DUANE BONING: OK. As you'll see with today's lecture, I made a small change in the lecture sequence. I basically combined the last part of Tuesday's lecture with what I was going to talk about next week because they really go together better anyway, which is multivariate SPC together with a couple of the advanced control charting.

And so next week we'll do the yield modeling lecture rather than try to squish that in today. So what I want to do today is first have a very brief warm up. Just get us back into the swing of things with conventional control charts and then talk a little bit about these alternative control charts with a little bit of increased sensitivity for rapid detection of shifts, as well as drifts. But primarily targeted at shifts.

And these are the moving average-- exponentially weighted moving average and the cumulative sum chart. And then we'll move into multivariate SPC, which is a little bit of a tutorial on multivariate statistics in general. So that's the plan. So to get us warmed up we'll do a rapid design of a control chart. How to do this relatively quickly. I will tell you that we have a model based on lots of historical data of this process that when it is in control it's operating as a normally distributed process with a mean of 1 and a variance of 1.

We're going to be taking samples every nine parts. And then the next 9 and so on. So what should the-- here's the easy one. Wake you up in the morning. What should the center line be?

AUDIENCE: 5.

DUANE BONING: 5. OK. Everybody-- how about the plus minus 3 sigma? We're quite happy with the 1 in 370 ARL with normal typical plus minus 3 sigma control limits. Where should the plus minus 3 sigma control limits be? I hear a 6 and 4. Why?

AUDIENCE: [INAUDIBLE].

DUANE BONING: I'm sorry?

AUDIENCE: [INAUDIBLE]

DUANE BONING: OK. Which is?

AUDIENCE: Mu plus sigma [INAUDIBLE].

DUANE BONING: Mu plus sigma plus minus 3 sigma. 3 sigma over square root of sample size. And so the 3 sigma and the 3, which comes from the root [? end ?] make it a nice plus minus 1. And there are some data. OK. So I plotted 99. We ran the process 99 runs. We have 11 samples. Here's the question. Is the process in control?

We have a yes. We have another yes. I contend you don't know. It's a little bit tricky here.

AUDIENCE: Not the 1 [INAUDIBLE].

DUANE BONING: Yeah, I guess. That's actually not what I have in mind. You could go in and start down the path of looking and saying, OK, what's the probability of having, for example three rising points in a row? And all the other [INAUDIBLE]. But I'll say none of that's happening. So I will say that the mean-- there's nothing too strange happening with the mean that you can tell with this chart.

Is the process in control?

AUDIENCE: [INAUDIBLE] variance [INAUDIBLE]--

DUANE BONING: Yeah.

AUDIENCE: [INAUDIBLE] variance is getting huge compared to earlier points.

DUANE BONING: Oh. So you're saying the variance is getting larger.

AUDIENCE: [INAUDIBLE] variance-- yeah.

DUANE BONING: Maybe. It's hard to tell with this. This Is a mean chart. All it's telling you is the tracking of the mean. We said that the process model is that we have this mean and this variance. And so I guess the main point here is you don't really know if you're just tracking the mean. What's happening also with the variance within your sample?

And in fact, on the next page this is the raw data-- all 99 points where we grouped 9 by 9 by 9 to get the sample. And if you saw this, you might be a little less comfortable. Now you could always go back and look at raw data. And raw data is always really good to do. But what you would also want to do is a typical S chart or a range chart or something like that.

And in this case, if you design the S chart with those apriori pieces of data-- sample size is 9 and the process variance is 1-- with the C4 correction factor and so on-- all of that just out of the formula-- you get these control limits. And sure enough, at about sample 10 we're above the control limit. Sample 11 is even more above the control limit. And that's indicative of essentially the increased variance of those two groups.

And I will tell you, I generated this data with JMP. And sure enough, right at the whatever it was-- the 77th point-- I re-sampled from a normal distribution with something like a variance of 1.9 or 2 or something like that. So the main point here was you need to monitor both aspects-- both the mean and the variance to know whether your process model is still operative.

We're saying that for a normal distribution the mean and variance are really telling you everything. Now, the other sub point in here is that if you were actually looking at this run data, there is a lot of noise. There is inherent process noise. And part of the goal here is you're trying to detect a mean in the face of that noise. And so one way of thinking about sampling and calculating x bar is that, indeed, you are filtering. OK.

So you want to filter to suppress the additional variance effect when you're looking for the mean phrase. But again, you've got to also monitor for the variance. But this noise makes it a little bit tricky in terms of being able to detect the means. So you want to filter. And one of the ideas of these alternative control charts is explicitly think about the filtering action that you're doing-- a little bit almost from a more signal processing point of view, but not necessarily have to wait for an additional nine runs before you call that a sample, aggregate it, and then plot it, and then wait for another nine runs.

One alternative approach here is to take a running average. So I take a window, say, of 9 points and measurements, calculate my statistics, and then for each new run move my window over and recalculate my statistics so that I'm in some sense updating what my best guess of the current state of the process is after every run. And the advantages of this are that you can get more averages per unit of data. OK.

You can go ahead and use the run data alone if you wanted to. You could actually plot for example, the individual measurements and even do a hybrid sort of thing where you did a running average just for the standard deviation. But the real advantage of these is they essentially improve your ability to detect. They improve the filtering action potentially because you can play with that window size to filter more but also be able to react potentially faster.

So I'm going to describe a couple of these alternative charts. And I'm going to focus really on the running average detection of the mean. But the same idea also applies to filtering of the variance across multiple runs. So there's also running S charts or moving range charts and those sorts of things. So don't forget the lesson we just learned a minute ago. You got to monitor both mean and variance.

But for the next 10, 15 slides I'm only going to talk about running averages on the average. So one idea of the running average is a simple running average or moving average where we simply take some n values of the past and calculate say a simple average. And similarly, we can do that for the variance. So in that case, this simplest picture, this simple moving average is just an equally weighted window.

You pick your window size. Maybe it's a window size of 9 and you calculate a new moving average statistic, which I've labeled m sub i here. And so that's pretty trivial formula. Now how do you set the control limits on a moving average chart? Well, you want plus minus 3 sigma control limits based on the variance that you expect in your statistic. Right?

