STAT 100 Notes March 30, 2010

Intro to week 11:

1. **Quality Control**. Snee 323-338: Improving the Accuracy of a Newspaper: A Six Sigma Case Study of Business Process Improvement. Doganaksoy, Hahn & Meeker 339-358: Assuring Product Reliability and Safety.

Overview of Snee Article:

W. Edwards Deming tried to interest the large American Companies (like GE, Ford, Du Pont, ..and many others) in implementing his product improvement process. This was in the 1940s. These companies were doing well and did not want his advice. So Deming went to Japan in 1950 and initiated a revolution in industrial style there that had the effect of seriously hurting American Industry. Eventually, other world economies including those in North America heeded the changes that were needed. This story is partly told in the Snee article (pp 323-337). The role of Deming himself is covered in the Wikipedia entry http://en.wikipedia.org/wiki/W._Edwards_Deming and many other internet sites.
The Snee article talks about "Six Sigma" strategies which is really the modern version of the strategies that Deming initiated. While some of the details may seem a bit ponderous, there are a few ideas that are both very statistical and very useful. One is that a focus on variance reduction in manufacturing is an excellent way to increase profit. Another is that "management by exception" allows a worker to monitor and enhance many complex processes at once. A third is that graphical displays can be usefully employed by workers with minimal statistical education.

Other things to note from the Snee article:

DMIAC – Define, Measure, Analyze, Improve, Control – p 324
Process map – steps in production process – p 326
Cause-and-effect diagram – "fishbone" for sources of variability – p 327
Process capability – how small can variability be – p 326
Control Chart – Fig 3 p 327 and today's lecture
Pareto Chart – causes of variability – p 328
Reduction of Variability – today's lecture and p 330

For each you need to find out: what is it and why is it useful for industry?

More about Control Charts:

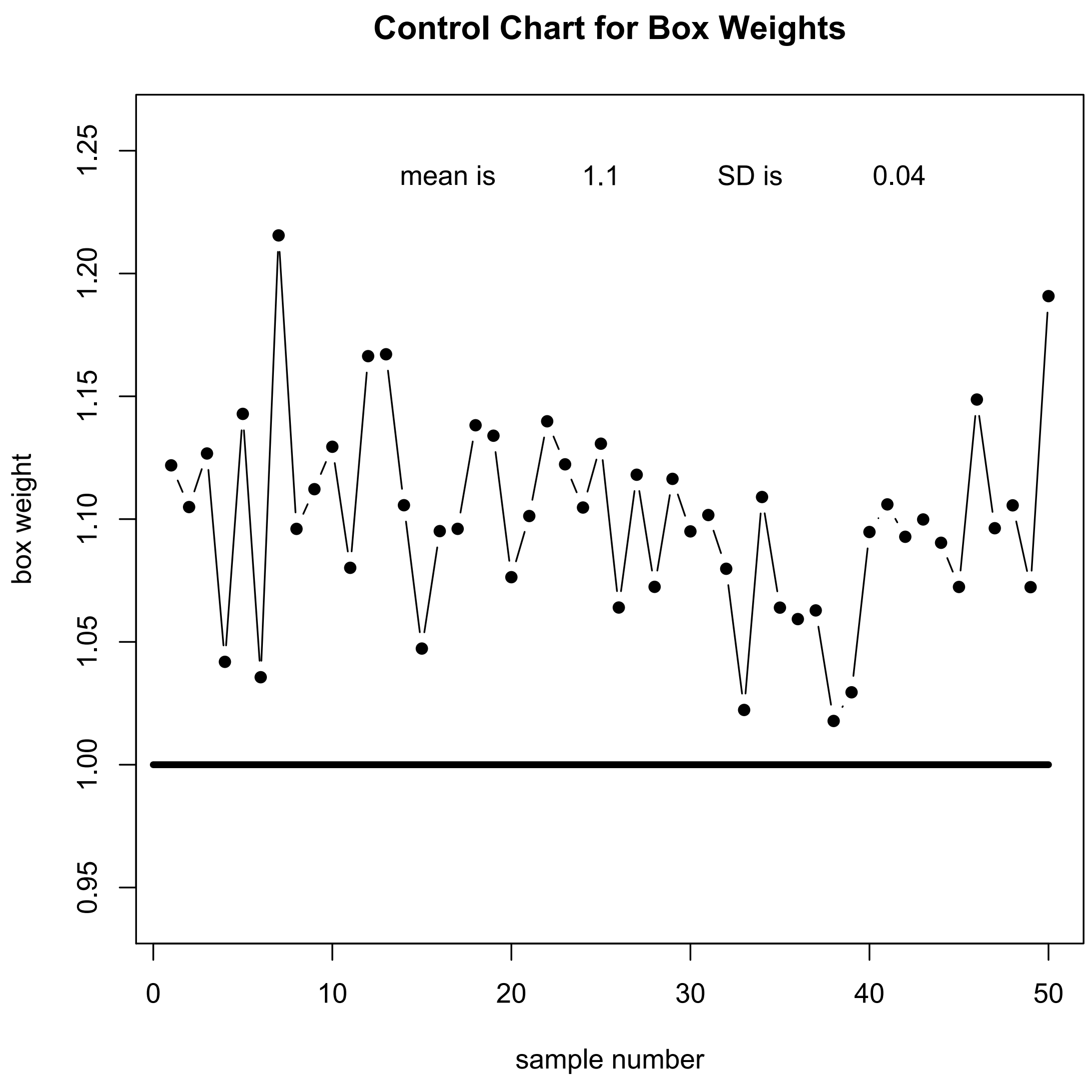
In the intro above I mentioned some points that are key to the quality control story – I will restate them here:

