your knowledge of the process also comes into play, and whether you might want to include or drop model terms.

OK. I think what we'll do is I'll go-- I might actually only go for about five more minutes and give you a little bit of a glimpse of blocks and confounding, talk about these nuisance factors a little bit. And then next time, we'll come back and dive in more on fractional factorial designs, because there are some additional design issues I've only alluded to in experimental design.

So far, we've really just done full factorial-- these corner points-- and then also the idea of center points to ask the question about curvature. But there may be other factors. Maybe there was the temperature in our break forming tool that we didn't really want to do a design experiment on.

We're not really sure it's important or not, but maybe it's out there. There may be all these other nuisance factors or nuisance effects. They may affect the output, but I didn't really want to include the effect in the model. And I'd rather suppress them, and for sure suppress them from making-- from me making a mistake and confusing the effect of temperature with, say, the effect of which metal I use.

So what are the approaches you can do? Well, in many cases, I may not be able to run the whole experiment, holding that factor constant. In general, you would really want to make sure, well, if temperature affects things, I want to hold the temperature of the tool as constant as I can-- try to keep everything else it's constant as possible.

Time can often be an important substitute for nuisance effects. Time is going on, and other things may be changing. So you can't entirely suppress potentially all these other nuisance effects. But in general, you want to keep everything else constant.

But some cases-- you may not be able to hold everything else constant. And if you know some of these factors-- like temperature perhaps throughout the day varies in my machine shop, say-- then it may be uncontrollable. But it's a known factor, in which case, I might want to think about that and randomize the order of my runs against that.

So I equally include some high levels of that nuisance factor and some low levels of that nuisance factor in each of my other settings. So I try as much as I can to randomize, basically convert that effect from a systematic effect into a random noise effect.

In many cases, if I can actually pick the level of temperature-- it's known and controllable-- you can actually design your experiment, separate out the data into blocks where, in each of the two blocks, the nuisance factor is held constant. And what we have the opportunity to do is transform the effect of that nuisance factor into what would only be a very higher order interaction term in our model.

And that's the intuition I want to just give you a little bit of a glimpse for, and then we'll talk about next time. First off, if the thing is controllable, we would just want to, say, replicate and randomize. Here's an example where we're looking for the hardness of four different test samples, but I'm doing it with an indentation test, but I've got four different tips that I'm using for the indentation.

I don't really care about the tip factor, but I'm worried that it might perturb my estimate of the hardness. So you can, first off, basically think of it as I'm just going to run with all four tips and take the average, and that's my best average of-- or estimate of the hardness in the four cases. So I'm just replicating around that nuisance factor-- not including it in the model explicitly, but on randomizing or replicating around it.

Many cases-- you don't have the opportunity. It's too expensive to do lots of replications. And what I want to do is arrange things so that, if there is a nuisance factor, it will have the least damaging effect on my model. So here's my one little intuitive example.

Suppose we've got this 2-by-2 design. So I've got two factors, two levels doing four runs. But let's say I have two arrange it where I do two runs at low temperature and two runs at high temperature-- have to do two runs in the morning, two in the afternoon, and during the day, my machine job heats up. That's just the way it is. I've got to live with it.

Which two runs should I pick to do with the high temperature, and which two runs should I do at the low temperature? Suppose I did it like this, where this was my low temperature and this was my high temperature. Any problem with that? Big problem with that-- if x2 has an effect, I can't tell whether that's a temperature effect or and x2 effect, because I did both of my x2 runs at the low temperature and both of my high values of x2 at the high temperature.

I don't know whether it's a temperature effect when I form that contrast or an x2 effect. Is there a better way to arrange or pick which two lines I want to do so that I don't get confused with the main effects in the experiment? And the intuition here is yes, there is a better approach. What if I pick block 1 to run the 1 and AB combinations?

So what if I pick block 1 to run-- what'd I say? The 1 and the AB? 1 and AB-- notice what's nice about that. The 1 and the AB-- so make that block 1 and make this block 2-- notice that neither block is corresponding to both high settings or both low settings for either factor.

What, in fact, the blockage is actually lining up with is the x1, x2 interaction factor. I don't want to get confused about whether there's a main effect or this nasty nuisance factor is having an effect. What I would am willing to do is be confused, or confound between my block effect and my interaction effect.

I won't be able to tell if the interaction is real or not, but that's OK. I believe it's not, and I'd rather mix in my block factor. So this gets to this idea of forming patterns-- or blocking, in this case-- to confound so that I can detect some things, and not others. And we'll talk more about that next time-- talk a little bit more generally about confounding.

And this actually leads into the main idea of fractional factorial design, where I may not want to pick all of my corner points. I might subset some of them, because I'm willing to be confused about whether there's a fourth-order factor that's also the same detection or same contrast with a second-order interaction factor. I don't care. And therefore, if I don't care, I can actually reduce the number of experimental points we want to do. So we'll talk about that next time.