MITOCW | Lec 11 | MIT 2.830J Control of Manufacturing Processes, S08

PROFESSOR: OK, I think we'll make a start since it's almost 10 past. Can everyone hear me in Singapore? Wave if you can. No? Oh, yes you can. Excellent. Right, I think that there's a couple of announcements before we start.

I'm going to have office hours tomorrow between 5:00 and 6:00 in case you have any questions about the current problem set, which is due on Thursday at 5:00 PM MIT time. And we are marking the quizzes as fast as we can, so you'll have them back soon. If you have any comments about the quiz, how you found the questions and so forth, then I'd be delighted to hear them. Any questions? OK. Singapore? No.

So anyway, today I'm going to introduce a tool that's going to be very helpful to us as we look to have ways of studying a process and understanding exactly how its inputs relate to its outputs, so that by doing that, we can eventually model the process, control it better, and improve its performance. And that technique is called the analysis of variance, or ANOVA. I want first to quickly review the tools that we've looked at in the course so far so that we can put ANOVA into perspective.

And we talked for the first couple of lectures about a systems view of manufacturing processes. We talked about a variety of parameters that were important in determining the outputs, some of which we have control over, some of which we don't-- some of which are actually disturbances or inherent properties of the equipment or the material that we're processing. And so we have this idea that there must be some relationship between the inputs and outputs. But we've actually focused on developing tools that just focus on the output-- on interpreting the output, the geometry, some other property.

And that's fine. That's an important tool. And here are some of the tools we've looked at for statistical testing purposes. There are, of course, the T and the F tests. And we looked at control charts, which are essentially running hypothesis tests that a processes in control. We're testing the hypothesis that the process mean, or variation, is of a certain value. We looked at cumulative some charts, and we looked at moving average charts. Then briefly, two lectures ago, we looked at Chi-squared and T-squared charts.

So let's look at the properties of those things just to remind ourselves. The key thing with the hypothesis tests and with the control charts is that what we're doing is we're interpreting changes in one particular specific output variable. So whether that be diameter, or threshold voltage, or something, all these hypothesis tests are comparing sample mean, sample variances based on one output variable. And with the Chi-squared and the T-squared squared charts, we introduced the ability to look at two outputs which might have different dimensions, and which we expect to be somehow interdependent. So we're allowing ourselves to interpret two different outputs.

Thinking about the number of samples that you can interpret with any one of these given tests, well, with a T or an F test, you're taking two samples and you're comparing them. So in the case of a T test, you're asking the question, is there evidence that there's a significant difference between the means of the underlying distributions of the two samples you've taken? And with the F test, we're saying, is there a significant difference between the underlying variances of two distributions?

With a control chart, well that's just a running hypothesis test. So you take many samples, and you're constantly doing a hypothesis test to see whether the process is at the place where you want it to be, so some predetermined operating point that your content is optimal, I suppose And with the Chi-squared and the T-squared charts, again, the same is true-- many samples.

But when we come to ask, well, how many inputs were we dealing with these things, we don't necessarily have any idea. We might be told that some input to the process was changed, here are some samples, is there evidence that the input has had an effect on the output? But sometimes, we might just have samples and have no clear understanding of what the relevant inputs were. So there's nothing inherent in these tools that allows us to get the relationship between inputs and outputs.

And similarly, we don't know how many levels or different values any given input to a process might take-- how many voltages we set a particular control voltage to. And so I've said that they're unknown. But in a sense, I put two with a question mark. Because you might think, in a very clearly defined T test, say, you might say, well, an operator varies one input. And here are some samples. Did the output change? So in that case, you'd have two samples-- two values for a particular input. And that might be the most sort of precisely defined type of test you can do.

We've also talked about yield modeling, where we're starting to try to get some physical concept of the link between properties of the process-- the actual nature of the process-- and what comes out at the other end in terms of yield; and process capability, where we're trying to ask questions about how well the process performs based on being centered at a particular output mean. But those tools, again, don't necessarily let us model the process, or indeed give us any insight with which to improve it.

So we need we need some tools that will start letting us do that, especially if we look at some typical data. Now you'll recall, this set of data, this was used in problem set two. This is from an injection molding experiment done at MIT for the course of a couple of years ago.

And so what we've got here is an output, which is diameter. We've got one output that we've measured, but we've got two input parameters that we think are relevant. There may be others. And in fact, there almost certain they are. Because when we look at-- when we look at the sample of parts with a given combination of these input parameters, velocity and hold time, we see that there is substantial variation. And that is either due to inherent random variation in properties of the materials, or some input variable that we haven't fully controlled.