So previously we'd have sample size of 9. And we'd say, well, what's the variance? It's 1 over-- it's sigma over root n is the standard deviation associated with our normal sample size. And that same formula ends up working here because we've got equal weighting. Right?

So if we simply apply, again, assuming each of the points are truly in process control and I do not have correlation in time between them-- it's very easy to see what the variance is. Now you'll notice something a little strange right here at the beginning. This is kind of an artifact of the startup of the chart. Let's say I have just started the use of the control chart and I've only got five samples so far, but I've got a moving average window of 9.

Well, I don't have nine data points yet. I have to form my moving average really with a window size only of 5 so far. And because I'm taking the variance divided by 5 instead of divided by 9, the variance of my average of a five point sample is larger. And you can see essentially each sample until I get to my full window size of 9, has slightly larger control chart limits. OK. So that's what's going on there. OK.

And you can also sort of see here now the filtering action at work. Notice that each of these samples don't vary dramatically from the previous one. They are no longer uncorrelated from one sample to the next. In fact, with this equal weighting, every new time point shares 88% of the data with the previous data point. So there's a huge amount of correlation from one point in time in the next.

So you should not try to use things like the WECO rules associated with three data points in a row or four or five data points in a row rising. Those things are right out because this is a running control chart or moving average control chart. And you've now got correlation. It's benefit though is that it is doing that filtering. And you can take whatever current point that you get as a sort of your best guess of what the current state of the overall mean is.

And so you would trigger on alarms above or below the control limits. So simple moving average-- this is relatively simple. Any questions on this before we do a unequally weighted moving average? Pretty clear.

AUDIENCE: I have a question, Professor.

DUANE BONING: Yes. Good. These questions have been wonderful by the way. The last couple of classes, I think the questions have really been right on target. It's been a lot of fun. So we have high expectations for your question.

[LAUGHTER]

AUDIENCE: OK. So that's a lot of pressure, but my question is that for the window size on 9, so do the two windows next to each other-- they overlap by 8 data points?

DUANE BONING: Right.

AUDIENCE: OK. That's it.

DUANE BONING: That's exactly right.

AUDIENCE: OK, thank you.

DUANE BONING: Now, one problem with that overlap of eight data points is, by the time you get to 8 points in the past, it may be-- that's a long time ago, depending on how rapidly you're sampling or whatever. And compared to the data point that you just took, you might think that the data point you just took represents the current state of the process a lot better than the 0.789 time points ago.

And so there is a more general approach here, where you can pick some weighting for how much you want to fold in data from different points in time. And a classic way to do the weighting is to use an exponentially weighted moving average, where you might define the weighting function-- these a sub i's here-- I guess I've switched notation to w here-- when I'm at the current sample point or the current time point t, I might weight points backwards in time by index i, i going from 1 to 9 to 1,000-- whatever-- by a function looking something like this, with some r factor, where r is weighting factor.

So what happens here, as I might think of r as being how much I believe my current data point-- might be I believe my current data point by 90%, and then the next data point, 1 minus r, is only 10%. So I'm only mixing in 10% of that. And then the next data point beyond that-- it's maybe 0.1 times 0.1. It's only 0.01, so I'm only mixing in 1% of the previous data point.

So very rapidly, I have this drop-off as a function of data points in the past-- this is my i index-- so that I much more heavily weight recent history, and then exponentially decay into the past. But I still have this degree of freedom. I can pick r. I can pick the relative weighting of more recent data compared to old data.

Now, one way of looking at this is, actually, I have access to all past data, and I would apply this out to 10 data points, or 50, or 100 with the appropriate weighting function. And I guess, with computer systems these days, it would be no problem to actually apply it in that fashion. There's a beautiful thing about the exponentially weighted moving average, which is I actually don't have to keep track of all of the raw data.

There's another way of thinking about what the EWMA, this exponentially weighted moving average is doing. And that's that I have my previous best estimate of the state of my variable-- my previous best estimate of the mean-- and now I have one new data point, and I use that one new data point to update my best estimate. My best estimate had embedded in it all of the past history, so now, with the additional new weight, I can just do a simple update of my best state estimate.

And there is essentially a recursive formula that makes this explicit, and it's exactly the same, thing if you now apply it iteratively out to all history, as actually performing that computation of applying the weights on all past data. So that is to say, let's say, a sub i is my best estimate of whatever my statistic is-- my mean. Again, I might weight my current data point, my most recent data point, say, 90%, which would be really rapid update. I really believe my most recent data.

And then my 1 minus r applies to my previous-- oops-- to my previous estimate. So I'm believing only 10% my previous estimate, and 90% my new estimate. So it makes it a little bit easier in terms of the computation, but I think it's also conceptually nice in terms of what the action here is.

Now, one can go through and do a little bit of the mathematics. It's actually a little easier to do it on the full summation version, but you just do basic expectation math on the variance of an EWMA-- again, drawing from a normal distribution that is in control. And one can estimate the standard deviation of this a statistic.

And it's kind of a little bit messy. It's got the underlying process variation in it, the fact that you are still-- you can still be doing sampling. For each x sub i, that might be a sample of n data points. Very often, you might simply be doing it one data point, so n, very often, is just 1-- and then this formula with r in it, including this 1 minus 1 minus r to the 2t, where t is how far along you are.

So it's kind of nasty, but one thing that's nice is, once you've start started up your control chart and you're away from that little artifact that we saw with the moving average, where the variance limits change-- once you get far enough away from that start up, your t is large, and that this whole term just drops out-- drops out more fast when-- more rapidly for large values of r, because I'm really weighting more towards recent history.

But even for small or modest values of r, after you get out 10, 20 data points, it typically drops away. So now you've got an estimate for large T for the variance, sigma a of your statistic, and you can simply plot plus/minus 3 sigma, or whatever control limit you want, depending on your false alarm rate alpha. So this is pretty cool.