1. A focus on variance reduction in manufacturing is an excellent way to increase profit.
2. "Management by exception" allows a worker to monitor and enhance many complex processes at once.
3. Graphical displays can be usefully employed by workers with minimal statistical education.

To understand these points we need to demonstrate some more details about control charts.

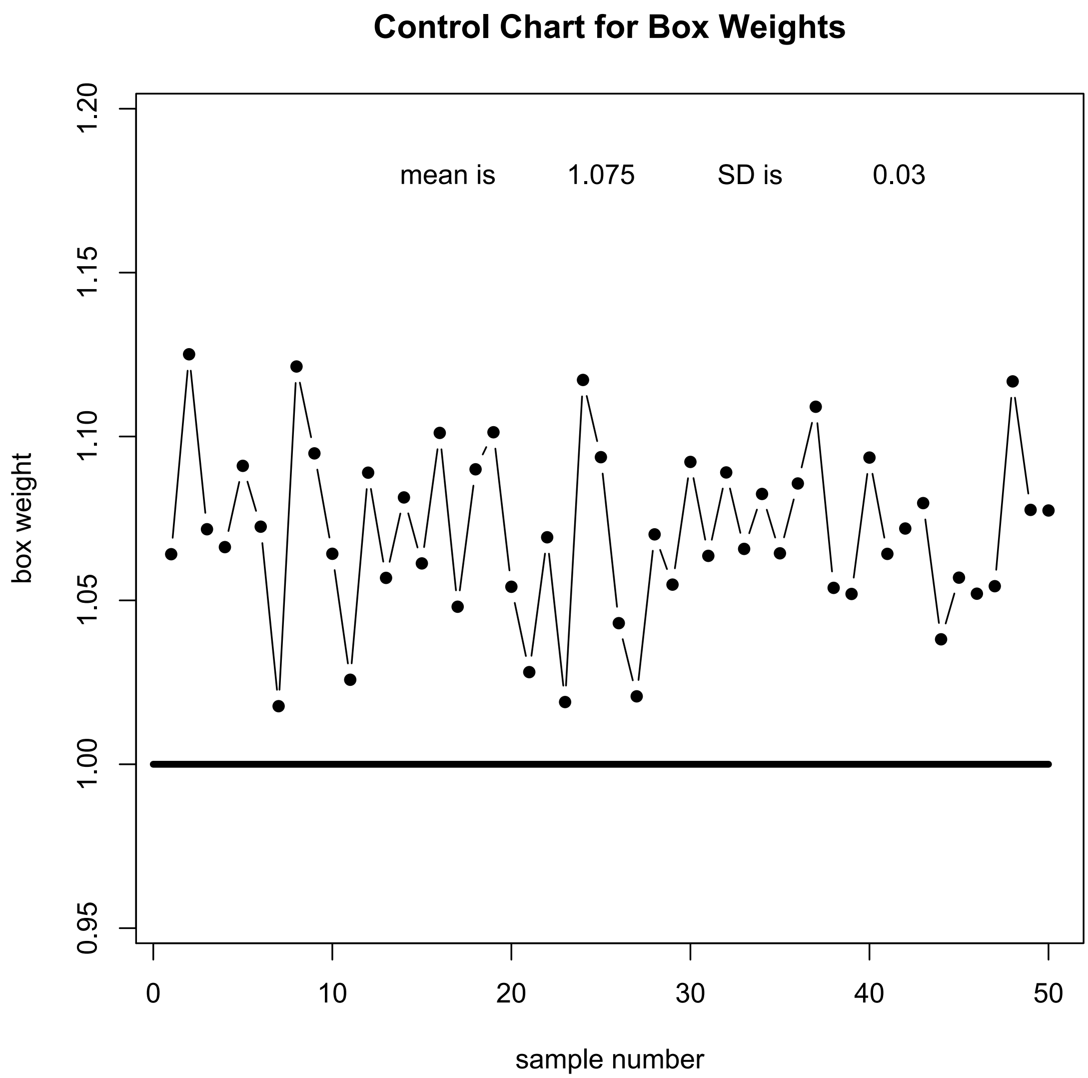
Suppose you are a manager in charge of a production line for a certain cereal box. The box is labeled that it contains 1 kg of organic wheat flakes. To avoid bad press, the cereal company does not want to risk selling 1 kg boxes of cereal flakes that are lighter than 1 kg. So part of your job is to ensure that the process keeps filling the boxes to at least 1 kg. There will be no complaints from buyers that have boxes overfilled.