But we've got two input certain parameters, and two values that we happen to have chosen for each of them. Now, we could start to look for significant differences between the means for each setting by doing a whole raft of T tests or F tests. But you know, that would become tedious quite quickly, especially if you're starting to look at many variables in many settings. So we need a more systematic way of going about this, and eventually being able to say which of the inputs is important in determining the value of the output.

Eventually, once we have those tools, well, there are a few aims that we might have. And first among those is developing a process model. So we want to be able to relate inputs and disturbances-- so inputs that we can't control-- to the outputs. And we want to know which inputs are relevant-- which inputs, when they vary by a reasonable amount, actually have a significant effect on the outputs that we're interested in understanding.

Once we've done that, we can think about optimizing the process. A target that we might have is to maximize CPK, minimize the cost associated with deviation from our target output value. We might have particular models that allow us to look at the impact of, say, a candidate process change to the mean outputs, to the variance of the outputs. These are all saying essentially the same thing, that if we can model a process, we can control it better.

If you had perfect understanding of the physics surrounding the process that you had, you could work forward from first principles. And you could think about every physical phenomenon going on in your system and predict what the output would be. But we don't. We don't have understanding of all the variables that are at play. And what this means is that to get control of a process, we need to work backwards from the physical output. We need to take measurements of what the process is doing, and model backwards to understand what the inputs do.

So when we're doing this kind of empirical modeling, there's some questions we need to ask ourselves. What are we trying to achieve? Do we want to minimize quality loss, maximize CPK? Do we just want to reduce the variance for some specific reason?

We need to define our system very carefully. What is the output? Are there multiple outputs? Are there multiple inputs? What do we want to vary? What does it cost us at least to vary, either in exploring the process space, or actually what does it cost us least to control in the factory? And we need to come up with some sort of model-- some sort of of form for the model that we can use. We haven't yet described any particular model. This function phi we've been talking about in previous lectures hasn't been specific to any particular process.

What's also important is to think about how easy it is to collect data from the process. If you have 20 variables and they can-- each vary over a factor of ten, one needs to start thinking about how one can simplify the process of doing these experiments and working towards a model, especially if there's a lot of measurements of the outputs involved, a lot of laborious measurement. So going back to that chart of the tools we have so far, what we're going to do today is add a new one. And that is the analysis of variance.

Now, the key advantage of this is that we're now starting to be specific about looking at the actual inputs that we're interested in and how they impact the outputs. And so we might just be interested in-- we might just be interested in studying one input. But equally, there are ANOVA techniques that will let us study multiple input parameters. Is there a question in Singapore?

AUDIENCE: Yeah, I have a question.

PROFESSOR: OK.

AUDIENCE: What is a Chi-squared chart, and why does it have two outputs?

PROFESSOR: Right, well this I think we covered in lecture two-- two lectures ago, certainly the T-squared chart. What that does is, if you have two output variables that you are maintaining control charts for, but they are functionally interdependent, you might not know exactly how, but you know that there's some relationship between them. A part that is out of specification may not trigger an out of control alarm on one of those control charts. But if you plot both of them together and look at the results in conjunction with one another, that allows you to infer more about whether the process is in or out of control.

So it's covered in Montgomery. We may have a problem set question about it as well to help you develop understanding. Anyway, I'm sorry. I want to focus on ANOVA today.

Anyway, so we have the option of looking at more than one input and taking more than one input value for each input. And when we have different levels for an input, we're going to refer to those as treatments. So a variable may be treated in one of a number of ways.

The outputs-- well, we're dealing in this particular case with one particular quantity that we're going to measure and analyze. And we're going to have two or more samples that we are able to deal with. So this is what we're going to do. We're going to look first for the ANOVA technique applied to one input variable. And then, we're going to work through an example of that. Then we'll look at multivariable analysis of variance where you have-- in this particular case, we're going to look at two input variables and how we look at the interactions of those two variables in setting the output.

So here, we're going to start with the single variable ANOVA. And this is a diagram that shows to you what might happen to the output of a process, which the output being this axis as we vary one particular input parameter. So we're assuming we have good control over this, and we can command that input parameter to be whatever we choose.

In this particular case, we're saying, well, that there are three possible values we might be interested in-- A, B, and C. And so these are the three treatments. And what we've shown here, schematically, is the distribution of the output variable under each of those three conditions. We're assuming that the variation in the output for each particular treatment is normally distributed, but that the mean of the output is determined directly by the value of this input parameter.

So there's a one-to-one deterministic relationship between the mean and the value of the input. So these tau values-- tau 1, tau 2, and tau 3-- these are actually determined directly by the populations. But then, if we imagine doing experiments under these three conditions, we're going to take-- we're going to set the input to the value for sample A, say. And then, we're going to take a number of output parts, measure them, and those outputs will have-- they will have a sample mean, y1.