One quick observation here is, again, we still have a lot of degrees of flexibility around the choice of r. And if we look at this formula here, and the effect of r-- I sort of plotted this out in this next chart. This effect of r on the sigma multiplier-- oops-- sorry-- is this factor right here. So in other words, as r changes, how do my control limits move? What is this multiplier of my sigma over root n from my normal sampling?

And you can look at it here. And as r gets larger, well, that gets larger-- as r approaches 1. Typically, you don't ever go above 1. You can't believe your most recent data point more than 100%. It's a ratio, or fraction of the percent. You can see that, as r gets larger, this multiplier starts to approach some limit. And finally, when r is equal to 1, this multiplier is simply 1.

In the limit, when r is equal to 1, what are we doing? We're doing 100% update based on my most recent data point. In other words, I'm only looking at my most recent data point. My variance is simply the process variance coming from sampling. This all reduces to [? simply ?] sigma a being by sampling variance, whatever size sample I was doing within each formulation.

So in that case, I have no factor or no multiplier on that sigma A. But for other values of r, what's happening? For smaller values of r, what do we see? The sigma A is smaller than the sigma x over root n. We're getting reduction in the variance because of filtering, the variance of our estimate. So that was the whole idea is I can tighten my control limits, essentially by the filtering action.

So as I get smaller and smaller r's, I do more and more aggressive filtering with lots of past data. I get tighter and tighter limits, but of course, I'm also suppressing recent data little bit more heavily, so there is this trade-off. Let's look at some examples. Oh, well, I guess I made these points already here.

Again, the variance of this estimate for the mean-- your best guess of the mean-- is smaller than you would have had with an x bar chart, because the x bar chart, remember, for a sample size of n, is simply that. So we do get this suppression. And it is very nice that it also, in the limit, works out for both the sample size of 1 case-- these formulas still hold-- and also the full update case, or the r equals 1, where we have completely unfiltered data.

You just plot in your pure run data on your chart. So what happens if I use much smaller r's, like 0.1? I got much tighter control limits, but I've got long delays potentially then in detecting-- I've got a lot of noise suppression, but again, I get back to long delays in detecting or responding to a true shift in the mean. So I want to look at a few examples here to give you a feel of the trade-off between different r's for some data, and compare it as well to a typical x bar chart.

OK, so here's a set of data, and we're seeing a comparison here between the x bar chart-- so you can see here are the x bar upper and lower control limits. And sure enough, after 150 runs, where at-- I don't know-- about run 90, we make this shift in the process-- so we shift the whole process by a half a standard deviation of the underlying process.

And sure enough, by about-- I don't know-- run 140, we finally got a data point in the tail of the mean distribution-- the regular x bar-- that fell out. And we could actually calculate the average run length to detect a mean shift of 0.5 using some of the techniques we talked about last time.

And we might not be happy with that very long ARL, and the question is, would an EWMA chart help us with this? And this is an example where we're using about a 30% update based on each new data point, and plotted in purple with the blue and the yellow are the upper and lower control limits for the EWMA.

And what you see is the effect of averaging. Now, I always find it's easiest for me to conceptualize the action of the EWMA chart is just kind of like a moving average-- the equally weighted moving average, but weighted a little bit more towards recent history.

But the whole idea is I get that aggregation from any kind of a moving average so that, right after this shift right here, now, when I take into account-- with an r of 0.3, I'm taking into account three, four, five data points-- I've got kind of an effective window weighted about 30% towards my most recent, but including some past history-- that all of the plus delta nu data points all average, and then my normal process variation just cancels out each other.

And so it's very intuitive, I think, that this EWMA is going to aggregate or respond to or sense the delta shift in the mean more than an individual x bar chart would have been doing-- so that it responds in this case, I don't know, after about 15 or 20 data points. You can start to see excursions beyond the control limit. Question--

AUDIENCE: At the beginning of the chart, why are the control limits [INAUDIBLE] spreading [INAUDIBLE]? The last time, they were [INAUDIBLE]

DUANE BONING: Yes, yes-- good point. I think you can go back to this formula and see if that works. I bet, if you look at the effective the t action, formulaically or analytically, I think that, as t starts out being very small-- so if t is 0, this is essentially 1.

Well, let's say t was 1 and r was 0.3, I've got 0.7. I've got a reduction. So I think, for small t, I've only got a limited amount of data, and so I'm not-- I don't have the long tail in the EWMA. So I think I've only got a small amount of data with a larger weight in my control limits I think are smaller. But then, as t gets larger, I've got a little bit more and more data until t-- so I think if I were to plot this 1 minus 1 minus r to the 2t-- kind of has a funky shape that dies out.

But I wouldn't worry too much about the startup times. There are artifacts when you start up a control chart, but once you just start a control chart anyway, you're still feeling out the process, so I wouldn't worry too much about the startup times on these. Yeah?

AUDIENCE: [INAUDIBLE] How do you get this side [INAUDIBLE]

DUANE BONING: Yeah. So the question is-- if you didn't hear that in Singapore-- was, when would you use an EWMA or what are the factors in using an EWMA from an x bar chart? And I think, if you've got a lot of noise and you're really worried about detecting rapidly a shift in the process, an EWMA-- and we'll see the CUSUM in a minute-- has improved sensitivity to rapid detection of the shift.

But it comes at one really important cost, which is a loss in transparency in what's going on. The x bar, I think, conceptually, is a lot easier to understand, have everybody on the line understand what it is. The data, if you also explode out and show the full data-- the x bar and the s-- gives you a good feel directly to the statistics of the process.

Here you look at this and you lose touch, I think, a little bit with the statistics of-- and the relationship to your raw data. It's a little harder to get your mind around. I don't know-- 10, 15 years ago, I fell in love with the EWMA chart and approach-- not only for charting, but also for doing simple run-by-run control. And so I kind of like it, but I can empathize with it being a little bit harder to explain and maintain on the line.

It's also got a little bit more calculation, so back in the days of-- and it's still true in some of the manufacturing lines. I know, when we went on the tour and some of the plants-- very, very up-to-date plants in Singapore and other places around the world, people are still often taking measurements by hand, and doing the hand calculation, and writing it on a chart.