One problem is that the process of filling boxes is not perfectly controlled. The actual amount in any particular box varies – so we have to set the average fill a bit higher than 1 kg to ensure that the 1 kg label is valid. The excess product we are forced to pack depends on the variability of the filling process. Here is a graph of a typical situation:



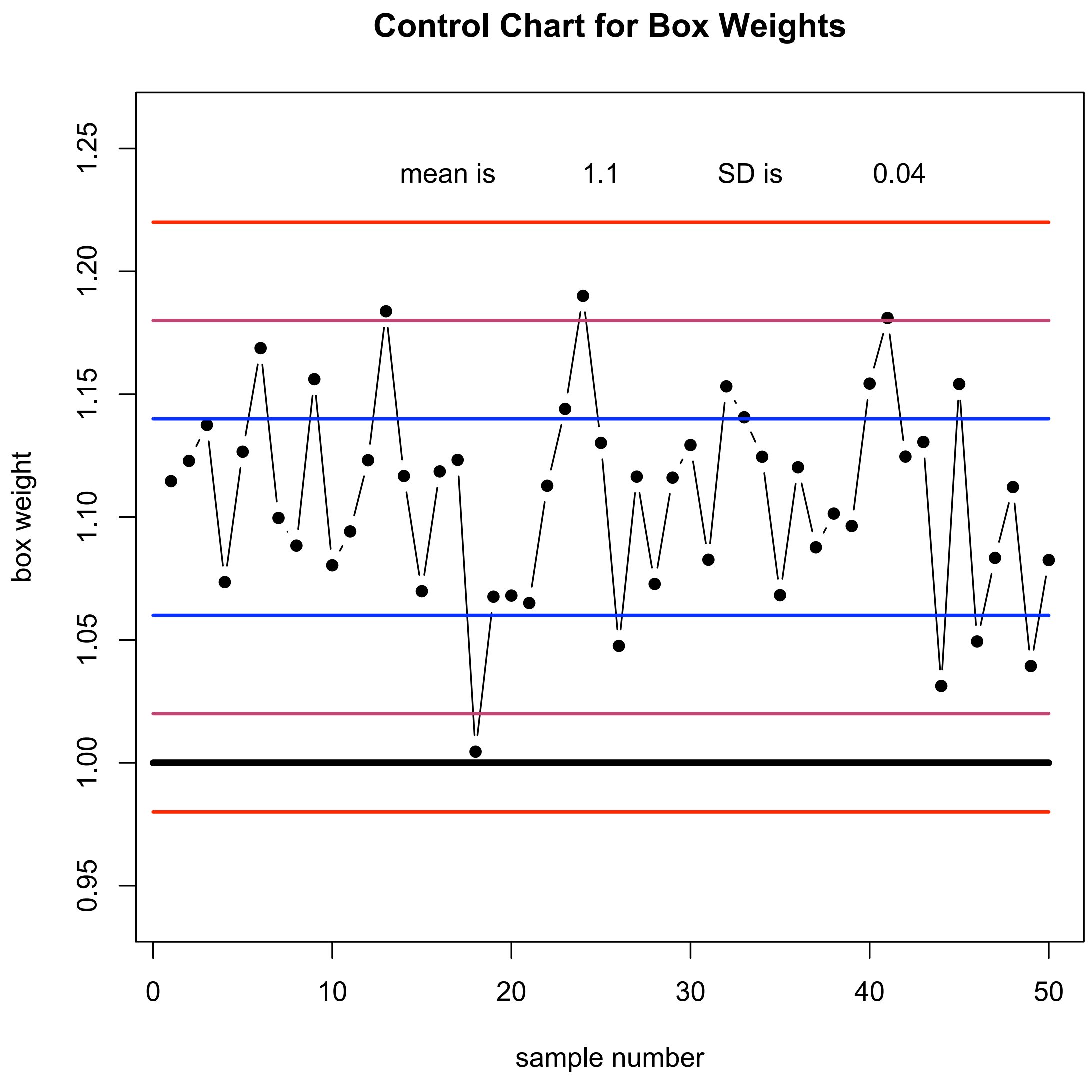
The graph might be thought of as the actual weights of a sampled box taken from the production line every 30 minutes. Note that all the weights are above 1 kg. The graph is fictitious data simulated from a normal distribution model with the noted mean and SD.

However, the variability seems quite high (SD is just under 4 % of the mean – if your own weight varied that much, you would have weight measurements varying day-to-day by ±6 pounds or more, which would be considered the result of a poorly constructed scale!). Clearly, if we can reduce the variability of our filling process, we can reduce the average fill amount without producing sub-weight product. Suppose we are able to reduce the SD from .04 to .03, perhaps by ensuring that small back-ups in the fill hose occur less often. Then we can reduce the mean weight by about .025 kg, and save a lot of product, increasing profit.