And there will be, for each of these samples, a sample mean. We can take all the data together and come up with a grand average for the collected data. Then the differences between those individual treatment sample means and the grand mean will be estimators for the values of tau-- the effects of the treatments, the different treatments that we're doing.

So here's what we're doing. We're considering multiple settings for some variable of interest. And there are these real effects, deltas between the output mean for different conditions. And we've observed these samples.

So the question is, based on these observations, we want to be able to see whether the observed differences in sample means are real. So we're going to do some kind of hypothesis test that says, the null hypothesis is going to be changing the input variable doesn't make a blind bit of difference to the output mean. And the alternate hypothesis is there is a significant difference, which, later, we want to model somehow, probably.

What's important in ANOVA is this assumption that the variance of the output for each treatment is equivalent. And here, we've written it as sigma sub-0 squared. So the width of each of those subpopulations is-- has to be assumed to be equal. And when one is doing ANOVA, that's something that one needs to check. And we'll talk about that a little later.

Now, I'm going to describe the underlying assumption that allows the analysis of variance to be done in this way. And this underlying assumption is that you can model the value of a sample part as the sum of three quantities. First quantity is the process mean, around which the output means of all the possible treatments center.

Second is the effect of the treatment. So that's how far is the output mean of set when you have this particular treatment, tau, taking place, whether there's a certain voltage setting of the inputs, or whatever. And the third is this residual term, which we write with epsilon. And that is describing the random variation in the output beyond the systematic effect of the treatment. Sometimes we take this mu and the tau sub-t together. We write it as a mu sub-t. That's the output mean for a particular treatment.

So at the bottom of this graph, I've just sketched that out to show what I mean. Here is the process-- overall process mean. Here is the output mean for a particular treatment. And say, if we take this particular sample, for example, then there's a residual. There's a difference between the treatment mean and the actual value of that sample, which is epsilon sub-ti. Sorry, that squared shouldn't be there. So the sum of these three quantities is thought of as giving us the output quantity.

Now, I'm going to describe the three steps that one is going to go through to test this hypothesis that changing the value of the input has no effect on the treatment means. And the first step is to come up with an estimate of the underlying population variance. That involves looking at each treatment sample in turn-- so looking at this sample, then this sample, and looking at the sample variance individually for each sample, and then taking a pooled estimate of the random components of variation based on all those sample variances. That's the first step.

Second step is to try to look at the between group variation. That means looking at the sample means, which I've denoted with these horizontal lines here, and looking at the variance of those, and making an inference from that about how much variation there is in the output as we change the input setting. Once we've got those two estimates-- within group variation, between group variation-- we want to compare those estimates, and infer from that whether there is a difference between the different treatments.

If there is a difference, then we're going to expect the second estimate that we make to be larger than the first estimate. And just by looking at this schematic in the top right of the view graph, you can perhaps see intuitively how that might be. The variance associated with the between group changes is going to be larger if there's a significant impact of the input on the output. But if not, then those two variance estimates should be equal. There's no effect had by the input.

I'm going to go through each of those steps in turn, and go through the mechanics of how we're going to do that. I've already said that we're assuming that the output for each group-- and by group, I mean the part sampled for a given treatment, so for a given setting of the input variable-- we're saying that that variation is described by a variance of sigma 0 squared, and it's the same for each group, which may be true or may not be. In which case, we need to go back and think again.

So the second bullet point here is giving us sum of the squared deviations for the teeth group. And so this is just saying, we have a sample value for the output-- y sub-ti. We've taken a mean for the group, y bar sub-t, which is here, and we're summing the squares of the deviations between these values and the sample mean.

Once we've done that, it's a simple step to go to the group variance. So this is an unbiased estimate of the group variance. The sum of the square is divided by the number of degrees of freedom in that treatment. So that's the number of parts sampled minus 1. And the minus 1 is because, well, the value of the mean-- y bar sub-t, is dependent on all the individual y sub-ti's. So there's one less degree of freedom than there are data points.

Once we've got each of those group variances based on the data, we're going to take a pool estimate of the common within group variance. So inherent in this, again, is the assumption that the variance could be modeled as being equal for each group. And actually, that should be a sub-2.

But what we're doing, this is effectively taking a weighted average based on the number of data points that we have for each group. So we might take more samples for a particular treatment, but we can deal with that by using the correct number of degrees of freedom for each treatment. And what we get out of that is this pooled estimate of [INAUDIBLE] variance s sub-r squared.

So that was the within group variance. And now, we're going to talk about the between group variance. As I said, we're testing the hypothesis that varying the input does not cause the output mean to vary. If that were true, then what I've sketched on the bottom left here would be the case. You would have three populations that looked absolutely identical. And it wouldn't matter what value you chose for the input. You'd get the same output.