This has an extra step. You've got to do an extra average. So in modern day, when almost all the calculations are automated. And the data collection is automated. I think there's even more reason to go with an EWMA. So the thing I like about the best about the EWMA is I think of it as giving me my best current estimate based not only on just my last data point, but a little bit of filtering, including past history.

It's giving my best educated state of the current process. And why I like that for then run-by-run control is, based on my current model, my updated-- on a recursive basis, updated model of the current state of the process I can make a decision. So here the decision in SPC is the process in control or not

'But if you think of this being not just the EWMA not-- giving you more information. It's not just giving you, am I in control or not, given the random spread? But if you interpret this little purple data point here as my best estimate of where the mean in the recent past has been, if there is a little bit of drift or movement maybe. I do active control and not pure SPC control.

I'm not making any longer just the decision. Nothing has changed. I'm saying there are small changes, and maybe I want to go and make a small modification or an adjustment on the equipment. Now. If, you are in true statistical process control, that would be the wrong thing to do. You'd be over-adjusting, and we'll talk about that later.

But that's another thing I like about the EWMA is very often, there is a little bit of time correlation in the process. And the EWMA is tracking those small changes and filtering out some of the noise. It's a good question. I'll take a question. There was somebody had a question in Singapore.

AUDIENCE: Yes. So under smaller r will lead to the narrow control limit, right?

DUANE BONING: That's right.

AUDIENCE: That's so the narrow control limit-- you will have a smaller beta?

DUANE BONING: So a smaller beta would be less chance--

AUDIENCE: Smaller beta?

DUANE BONING: Yes.

AUDIENCE: Yeah, [INAUDIBLE] type II error.

DUANE BONING: Right-- but its power is then larger, so the chance of detecting a true out of control point increases. But there's a very important additional factor, which is there's additional time lag. So I'm going to show a couple of examples, because you can see in other examples the effective the filtering and the time response that it takes to detect can go up.

AUDIENCE: OK.

DUANE BONING: So another question here--

AUDIENCE: I'm was just curious. What sample size were you using for the x bar chart?

DUANE BONING: I don't know. I'm not sure what was used in the-- as a sample size. I think you're just seeing the plot of the sample, so the sample size is actually the same in both cases. But I don't what was being used. In some sense, it doesn't actually matter too much. The EWMA is looking-- is an EWMA over sample.

I might have 1,000 parts an hour, and every 15 minutes, I might do a sample size of 10. So I'm only getting sample size 10, lots more. And what's being plotted here, in some sense, is the sample. So I have an x bar where I might plot those samples. Here what I'm doing-- the EWMA has enough-- it has the flexibility to extend to n equals 1 sampling, if that's what I'm doing.

But in some sense, it could also be used where, instead of each of those n equals 10 samples was the only data I looked at. Each one of those is my t-- point in t, and my previous sample of size 10 is t minus 1. So in some sense, it's doing the whole sample-by-sample comparison.

So my point here was both the EWMA and the x bar are using, in this derivation, the same sample size. I'm not differentiating in that. All I'm doing in this case really, with the EWMA is using multiple samples back in time, and the x bar only uses one. So in fact, if I go to r equals 1, the two converge. They are the same. The r equals EWMA is exactly the same as the x bar.

AUDIENCE: But the larger my sample size is the less of a spread I get between my two-- [INAUDIBLE]

DUANE BONING: They would both go together, though, because I had to have the factor. So back here on slide 12, if I change my sample size, both sigma A and my upper control limits and lower federal limits for the x bar shrink exactly the same. So I would scale both of them, if I increase.

So what I was talking about here is with different values of r. So this was an r equals 0.3. Here's a smaller r, r equals 0.1. So what's happening here? I'm only accepting my most recent data point 10%, and mixing past history and 90%. That's more averaging of the past. I'm using more past history, and so in this case, what you see is kind of a more slowly drifting, more heavily-- excuse me-- more heavily filtered version.

And here, with the mean shift occurring about here, you can start to see in this case a big mean shift at 0.5 with, I guess, a sample size of 5. You do need to know the sample size in order to know the effect of the mean shift and ability to detect with either chart. So somewhere in here-- I guess around there-- I got I got lucky, and I'm starting to sample now with the x bar picking up that detection. But with the EWMA, in some sense, I've still got 90%.

It takes 9 or 10 data points for the more recent data to outweigh the past data, and so I've actually got, in this case, a time lag in detecting the mean shift with the EWMA compared to the x bar. So a question here first--

AUDIENCE: I don't see how we can compare them because they actually are different time scales. All the x bar will effectively-- let us say the sample size is [INAUDIBLE]

DUANE BONING: So the question was, how do you compare the time scales on these? So here's what I think is going on. And this is sort of what I was talking about a second m but it's probably still a little confusing. I believe that this n equals 5 sample size means that, inside of this data point, if I were to explode that, it actually has five data points inside of it. And I'm only plotting the average of those five on here.

This data point also has those same five data points, but then it also grabbed the data that was lurking in the previous one. And it has its five data points and its five-- whoops. Yah! Don't want to-- I think it was trying to exit me out here. So while it's doing that, the point here was that the n equals 5 is not looking back five data points in the x bar. It's buried in each of the individual points-- so that both EWMA and x bar are not affected by this n equals 5 sampling differently.

But then the time scale is correct in terms of-- x bar only looks at my most recent sample. The EWMA does look back in history, and that's why you've got this additional time lag. Was that getting at your question or your observation, or is your question a little different?

AUDIENCE: [INAUDIBLE]

DUANE BONING: OK. OK. And there was a question in Singapore.

AUDIENCE: Yes. So my question is sort of built on what [? Stephen ?] was asking. So for a smaller r, you have a higher probability of detecting a mean shift. But at same time, for a smaller r, the time it takes for you to detect the mean shift is longer. How does these two coexist? Well, you have a higher chance of detecting it, but it takes longer as well.