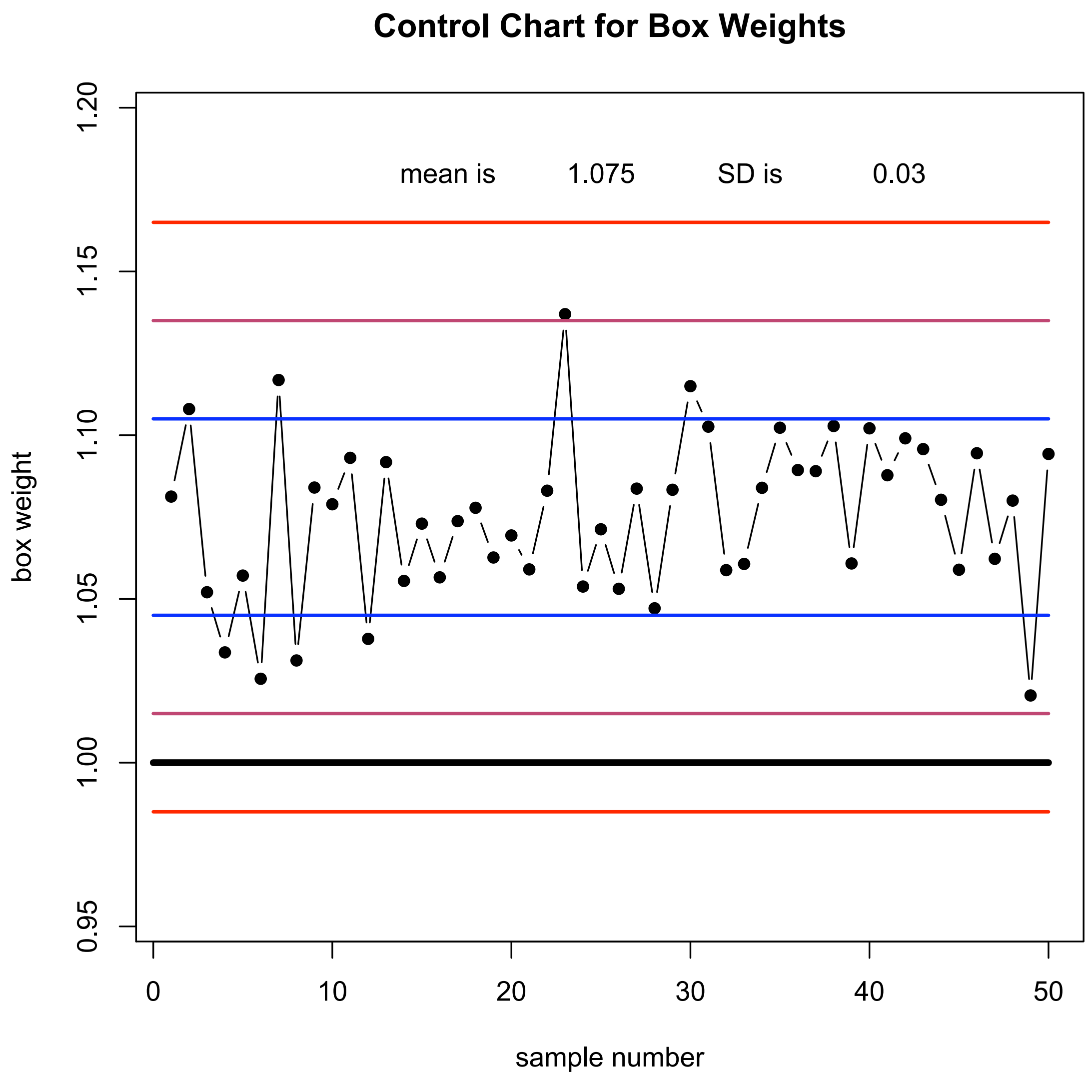


So this chart suggests that we are still staying out of trouble, even though saving almost 2.5% on our product cost.

So the big question is, what statistical strategy can encourage workers to find ways to reduce variability? One device is the control chart:

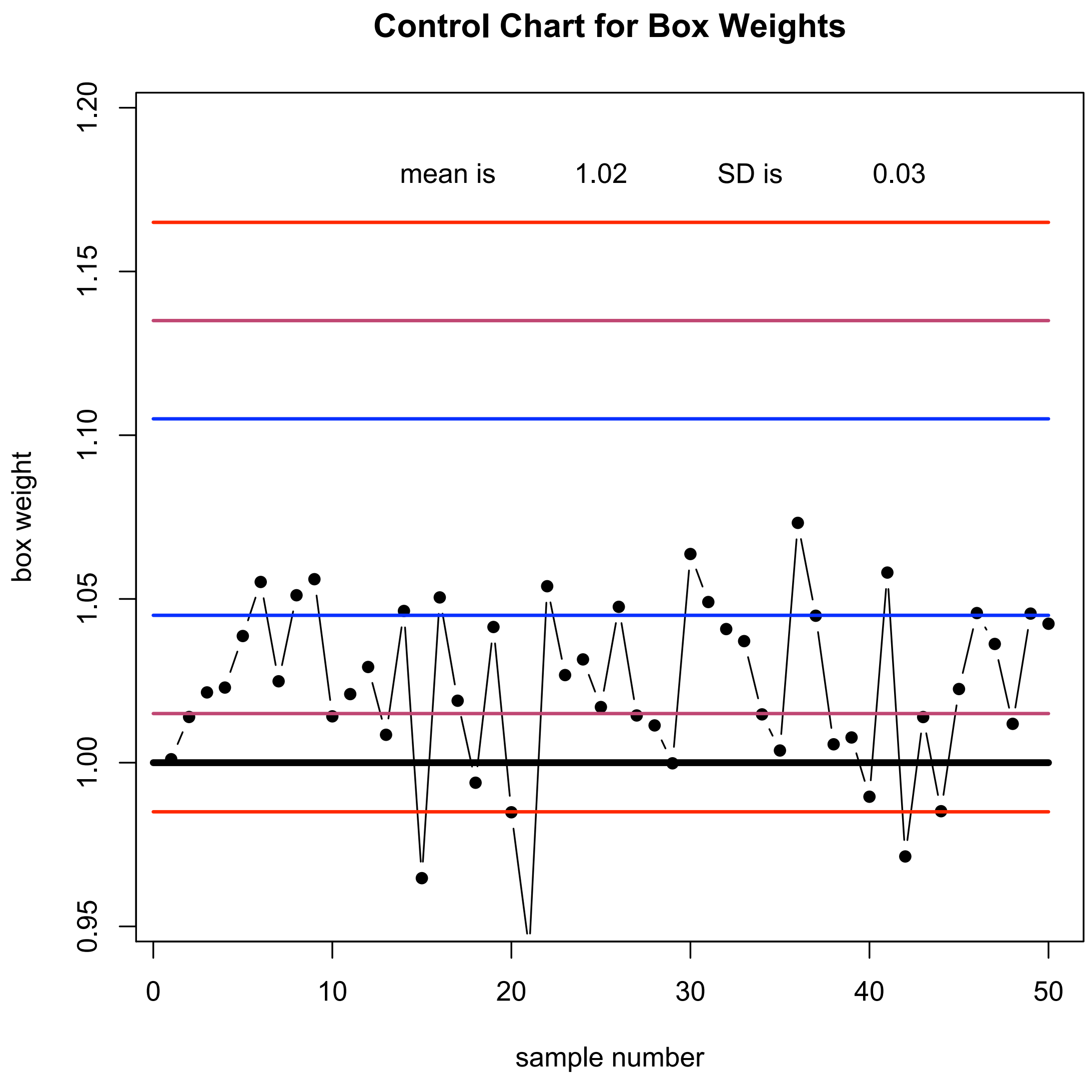


The coloured lines are at 1,2, and 3 SDs above and below the mean of 1.10 kg. The black line is our "specification" value of 1.00 kg. If we can figure out how to reduce the variability from say an SD of 0.04 to 0.03 kg, then we could lower the mean fill to say 1.075 kg and the 1.00 kg spec limit would still be 2.5 kgs below the mean, and underfilled boxes would occur with the same rare frequency as before. (1.10-1.00 = 2.5 * 0.04 and 1.075-1.00 = 2.5 * 0.03). See chart below:



As you can see, although we have reduced the average fill to 1.075 kg, underfills below 1.00 kg are still going to be rare (which is good). So reducing the variablility by 25% (.01/.04 *100), we are able to cut product costs by just under 2.5% (.025/1.010 *100). Put another way, variability reduction will result in increased profit.

Our premise above was that we were able to find a way to reduce variability. But how does the use of control charts help in this? To see how this might work, suppose that, continuing with our new model of mean 1.075 kg and SD 0.03 kg, we have operated for some time without any unusual test values. Then one day, unbeknownst to us, a malfunction in the filling process causes the average fill to suddenly drop to 1.02 kg. Because the SD might still be 0.03 kg, we may not see this reflected immediately in the data. In fact we can simulate what kind of data we would see following this malfunction:



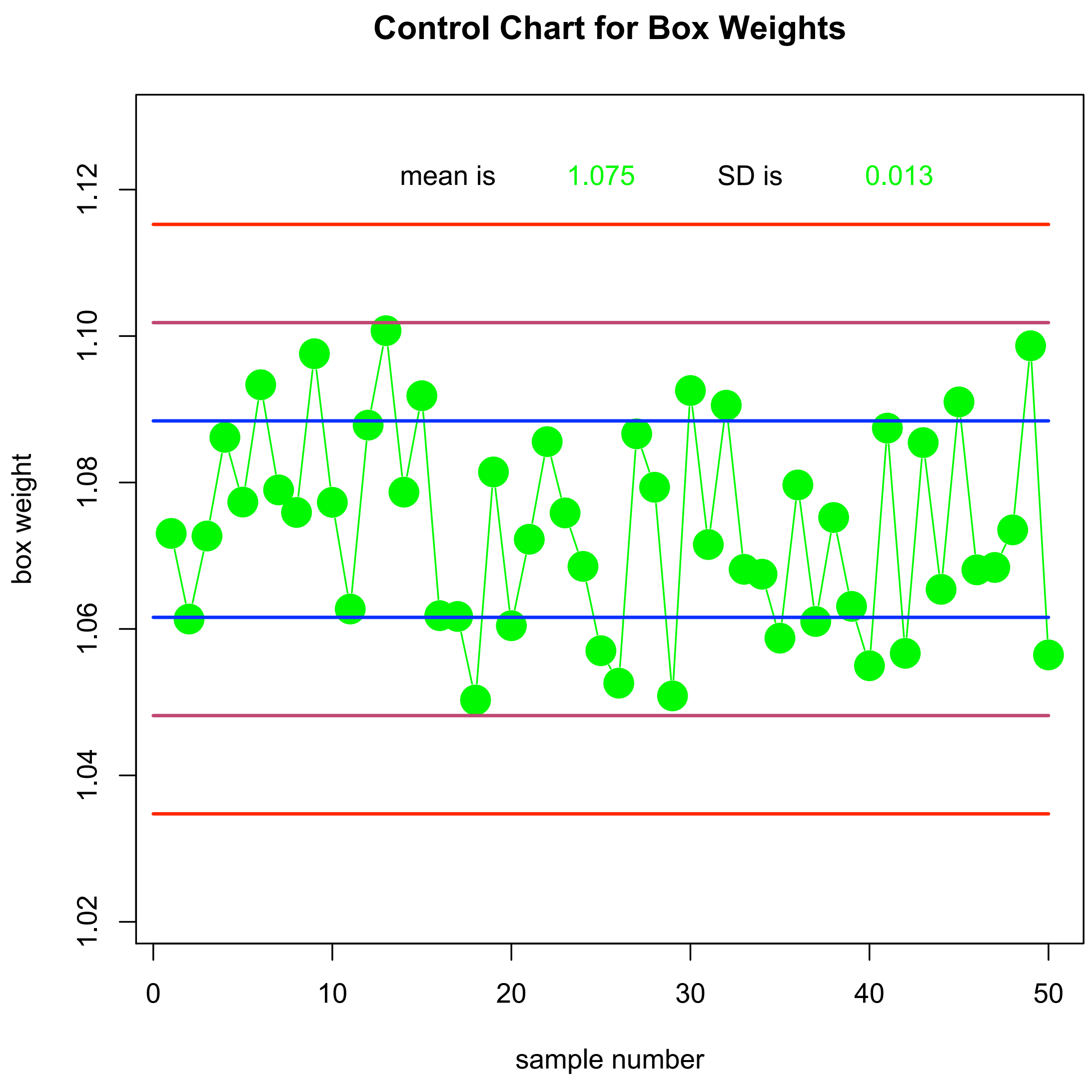
We can see that the -3SD line is breached fairly soon, even though, if the mean were still at its proper value of 1.075, the chance of a value below the red line (at -3SD) would only be about .13% (about 1 in 740). The lowering of the mean by just under 2.5% ((1.02-1.075)/1.02 * 100) means that the chance of a -3SD signal value is increased from 0.13% to 12.16%, no longer a rare event.

In this case it took 15 samples to detect the change, but there were none in the first fifteen samples that were below 1.00 kg. This was a bit lucky but in any case the rule "Stop the process to investigate a problem" would occur before too much underweight product is produced.