However, when you sample, we know that, of course, because there's random variation in the output, the sample mean will not always be the same. And that will vary. The sample mean will, itself, be normally distributed with a variance of sigma 0 squared over n, where n is the number of samples in-- the number of data points in the sample. So this is looking a little bit-- on the bottom right-- a little bit like a control chart, where the y-axis is could be thought of as being linked to sigma 0 squared divided by the sample size.

And so what this all means is that if the hypothesis was true-- inputs don't affect the output mean-- then we could form a second estimate of the within group variation by looking at it this way, where we are taking the treatment mean, which are these values here. We're subtracting the grand mean of all the data from each treatment mean in turn. We're taking the square of those deviations. And this n sub-t here is counting for the fact that, because of the central limit theorem, the variance of the sample mean is inversely proportional to the sample size, and k is the number of different treatments. So in this case, k is 3.

So we're making this estimate, s sub-t squared, which, in the case I've sketched here, we would expect to be equal to s sub-r squared. However, if that isn't the case, if there is a significant relationship between the input and the output, then s sub-t squared will be larger than s sub-r squared. And that's the quantity by which it's larger-- this sum here. And this value, tau sub-t, is the real difference-- the actual systematic change in output mean that occurs when we vary the input. So we're causing an inflation in the value of s sub-t squared based on the fact that changing the input changes the output mean.

So we've got these two estimates-- that within group variation and the between group variation. We want to look at them, and compare them, and say, is there a significant impact on the output when we go through these different treatments? So we're comparing two variances. And that means that what we're interested in doing is an F test.

So we're asking ourselves, how big is the chance that, if the null hypothesis were true, we would observe these two variances? We're going to make the test statistic the ratio s sub-t squared divided by s sub-r squared. So that's going to usually be a value greater than 1. And we reject the null hypothesis if that value is significantly greater than 1.

What it's worth remembering here-- go ahead.

AUDIENCE: When you say significantly, you mean factor of ten, factor of--

PROFESSOR: Well, we're going to do an F test at a chosen level of significance to work that out. And that's exactly what this slide says, so good question. We have this F statistic. And we're now going to interpret that. We're going to pick a particular significance level, which we want to say, is it-- at the 5% level, is it significant that there's a relationship between the treatment that's chosen and the output mean?

It's important to remember that this is a one-sided F test that we're interested in doing. The possibility we're considering is that sigma sub-t squared is greater than sigma sub-r squared. The case where the real value between group variance is less than the real value of the within group variance is not something that has a physical meaning. So it's a one-sided F test where the statistic we're expecting to be bigger than 1-- much bigger than 1.

And here are the degrees of freedom that we need to use for that evaluation. And that's, based on what we've learned so far, fairly straightforward test to do. We can also make this additional estimate, which is based on the sum of the squared deviations from the grand mean among all samples. And that can be useful in a number of ways, although it's not central to evaluating this F statistic and testing the hypothesis that we've described.

This slide is pretty important. So this shows you how one usually would lay out an ANOVA analysis. And we would put real quantities in the spaces shown by expressions here. We'd evaluate the sum of the squares between treatments and within treatments. And we'd figure out what the number of degrees of freedom was. We'd look at the estimates of our variances, we'd take their ratio, and we would then use F tables to find the probability that, by chance, that value of f sub-0 was observed if the null hypothesis of 0 treatment effects were true.

And it's worth remembering this word "residual." So before, when I highlighted that quantity epsilon, which is the variation associated with randomness, the thing that we're not trying to model in looking at these different treatments, that is accounted for by this sum of squares within treatments. What I'm going to do now is work through a very simple example to show how this single variable ANOVA can be done. And it will hopefully give you a feel for the steps that have to be gone through.

This sort of thing is automated in a number of programs. There are macros available in Excel to do it. But it's worth knowing exactly what the steps are so that you understand what's going on. What we've got here are three samples for each of three treatments of a particular variable. So this is the output that we've measured in some arbitrary quantity. And here are the treatments.

So for each particular treatment, we have three samples. We evaluate a sample mean. Here, it's 11. Here it's 8. And we take the grand average of all samples, and we know that as well.

We can also evaluate the sum of squared deviations of the sample values from the sample mean. And in this case, sum of squares for treatment one is the square of that difference, plus the square that difference, plus 0, because that particular sample lies on the mean. We do that for each particular treatment, and evaluate these sums of squares.

We know that there are two degrees of freedom for each particular sample. So we can go straight to estimates of the within group variances for each group or treatment. And based on that, we can evaluate that pool estimate of within group variance, which is the thing that's meant to be excluding the effect of the treatments.

Now what we're going to do is make the second estimate-- the between group variance estimate. And here, what we're doing is looking purely at-- we're interested purely in the sample means that we've evaluated. Take this value, this value, and this value. That sample mean is 11. The grand average for all samples is 10.