DUANE BONING: Yes. They're a trade-off. And so what you are basically given is a degree of freedom that you get to play with, or experiment with, or adapt to your process. So what are the factors in adapting to your process? If your process is really noisy and you think it's going to be hard to detect a mean shift in the phase of that noise, then you do more averaging.

You'd have a smaller r. If you're trying to detect-- maybe with a higher risk of a false alarm, but you're trying to detect more quickly, then I'd have a larger r, so it responds faster, but doesn't filter as much. So you can see here, with the r equals 0.1 versus the r equals 0.3-- this is the same data, I believe. Is this the same data? I guess not-- it's just another example.

Here's an r equals 0.3 kind of case, where, for this data, this looks like about the right trade-off. I'm getting reasonable filtering. The black dots are my EWMA. And for a mean shift here, I've got very rapid-- I've got rapid detection, much more rapid than, in fact, the x bar, which doesn't really trigger finally until way-- another 20, 25 runs later.

So I think what you'd have to do is, again make the trade-offs between responsivity, the size of the shift that you're trying to detect, and how much cost you to have different delays in detecting the shift. So we've got two questions here-- one here first.

AUDIENCE: [INAUDIBLE] It's almost like you are taking a larger sample size. [INAUDIBLE] You're taking the 5 samples for-- 5 measurements per sample, plus 5 previously or slightly lower [INAUDIBLE] plus 5 [INAUDIBLE] plus 5 [INAUDIBLE] So you're just taking more samples per-- more measurements per data point, right?

DUANE BONING: Yes. So that's a fair way of looking at it. That's a good way of looking at it. So just so everybody could hear, with the x bar, I take my n equals whatever sample, but with the EWMA, I still got that from my most recent update, but I'm doing additional filtering or averaging of past history in.

And the net effective that is I've got a net aggregate-- more number of samples, if you will-- unequally weighted, but a more-- and therefore, that makes perfect sense that your control limits come down. That's a very good way of looking at it.

AUDIENCE: So in the previous example, does EWMA also trigger before x bar? Because x bar [INAUDIBLE]

DUANE BONING: Oh, yes-- good point, good point-- yes.

AUDIENCE: Is there ever a case when x bar will trigger [INAUDIBLE]

DUANE BONING: Good question-- are there ever situations where x bar might trigger still before an EWMA for an x bar chart? I don't quite know the trade-offs and where you get in the limits of the trade-off. You've always got chance at work, and I suspect that, with a large enough-- so the following scenario probably comes closest.

Let's say you had a very large mean shift, but a very small r. With an x bar, I react entirely to a sample drawn from the very large mean shift. Let's say it was a 2 sigma or a 3 sigma means shift. But I have a very small r. Maybe I'm only calling that 0.05. That 5% mix in of that most recent data point is probably swamped out by my past data still, and so in that case, I think my EWMA would likely not detect the mean shift.

So I think of the EWMA as especially good for improving responsivity and detection ability for small effects, because it's aggregating, just like the CUSUM that we'll see in a second. But I think, if there are really big-- bigger effects, you lose by the filtering in the time response. Was that an example you're going to suggest?

AUDIENCE: [INAUDIBLE]

DUANE BONING: Oh, good point--

AUDIENCE: [INAUDIBLE]

DUANE BONING: Yeah.

AUDIENCE: [INAUDIBLE]

DUANE BONING: Yeah. Yeah. So other rules here would help-- yeah, absolutely. OK, so I guess this is sort of saying a little bit the same thing. This, I think, is really good for small mean shifts-- that are persistent. It's not a one-time shot, that I've got a delta mu mean shift, and this allows us to aggregate. There are additional ideas we can use. Yeah-- question [INAUDIBLE]

AUDIENCE: [INAUDIBLE]

DUANE BONING: No, you would not want to-- so I think Hayden's point here was on the purple beta points you might apply, which was just an x bar. You might apply the WECO but you should not-- there might be variants you could derive some of the other WECO rules, but I don't know how you would do that. It's not intuitive.

I think, for the EWMA, you pretty much are just trying to track the mean, and not additional trends or-- I would not use any of the additional WECO rules with an EWMA chart. So there's another chart that is very often used. It's called a CUSUM I just want to give you a quick feel for that, and then move to some multivariate issues, which is also targeted at this idea that, when I have a small mean, shift what I'd like to do is use the fact that it is a persistent mean shift that I'm then sampling around within the process noise.

And if there were a simple way to aggregate the mean shift algorithm many data points in the past, I'd build up a signal that's larger and might improve my ability to detect that effect and reduce the average run length for detecting that effect. And this CUSUM, or cumulative sum, is another way of thinking of it as a filter-- and in this case, just a discrete time integrator-- which is looking at the sum of deviations of each of my sample point from the assumed grand mean.

So I'm not just filtering and plotting the average, but I'm saying-- asking the question, am I still in control at my assumed x bar bar, my mu 0? And if I'm not, then this statistic, this sum of being to one side of it, the c sub j here will grow if I've got a plus mean shift, or become more and more negative and grow if I've got a negative mean shift-- so that I'll convert a mean shift to an integration of a mean shift.

The mean shift is a step. The integration of a step response is a ramp. So what we would hope to do with a CUSUM chart is improve our mean shift sensitivity, and basically have charts that look like this. When I'm in control, I might wander around the mean. Sometimes I'll be above, sometimes below, so this c sub j, this cumulative statistic, does wander around a bit. But if I do have a mean shift, on average, I keep adding to that C sub J.

Sometimes I might go down a little bit, because I still have something a little bit below the mean. And so what you create is this ramp. You create a slope caused by that mean shift. Now, one can also now apply some statistics and design, control limits for that C sub J statistic, or use a couple of other approaches.

And one approach is this so-called V-Mask that says it-- what I would like to do is, if, in fact, I've got a mean shift, this slope-- let me erase a little bit here-- whoops-- so there are two different approaches-- and I just want to briefly mention both of those-- for detecting things on these charts.

One is draw a control limit for the aggregate of the C sub J. The other is to look and say, once I've got a mean shift, I start building data points with, on average, some slope that is different than a slope that I would expect by random chance alone, just forming the sum statistic from a normally sampled process.