We're looking at that difference, and we're squaring that deviation here, and then we're scaling it by the number of samples for that particular treatment, which is 3-- 3 samples. And that's, again, dealing with the fact that the standard deviation of a sample mean is inversely proportional to the number of data points in the sample. So we do this for each treatment-- t equals 1, 2, and 3-- and we evaluate our estimate of the between group variance. So this estimate, in a sense, is trying to look-- well, folded into that estimate will be the effect of the random variation within the group and the effect of changing the treatment.

We have these two estimates-- sigma sub-r squared, sigma sub-t squared. Now, we're going to do the F test to see whether there's a significant evidence that changing the input changes the output. Here's how we might lay it out in Excel.

We've evaluated the sums of squares between groups within square-- within groups, sorry. We've got the degrees of freedom here. The mean squared value is just taken from over here. And the F statistic is just the between group estimate divided by the within group estimate, so 4 1/2. And then we can go to the tables and say, the 5% level, what's the critical value of f? It turns out to be 5.14. So in this case, we would-- the 5% level rejects-- accept the null hypothesis that there was no significant effect of inputs on outputs.

AUDIENCE: Is the ratio always larger than 1? How can it be smaller--

PROFESSOR: There's a chance that it will be smaller than 1, even if the actual output means were unaffected by the input. But again, that would be-- so if this is the value of f and this is the PDF-- I forget. This would be what the f distribution looks like. So there are values of f that are less than 1 down here.

But the 5% level, you would only reject the hypothesis that-- so f crit here is 5.14, the 5% level. You would reject the hypothesis that there was no significant impact of inputs on outputs if you-- this particular variable input, obviously-- if the value of f were greater than that. So yeah, there's a small chance that f will be less than 1. I think that's right. Is that right?

AUDIENCE: Yeah, certainly. If the observed f is less than 1, that would for sure tell you that the treatments are not having an effect. It means your treatment deltas are so small that in fact, you got lucky, your sample with 0 treatment effects in that case was of course even smaller variance than the in group variance, which happened purely by chance. So if you actually observed an f less than 1, exactly as Hayden said, you would reject the alternative hypothesis. You'd just have to say, yeah, there's no effect.

PROFESSOR: OK, well, that is the whole example. That's what you would do if you had nine data points, and you could write it down on one bit of paper. Obviously, not everything is that simple. Yeah?

AUDIENCE: OK, go back to previous slide. So the p value mean says if the r is larger than 0.064, we will reject the null hypothesis, right?

PROFESSOR: Yeah, the p value is the level of significance at which you would *just* reject the null hypothesis. So that's the level of significance for which f is equal to f crit. And in this case, it's 6.4%. So if the level of significance is 5%, that's a more stringent-- places a more stringent requirement to reject the null hypothesis than than 6.4%. So you would keep the null hypothesis at the 5% level.

AUDIENCE: So in fact, this is a good example where, if I asked you just directly the question, did the treatment have an effect, your answer is dependent on what level of confidence you wanted [INAUDIBLE] of the evidence tells you there was an effect. If I asked you, I want to be 95% confident there was an effect, your answer would be I don't have enough evidence. No, there's no effect. If I asked you instead, 90% confident that there's an effect, your answer is yes, there is an effect to 90% confidence. So you're right at that interesting point with that 6.4-- what is it?

PROFESSOR: Yeah, 6.4.

AUDIENCE: 6.4%. And if you just look at the scatter of the data, it's kind of fuzzy. You don't have a lot of data. You only have three data points in each sample. So that whole idea of confidence interval we talked about early in the term. It's very important.

PROFESSOR: OK, any more questions from anyone? Right. Well, we mentioned at the start that inherent in doing this analysis is the assumption that the within group variance is the same for every group. And that this actually should be a sigma, not a 3/4. It's important to check that to make sure that the ANOVA is valid. And there are ways of looking at the problem differently if you can't really make that assumption.

But what we also want to do is try to make sure that the analysis we're doing really does capture as much as possible of the treatment effect. And there are various ways of doing that. We can take the residuals-- in other words, the difference between the sample value and the sample mean. We can plot those residuals either against the time order in which the samples were taken. We could look at the distribution of those residuals, do a QQnorm plot or some other kind of plot. And we can make checks that this underlying assumption is reasonable.

So that's all I'm going to say about this. But I think it's an important thing to bear in mind. So ANOVA for one variable, though it takes quite a lot of effort to get it conceptually, the actual mechanics of doing it are fairly straightforward.

It's not always the case that there's only one input variable that we're interested in changing. We might, of course, want to do it with two or more input variables. There are good reasons why we'd be interested in exploiting a number of inputs to achieve the output we want. If we have several variables we can control, we can do things like controlling the output mean and the output variance to be what we want them to be. Or we can control the output to improve CPK, while at the same time reducing sensitivity to some other disturbance. And that is why we need to be able to do analysis of variance for multiple inputs.

One way of modeling the effect of these multiple inputs would be to have this simple additive model, where we are saying that the output value is the sum of four quantities-- process mean, as we mentioned before, and then two separate treatment effects, tau sub-t as we described for the one variable case, and then an equivalent value for another input, beta sub-q. And in this case, we would say that there were there were k possible treatments at first variable, and n possible treatments for the second. So you could have any one of the treatments for the first variable, with any one of the treatments for the second variable in combination. The fourth quantity is, again, this residual-- this random variation that's not something we're trying to account for with these two input variables.

And again, this is what we call a fixed effects model. So we're saying that there's a deterministic relationship between the input values and the value of tau sub-t and beta sub-q. There are cases in which there's a probabilistic relationship between the input and those treatment effects. There are ways of dealing with that that we won't describe here.

The model that is up on the board now is pretty simple. It assumes that the effects of these two inputs are additive. There are plenty of cases, plenty of processes where that simply isn't realistic. And there's some synergism between the two inputs.

An example that I can think of that's been be relevant in my research is in modeling the etching rate in a silicon plasma etching chamber, where there are-- I guess you could think of it as there being two really important input quantities, which are the flux of ions of reactant onto the surface of the wafer, the flux of uncharged fluorine radicals, which are the chemical species that are responsible for chemical etching of the circuit. So you have these two fluxes approaching the surface of the wafer.

And it's not just the case that the rate of removal of silicon is proportional to some weighted sum of those two fluxes. You can't really have etching unless you have a substantial flux of both ions and these fluorine neutrals, so that the ions provide the energy for the reaction to occur. And that means that the model I just described wouldn't get us very far. We need to be able to deal with an interaction between those two inputs.

This is how we can represent it. We can add in a fifth term, which is specific to the combination of treatments t and q. Let's look at how we might incorporate that into ANOVA. Here, we look at the within group variance for a particular-- for particular input variable. So this is for the first input variable. The treatment is tau. And then, we have the equivalent expression for what happens with the second variable.

Then, we have this interaction term, which is an estimate of the variance that's to do with-- that can't be explained by this additive idea, and finally, the residuals-- so taking the actual value of the of the outputs minus the grand mean, squaring those residuals. And what this leads us to is a two-way way ANOVA table where we can evaluate 3 F statistics, and apply 3 F tests to look for significant relationships between factor 1 and the output, factor 2 and the output, and for a significant amount of interaction between those two factors in setting the output.

Katerina, you had a question?

AUDIENCE: [INAUDIBLE] or whether [INAUDIBLE] graphic you were showing us earlier, with a [INAUDIBLE] Would this be between groups, or is it within one?

PROFESSOR: Yeah, this is-- sorry, this is between groups, isn't it?

AUDIENCE: Could you repeat the question?

PROFESSOR: Oh, right. Yes, Katerina was asking, are these estimates within group estimates or between group estimates? And yes, absolutely, you're right. These are actually between group estimates. So we're taking a sample mean and looking at its deviation from the grand mean for all data. And we're making an estimate, based on that, of the between group mean. So this is folding into it some of the effect of varying the input. Question in Singapore?

AUDIENCE: Yes, for si squared, shouldn't it be ytq minus yt, minus yq, plus y instead of minus y, according to a slide you showed just before this slide?

PROFESSOR: Did I show-- oh, yeah, OK. Yes, I'm sorry. That's a mistake. Yeah, you're quite right. Thank you very much. Absolutely.

Anyway, we have these estimates of the between group variances and the interaction variance. And we have these three F statistics that we can use separately to test separate null hypotheses that there's no impact to factor 1 on the output, no input of factor 2 on the output, and get to test the hypothesis that there's no importance in any interaction between those factors.

So now, I want to give an example of where this analysis could be relevant and useful. Now, we often think about the relationship between inputs and outputs being described where we're varying some inputs over time, and the output is changing over time. The output mean could be shifting over time. So we have a machine, and we're trying different settings for it.

But in fact, a lot of cases in semiconductor process control, we're interested in spatial variations. And you can think of the case I'm going to show you. This is for a metal etching process where we're-- some work that we've started with one of our collaborators. The idea is that we want to be able to model. And we're etching a metal layer to form interconnect wires to be able to model how uniformly that etches across a wafer, depending on how densely the features are packed, what their individual sizes are, and where they're situated on the wafer.

One problem that you can encounter when processing these metal layers is that the metal, you're masking a blanket layer of the metal with photoresist, and then applying a plasma to the wafer to etch the exposed metal away. But as that process happens, you can you can get sideways etching of the metal. And imagine you have a photoresist layer. This is a cross-section I'm sketching here. And you're etching a trench into the metal down to some insulating layer.