The tabular CUSUM-- this idea of actually looking at the statistics of-- let me skip this for a second-- this normalized statistic z and s, and looking and deciding on what control limits are, and so on-- I'm actually going to skip that. There's about five slides in here, but it's described very nicely in, I think, chapter 8.1 in Montgomery.

And I think that's pretty natural. It's just sort of a statistic-- a little bit confusing derivation of the statistics that-- I just trust the statisticians who have done right. But there's also another picture here, which is based on the slope idea. And you can similarly derive some formulas based on alpha and beta risks, what size of a delta shift in the mean you would hope to detect, and what kind of a slope you would expect to see, in those cases.

And so you've got these two parameters that go towards the formulation of this V-Mask that you overlay on your data points based on an expected slope, and you need a certain number of data points in order to be able to build up your estimate of the slope. That's the dead zone.

And what you do for each data point here is overlay the V-Mask and say, look back in time and see, did any of these points intersect with the mask-- have a higher slope than you would have expected? And then you update that on a running basis. So in this case, here at time point 18 or 19-- whatever that is-- everything looks fine.

I go up. Everything is still looking fine. Still here, I'm starting to get some evidence maybe that things have changed, but statistically speaking, that's still within the range that I would expect with some alpha error. But then finally, I get to a new data point here at time 23, where, in aggregate, I've got some indication that the slope is too large.

So this is kind of a intuitive graphical approach that's talking about the slope associated with a mean shift in the CUSUM statistic. Yes?

AUDIENCE: Would you also perform a linear regression for the last three or four parts and check the [INAUDIBLE]?

DUANE BONING: So the question is, can you perform a linear regression and calculate the slope of that? And I think that's identical to what this is doing graphically. So in fact, if you were numerically-- or implementing an automatic version of this, that would probably be how you would implement it. That might even give you some more flexibility in terms of looking at the goodness of that fit, but I think it's essentially identical to this.

OK. So just to summarize these alternative charts, there's a few simple messages. Noisy data-- if you're trying to detect a mean in the noisy data, you might want to use some filtering. We have the sample size that does some filtering, but I might want to use more data in the past and this linear discrete filtering, either with the CUSUM as a running integrator or an EWMA-- which is an iterative discrete filter-- can be applied.

And I think, conceptually, they're fairly nice in terms of that noise suppression for detecting of the mean-- got to remember that your choice of factors, like r and the EWMA, depends on your process, and the mix of noise, and how fast you want to respond to be able to detect a mean shift.

And then, going back to what we started with our little example at the very beginning. Noisy data need some filtering. We've talked about it with respect to the mean-- that, if I'm filtering that data, you also still want to look back and monitor the variance as well. So you can apply these charts also to moving, S bar charts, and so on. So what I want to do next is move a little bit to multivariate process control, control charting when I've got more than one parameter that I may be monitoring on the unit process or on the product.

And what's really interesting here is there are essentially two classic mistakes that are made again and again. And I want to give you a feel for those two mistakes, and strategies for avoiding them. OK? One classic problem is you've got more than one parameter you're monitoring-- you implement more than one control chart-- seems natural.

But the problem is you can get many, many, many more false alarms than you might have expected, or believed you are going to get, because you're just monitoring more parameters. And so I want to talk a little bit about that. That's even if there's no correlation between your parameters-- or especially if there's no correlation between your parameters, you can get-- just by the fact that you're monitoring more than one thing, you've got to think about false alarm rate a little bit differently.

And then the second classic problem is, well, there is correlation among the parameters-- the multiple parameters. How do you deal with that? And in fact, in this case, you often run into the opposite risk. Instead of too many false alarms, you may have points of data or product that is truly out of normal operation that you may not detect.

It's a highly unusual point that may in indicate something has changed in your process, but if you're just monitoring with univariate control charts, you'll never see it. That is a common mistake number two. Multiple charts is common mistake number one first.

If I truly have independent parameters being modeled-- monitored at my process step, and I set each one of them up to have some acceptable alpha rate-- false alarm rate-- 1 out of every 370 process steps going with a plus/minus three-sigma-- three-sigma control limits, alpha's about 0.0027-- sometimes talked about 0.003 as an approximation.

But now I've got p of those separate charts. What is the aggregate false alarm probability? Well to be careful about it, if they are, in fact, independent, that alpha prime, might aggregate false alarm, is, 1 minus 1 minus alpha to the p. This is the probability that I don't get a false alarm for each of-- for all of the other p control charts all at the same time.

And then 1 minus that is the chance that I did get one of them. Now for very small alpha, you do that expansion, the 1 minus x for small x to the p is simply 1 minus x So you do that and you get sort of the nice intuitive result that if I've got 10 independent control charts, I've got about 10 times the rate of false alarm that I would have from one. So what do you do?

What would you do? Let me say it this way. You've decided, based on your cost analysis or whatever, that you're willing about once every 370 runs to have a false alarm somewhere on this process, that you want to at least-- look at maybe you look at it real quickly. It doesn't take a long time.

But that's about how often you want the operator to stop and take a look, on average, even when the process is in control. So that's your false alarm rate that you're willing to accept, but you're monitoring 10 parameters. [INAUDIBLE] some approaches you would use?

AUDIENCE: You have to increase your control limits.

DUANE BONING: Right, perfect-- you would increase your control limits. If I expand my control limits out a little bit, I have smaller false alarm rate on each chart. Well, I can basically decide what my aggregate false alarm rate would be-- 1 out of 370. And to set my false alarm rate on each individual chart, just divide by how many charts I have.

Now, I might have a 1 in 3,700 false alarm rate on each chart, and I use that new alpha to set my control limit. They might end up being, in that case, plus minus 3.4 sigma control limits. So the individual false alarm rate is lower on each one, because really, the question you are asking-- the key question was not that individual parameter, but rather, the question, is my process in control?