Agents have to enter this gap. Depending on the size of the gap, that transport process will vary. It will be harder in narrower gaps than in wider gaps for the reaction to get in, for the products to get out. But also, there's going to be some lateral etching of the metal. And the rate at which that lateral etching happens will depend on the availability of reactants in the region of that feature, which might vary across the wafer.

If there is this lateral etching, if it varies, then it's going to affect the final width of the wire, its final resistance, and therefore the speed at which an individual capacitor, parasitic or otherwise, in the circuit will charge. So you might find that if you can't reduce variation in this lateral etching process, the circuit properties of the devices produced will vary substantially across the wafer. So there are several things that can affect the availability of reactants at a given feature.

Firstly, there is the position in the chamber. And what I've sketched in the top left is a plan view of a wafer. This is, say, maybe one of several wafers sitting in a large plasma etching chamber. The gases flow through this chamber with some path, some velocity distribution. The design of the chamber will have an impact on the density of reactants and how it varies across the wafer.

So you might find that there's greater availability of reactants at the center of the wafer. That's because there's an inlet above the center. And so that's one thing that can affect the amount of lateral metal etching.

Then, you have the actual geometry of the pattern that is being etched. If you're trying to etch a large amount of metal in a given region of the wafer, that will act as a sink for reactants. There will be a lot of competition for these reactants. The concentration locally at the surface is likely to be depressed. And that's going to reduce the lateral etching rate for any individual feature. So we have this what we're going to call pattern density effect-- density of exposed metal for etching.

Finally, the thing that can affect the availability of reactants for this lateral etching is the size of the feature, as I mentioned. Narrower features provide a greater impediment to the transport of reactants to the side wall of the feature. So in a way, you could think of these phenomena as being input variables. Some of them you can control. If you wanted to, you could place constraints on what kind of density of features were available, what the smallest feature available was to the designer, these chips.

And to an extent, you can control the tool-related variation. You can choose a process that will give a more uniform distribution of gases in the chamber. You could redesign the chamber. And some of that variation you don't have an easy way to control. Some of that-- you don't have complete control over all the inputs. But what I'm saying is that this is a case where you have multiple geometrical input variations where you're trying to manufacture many identical chips.

Each square in this graph is one chip. And they're all supposed to be identical to one another. But because of the effects I described they will not, in fact, be identical. So you can think of this as being-- a wafer as being many interdependent samples of the output of a process where the input is varying in, in some cases, an uncontrolled way, in some cases a way you can control. So this would be a really meaty problem for multivariate ANOVA to deal with.

What we've got within each chip-- this is actually a test chip that was designed to do some experiments and build a model. So it's not actually a product. But within each chip, what we have is many copies of actually the same sorts of features. In fact, what they are is just snake-shaped wires which have a total length that amplifies any resistance variations caused by the lateral etching. You can go into the chip and you can probe the resistance of these wires. As the lateral etching gets faster, the resistance gets higher, because the wires are narrower.

So you have many of these snake features within each of these individual squares. And what we do is we surrounded those the same features by a different amount of padding metal, which is not electrically connected to these snakes but is sitting right next to the snake structures, so that they're perturbing the transport of etching gas. So in areas where there's a larger amount of metal exposed for etching, there's going to be greater competition for reactants locally, and there's going to be a lower etch rate, including a lower lateral etch rate.

It's not necessarily true, of course, that the pattern density effects are confined to one of these squares. You may find that the lengths over which competition for reactants occur are larger than the diameter of one of these patches. And that is something that would need to be modeled and dealt with in understanding the process.

So I'm not going to go through the analysis of variance for this problem. I just wanted to highlight the fact that there are these complicated sets of geometrical input variables that we want to try to understand often in semiconductor manufacturing processes. And very often, we want to just go beyond finding a functional relationship between some geometrical property and the performance of the products. We want to build a physical model that will work as far back as the settings on the machine, the flow rates of gases, the amount of electrical power that's going into generating the plasma in the chamber, to try to work back with enough detail that we can start to decide what good input variable values would be.

I'm just going to show you a little bit of data from one of these test wafers. You can see that we've identified a clear relationship between the pattern density-- the amount of padding metal within one of those sets of features-- and the average resistance of one of the snakes. There are several hundred snakes within each given patch of features. And what we've done here is just average the resistance of all of them to give a quick estimate of the effect of local pattern density and input variable on one important output.

So we can see that the pattern density has an effect that we could think of inventing a functional model for. But what we also have is the wafer scale non-uniformity. This is to do with the way gases flow around the chamber, approach the wafer, are transported across the wafer. And this graph shows you a subset of the data we have.