In aggregate, do I have an event that looks unlikely-- so unlikely that I have 1 minus alpha confidence, 1 minus alpha prime confidence, that, in fact, something unusual is going on in the process? So that's a strategy. I want to also point out-- and this is kind of a neat little article that I found last year-- that this is a more generic problem, a more generic common mistake than just control charting.

It applies all the time, when people are doing significance tests or hypothesis tests, but doing more than one of them at a time. Control charge is just a running hypothesis test. Now, if I'm checking 10 hypotheses, 10 control charts, and 10 parameters. My aggregate false alarm rate is higher.

And there's this neat little article pulled from-- that I found in*The Economist* last February that's called "Medical Statistics Signs of the Times." The subtitle here-- you may not be able to read it-- is, "Why So Much Medical Research is Rot." I'll read you the first couple of sentences here. People born under the astrological sign of Leo are 15% more likely to be admitted to hospital with gastric bleeding than those born with the other 11 signs.

Sagittarians are 38% more likely than others to land up there because of a broken arm. Those are the conclusions that many medical researchers would be forced to make from a set of data presented to the American Association for the Advancement of Science. At least they would be forced to draw them, if they applied the lax statistical methods of their own work to the records of hospital admissions in Ontario, Canada.

And the basic confusion drawn here in the red arises because each result is tested separately to see how likely in statistical terms it was to have happen by chance. If the likelihood is below a certain threshold-- typically 95%-- so we've got a 95% confidence, which says, 1 in 20 times, by chance alone, I might see a signal that big-- by chance alone.

And now I check 12-- maybe astrological sign is your 12 things I'm checking for as factors-- and I've got a 1 in 20 chance, and I check 12 of them-- probably I'm going to pop up and say, oh, that one's significant. And so that's exactly the same common mistake that you can see with multiple control charts.

And so the point here would simply be you should-- if you're checking multiple factors, you got to have a stronger indication at any one factor-- otherwise, you're just fishing, and by chance alone, you're going to find stuff. So that's kind of a neat example. And so here was the point I made about how you fix it. You just fix your aggregate alpha prime, you have your individual alpha, and then you pick your upper and lower control limits based on that alpha.

There's a second common problem, and as I said, it's the obverse of this, which is you very often assume that, if I'm monitoring 10 process control parameters, I assume they're all independent. But very often, they are not. There is correlation among those parameters, and I want to give you a little bit of a feel for what that looks like, and what the risk is, and what approaches are for control charting and particular of this.

So here's an example drawn from some data that came from an LFM thesis a few years ago, a body and white assembly. So body and white-- this is when they are basically putting together all of the different panels on the car and measuring many, many points with all kinds of laser, automated apparatus to be able to look at dimensional control of these panels in the assembly-- of some of these panels.

Other correlated data comes up all over the place-- injection molding-- I think we're missing-- I can't remember his name. He asked a question last time about critical dimensions on a semiconductor wafer. Yes, they are often correlated. But here's the body and white example, and what I'm sort of showing are these little coordinate measuring machine, CMM kinds of measurements. I have a few of these different data points.

And the point is, if I think of this as a metal panel in some way, and I'm measuring the position, say, of these four corner points, you would expect there to be correlation among them. And certain kinds of distortion might change relationships between them. And in fact, you might even come up with other statistics, like a y1, and a y3, and a y4, that try to detect or look at tracking of a couple of these parameters and deviations.

But the point is I wouldn't necessarily want to monitor those four corner locations as independent parameters. Here's another example. This comes from an ultrasonic welding set of data. And you may not be able to see this. This is a wonderful correlation plot of three, four, five, six, seven, eight, 10 different parameters against each other.

There are different things, like P2 weld energy, P2 peak power. These are simply scatter plots of data drawn from that process. You can see some of these parameters that track down-- that's total weld energy versus P2 weld energy. I don't even know what those things mean, but they sure sound like they're going to be correlated, don't they?

And sure enough, there's extremely strong correlation in that data. There's other cases where there's also fairly strong correlation. So again, these kinds of correlated data, where knowing something about one of the measurements tells you about the likelihood of the other measurement, occur all the time.

So what if you have truly random independent variables? What would a scatter plot look like, and how do we go about setting the control limits? That's pretty natural, right? So here's an example where we had x1 and x2. The scatterplot basically looks like a cloud of data-- and if I added more data, you start to-- depending on the scaling, if I were simply to draw, say, a joint 95% confidence integral, it's just kind of a cloud of data, but there's no cross information between x1 and x2.

In other words, if I told you that x1 was right here-- the value of x1 was this value-- that really doesn't tell me very much about what value of x2 could be. It still falls anywhere within its distribution. These are two independent random variables. And you can decide what the proper limits are. Now you know enough about picking x1 and x2, and I would do the alpha prime idea to pick independent limits on those two cases.

This is an example scatterplot for correlated random variables. And in this case, it starts to look like an oval or a football kind of shape, where there is statistical correlation between these two parameters-- by which I mean, if I tell you that x1 is in this range, that tells you something about where the value of x2 is likely to fall. It might not fall within its full natural range. There is cross information or cross-correlation between the two.

Now, if I were to plot x1 and x2 separately, it looks an awful lot like the uncorrelated case. So if I had independent control charts for the two variables, essentially what I'm doing is just projecting all of these data points. I project all of these data points to get my range for x2, and I project all of these data points.

And in fact, if I plot both of those, x1 and x2, I would likely get-- yikes-- don't want to do that-- I would likely get two separate, normally distributed variables, and I might be inclined to simply plot them both as x bar and S kinds of control charts. The risk here, though, when you do that is you might miss outliers, data points that do not fall in the normal or the natural variation of the process-- this being the natural variation of the process that I will not detect.

This data point still looks like it's within that distribution. In, fact, it's very close to the mean of x2. And it's still within this distribution. But within the natural variation of the process, within that oval, its distance, in some sense, from the central moments of that distribution-- it's pretty distant. It's not indicative of the normal operation.

It's a highly unlikely, by chance alone, data

point, and I would never see it on the control chart. So there are multivariate charts that let us get at and chart on a single chart-- multiple factors that are correlated together-- and detect those kinds of problems.

What I want to give you is a notion first of a statistic that aggregates, in some sense-- numerically aggregates exactly what I was talking about qualitatively here, that tells you this notion of, what is the distance from the correlated variation cloud of each of my measurement points? And if I have a statistic and I know it's PDF, then I can set control limits on that statistic. And that statistic is the Hotelling T squared, is one very important one.

We can also monitor-- have EWMA and CUSUM extensions of these. Which I'm not really going to talk much about. So what we have to do here is have a slight extension to our notion of a probability density function to have joint probability density functions that tell me if I've got, say, two variables-- or three, or four, whatever-- what is the probability density associated with those?

And what we want is a single scale control chart that aggregates these. Now, I've got a little bit of mathematical notation in here that I'm not going to go through the careful full explanation or derivation of. What I want to simply show you is how a multivariate expression of the same variables we've been using all along gives you a nice qualitative feel for what this statistic looks like.

So if we have a vector of measurements-- let's have p parameters-- I'm measuring 10 different parameters-- body in white or whatever it is for that particular process-- and put those in an x-- underscore, an x vector. And then I can define a vector of means. So I have a set of means for all 10 parameters.

And I can also calculate, if I have lots and lots of historical data. There isn't a true underlying covariance matrix that talks about or characterizes how these data go together with each other. Now, I can estimate that covariance matrix based on sample data, so I similarly have a true mean and a true covariance.

Then I might estimate with an x bar and an S-- here I'm using a capital S-- to be a sample covariance matrix. And there are formulas in the book using a very simple linear, matrix-- linear algebra for how you would do that, given a particular set of data. And they look a lot like the univariate example.

But I did want to point out that there's both embedded in this the notion of variance-- how variable 2 varies together with variable 2-- that's just our normal notion of variance. And there's also covariance-- that is, a measure of how these two variables move together.

We can actually write out what the PDFs probability density functions-- are in these joint cases. And if you write it in matrix form using this covariance matrix and the mu vector, it looks almost exactly the same as the univariate case. Notice that, instead of variance in the univariate case, now we have the covariance matrix.

And similarly, instead of just x and u being univariate values, I've got those as vectors. But the PDF looks almost exactly the same. And qualitatively, you're getting an exponential drop-off, this e drop-off with the square of the distance of a point from the mean-- the normalized distance from the point.

So if we look back at the univariate, the whole idea is I subtract off the mean when I standardize, divide by the natural variation, and that tells me the shape of the bell curve-- shape of the normal curve, how fast things drop off.

And I can also, for any particular multivariate u-- multivariate data point x that has its 10 parameters in it-- what's really cool is this formulation here of the distance of that x vector from the mean point-- the main vector squared-- that squared distance, but weighted then or normalized to the natural spread-- that natural spread or width of that cloud-- is just a distance measure-- squared distance measure of each data point from the central moment of the central point of the distribution, scaled to how much natural spread or scatter there is in the data.

OK, so that's the PDF. Those are just the formulas for the sample use cases. What I can now go in and do is ask the question, what is the distribution associated with that distance metric? And that distance metric is essentially-- if I expand out that distance formula, it is a sum of squared distances normalized to the correlation, which starts to sound like sum of squared unit normal variables, which was our chi-squared distribution-- sum of squared normal-- so what's cool here is that the chi-squared distribution with p parameters-- p degrees of freedom-- is the right statistic for measuring how likely it is to see those distance-- the T squared, those distance metrics.

So one can actually have the PDF. One can have the distribution for that PDF, and based on that, you can formulate or form a-- that statistic based on my actual data x. and then I can have an alpha squared that I can pick my cutoff points from appropriately. Oops-- I thought I had a PDF in here somewhere.

So when I'm making a control chart, I use the same ideas that we saw before. I have a false alarm rate. In this case, I'm going to aggregate all my data onto a single control chart so I can just use my aggregate alpha. And I pick the tail points on the chi-squared distribution.

Remember, a chi-squared might have looked something like this, and I might pick a way out here as my chi-squared so that I've got my alpha out there in the tail. That picks my cutoff point for my upper control limit. Generally, you can also do that for a lower control limit. Very often, if you do some-- apply some of these formulas, the lower control limit is negative.

And you can't have negative square, so you pick 0 as your lower control limit. And you're really only monitoring four data points that are larger than-- I've got more variation in my system than I expected. So just to close up on this univariate versus chi-squared chart, here's an example-- comparing the univariate case up here to the multivariate case down here.

And what I'm plotting here is, again, the scatter of data. You might see there's a slight shaded oval case. And all the blue data points are what I would do if I just projected those down to an x1 and x2, and did univariate control charting. If I then had this little red dot that fell outside of the joint confidence interval-- on my individual charts, notice where it falls-- perfectly within the control limits-- nothing that tells me that that point is unusual.

But if I then plot that data point-- that red data point there and on the chi-squared chart-- I calculated the chi statistic, that distance statistic-- that would trigger, because its distance from the mean is unusual. Now, if I actually didn't know my true covariance matrix, but I had to estimate it, I have to estimate sigma with an s.

I'm right back to this whole difference between a t and a unit normal. So just to close this out we often talk about these as t-squared, Hotelling t-squared statistics, but it's exactly the same idea. And it's got its own cut-off points.

OK, so with that, I think we've pulled to a close on control charting for the-- at least for the time being. Next time, we'll start diving in on yield modeling and different uses of some statistics, but I think what you've got now is some exposure to EWMA, CUSUM. You'll play around with those with the three or four problems on this week's problem set. And start looking at the quizzes. We'll have the first quiz a week from today-- in class.

AUDIENCE: Professor, I have a question.

DUANE BONING: Yeah. Go ahead and break. I'll take your question while people are dispersing.

AUDIENCE: So for the [INAUDIBLE] you taught us today, are they going to be included in the quiz?

DUANE BONING: It's possible you'll see an EWMA or a CUSUM, but they would only be fairly simple questions for those. It's not likely you'll see a multivariate problem.

AUDIENCE: OK, thank you.