Down in the bottom left is a diagram of the wafer. And what we've done on this graph is plotted the resistance of a particular test feature, a resistive snake, within each chip on the wafer as a function of location. We slice up the wafer. And the x-axis corresponds to each slice of the wafer being concatenated. The first slice is here, the second slice is here, and so forth.

What we see is that the resistances that result tend to be larger in the center of the wafer. Here's a central part of the wafer, and here is an edge. So that relationship is clear. But then, if we look within each chip, if we look at the features that are near this 5% local metal density, then we see this amount of variation. And if we look in the region where there's a much higher amount of metal density, therefore less of the metal is being etched away, we see that the resistance is higher, and-- I'm sorry, by pattern density we mean the proportion of the chip that is open for etching. And we see that the resistance is higher because there's more there's more lateral etching.

But what this gives us an indication of is that there must be an interaction. Because the size of the wafer scale non-uniformity depends on the local pattern density. It's not as if we're taking the 5% variation pattern and we're shifting it up. When we change to 85% pattern density, there is a change in the shape of the location dependence as we change the local pattern density. And that is an example of one of those interaction effects that we would need to capture, so either through some multiplicative model or something more physically-based.

Anyway, that is the end of today's lecture. Next time, we're going to use these techniques, ANOVA, as the basis for starting to build real models where we're actually fleshing out the functional relationship between inputs and outputs, and designing experiments that will give us that information as efficiently as possible. OK, are there any questions from either side? Hello.

AUDIENCE: Yeah, I have one question. As for the ANOVA, you have, like, three parameters-- a k, n, and m, right? So does m always control n multiplied by k?

PROFESSOR: Actually, yeah, m in that case was the number of samples per combination of treatment. So actually, in MANOVA, the quantity that I termed m is a bit like n in the single variable ANOVA. I know that's confusing. But here, m the number of replicates for a given combination of input variables t and q. Does that make sense?

So we have a combination of inputs t and q. We keep them constant. We sample a few parts for that combination of inputs. And there are m of those parts.

AUDIENCE: OK.

PROFESSOR: Thanks. Anyone else? OK, well, thank you. The problem is due on Thursday. Let me know if you have any questions about it. I'm sure you will. Oh, hello?

AUDIENCE: Can I check in with you?

PROFESSOR: Sure.

AUDIENCE: For the quiz, I'm not sure which question. I think it probably be the last section of problem 1. I think I needed a [INAUDIBLE] table with r for equal to 0.025, but we were not provided with that.

PROFESSOR: Yes.

AUDIENCE: I'm not sure whether I'm wrong, or--

PROFESSOR: About ten people asked me at this end about that. Well, you know, I actually think that a one-sided F test would have been appropriate in that case. In which case, the table-- the 0.05 table would have been appropriate to do a one-sided 5% F test.

It was also possible to answer that question by looking at the confidence intervals from part B and seeing whether they overlapped. And I think they didn't overlap, did they? So they didn't overlap. So we said there was a significant difference. But anyway we'll publish solutions. Thanks, anyone else? No? Good.

AUDIENCE: One question.

PROFESSOR: Yeah?

AUDIENCE: I'm sorry.

PROFESSOR: That's OK.

AUDIENCE: For the MANOVA, there-- OK, for ANOVA there's a term that is the 0 mean normal residual. And for MANOVA, there's the same term. But in the second line, the term disappeared. Do you know what I'm talking about-- on previous slide?

PROFESSOR: On the previous-- this slide? This slide?

AUDIENCE: No, previous, previous slide-- two slides ago. Oh, yeah, this one.

PROFESSOR: This one.

AUDIENCE: The term disappears later. The last term on the--

PROFESSOR: Yeah, oh, you mean the term-- there's a term here, but there isn't a term here. Is that what you mean?

AUDIENCE: Yeah.

PROFESSOR: Yeah. Ah, right, well, the second line is an estimate of the value of the output for the combination of inputs t and q. An estimate is not trying to make any predictions about what was a residual will be. It's really dealing just with means for a given treatment. And it's saying, if, say, you set your machine to input 1 having value t, input 2 having value q, what's your best estimate of what the output will be?

And that estimate has to be the mean output. There's no point making anything other than a mean. And so the residual, that epsilon term, is present in real data, because there's random variation in the output around the expected mean.

AUDIENCE: Another way to say that is your best estimate of epsilon [INAUDIBLE].

PROFESSOR: Yeah.

AUDIENCE: So you could put, like, a-- you could put a plus 0 there if that's your best guess.

PROFESSOR: OK, did you hear that in Singapore?

AUDIENCE: Yeah.

PROFESSOR: Yeah, good.

AUDIENCE: Thank you.

PROFESSOR: Anyone else? OK, see you next time.