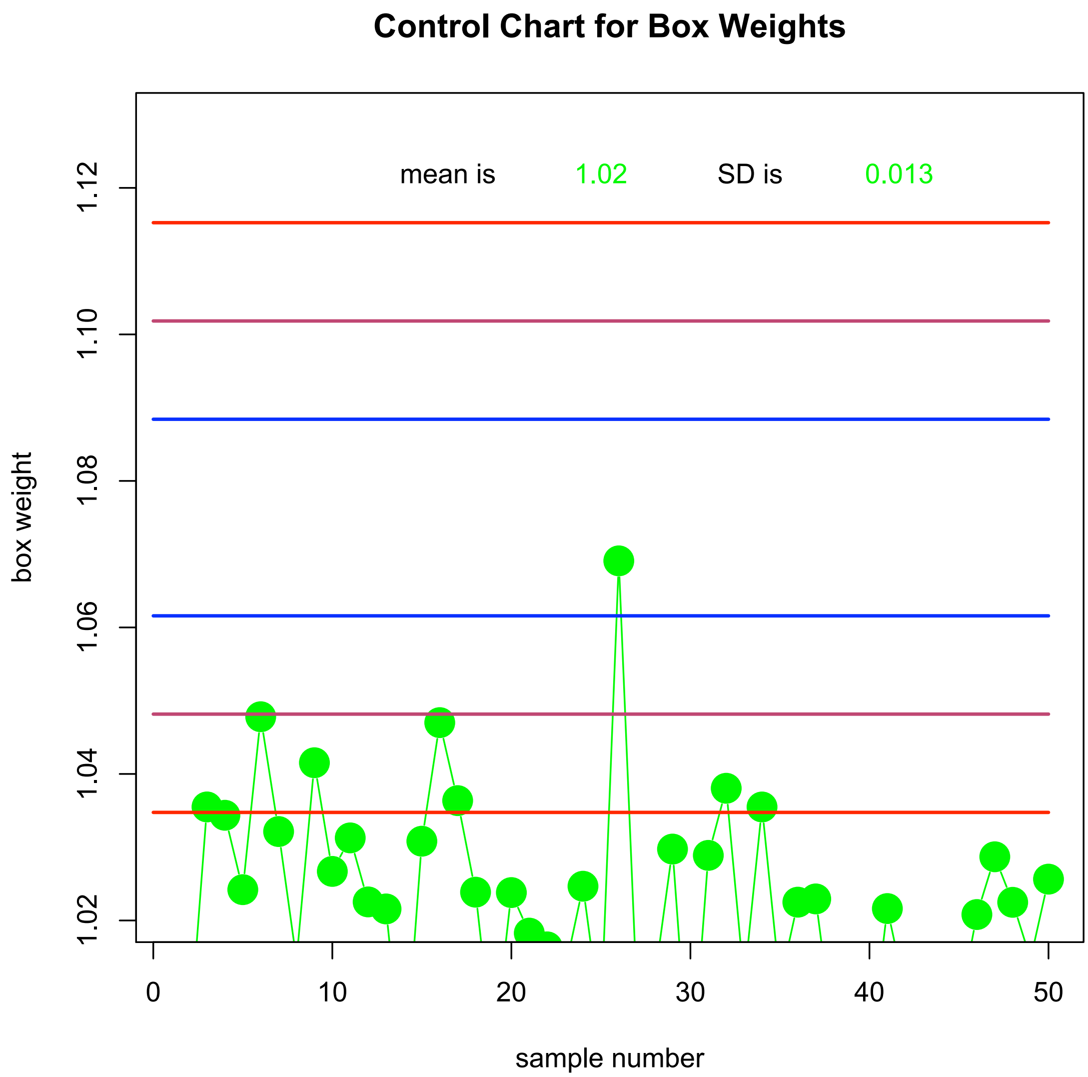
So this suggests the value of the control chart – *it signals in a simple way when a process is doing something wrong, and moreover it gives a hint of what has gone wrong by recording the timing of the onset of the problem*. Knowing this timing helps to identify what the problem was (for the engineers who are investigating).

So variability reduction is a something that can lead to increased profits, and control charts help to reduce variability.

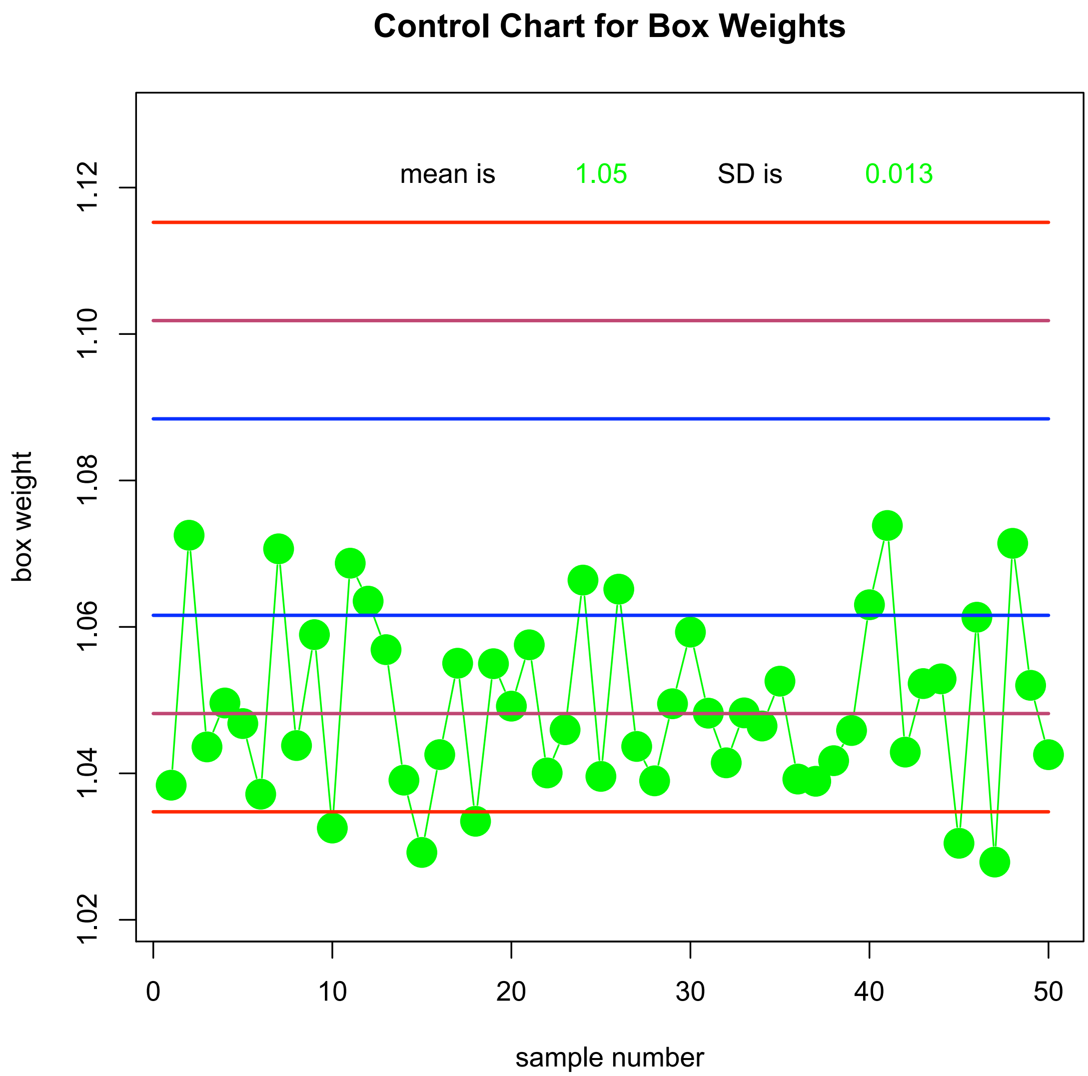
Now recall that taking averages also reduces variability, but in this case it was variability of means that are reduced. Q: Why do we need to find a physical cause of variability when we can reduce it by taking averages? The answer is that keeping the average box weight above 1 kg is not good enough, we need the individual box weight to be as labeled on the box. However, it does raise another possible use of control charts: suppose we are not concerned about individual values but we are concerned about a fluctuating average of the fill process. Then obviously we get a better idea of what is happening to the average of the process if we use more than one box per chart point. Given that even measurements cost money, we might explore the effect on our chart of using a small number of units at each sampling time, say 5. See below:



The points are enlarged and coloured to remind you that these are each an average of 5 measurements. However, the chart looks the same as our original control chart, but notice the difference in the scale: 95% of the values are in 1.075 ± 0.025 – the maroon lines – whereas for single values the maroon lines were at 1.075 ± 0.06. If the mean of the process fell from 1.075 to 1.02 as we postulated with the single measurement scenario, this would be a huge change in the distribution of the mean of 5 measurements, and we would start seeing the following:



So alarm bells would be ringing immediately. Averages, even for small samples of size 5, are much more sensitive to changes in the underlying population average than individual values. In fact, even a small change of population mean from 1.075 kg to 1.05 kg would cause a chart like the following:



and we would quickly detect even this small change (using the -3SD warning line, the red one).

So whether you use an individual values control chart or a small group control chart depends on the target you are trying to control.

Some further comments:

1.We used a small group (n=5) since this frequent measurement would become too expensive with a larger n – and also some measurements (testing of firecrackers, for example) are destructive.

2.We used upper control lines since, if we are overfilling boxes too much, this would also be wasteful.

3.Whether we use the ±1SD, ±2SD, or ±3SD control lines to take some action depends on how sure we need to be that the mean of the process is stable. Usually ±3SDs is used.

Now revisit the comments made earlier in the introduction to the Six Sigma article:

1. A focus on variance reduction in manufacturing is an excellent way to increase profit.
2. "Management by exception" allows a worker to monitor and enhance many complex processes at once.
3. Graphical displays can be usefully employed by workers with minimal statistical education.

I hope these claims make sense in view of the discussion of control charts above. Of course, there are many other techniques in the Six Sigma arsenal to achieve these goals, but the QC (Quality Control) chart is the one that depends most directly on basic statistical theory.

In the course outline, I listed the "Reliability" article (pp 339-358) as relevant to this topic of Quality Control. We have already discussed it in connection with Graphics although the main connection was the one graph having to do with the Challenger Disaster (p 341). The other graph on page 348 was a bit complicated at the early stage of the course so I wanted to mention it again at this stage.

But before I get to that, what is the connection of "Reliability" and "Quality Control"? Well, a high quality manufacturing process tends to produce reliable products, so is not a deep connection. Of course, if a product is badly designed, then no amount of control of the manufacturing process will produce a reliable product. But that is why "Design of Experiments", another statistics subject, is so important. The Six Sigma "DMAIC" process does include an "Improve" phase before the "Control" phase, so experiments to improve product design are an important part of Six Sigma. We have not said too much about experimental design except its basic structure (assigning treatments at random to experimental units) and its application to clinical trials. But there is a huge list of statistical strategies aimed at getting the most information out of every dollar spent in experiments. There are upper division courses in this at SFU.

Back to the graph on p 348. Ignore the orange line for now. The blue dots are just some "survival" data from the washing machine reliability study. The x-axis is simple enough – just the time until failure. The y-axis is designed to provide a straight line if the failure times follow a certain probability distribution model called the Weibull distribution. The orange line is the largest values for the failure rate believable in view of the data, and it is based on the Weibull model being correct. The reason for explaining this much is just to show that graphs are used for data analysis, and not simply for data summary. But maybe you already knew that from other examples.

OK.. enough about reliability …

Assignment?

No assignment this week. Spend some time on the lecture notes! Also, review your midterms – make sure you know now what you did not know at the midterms – it could pop up again on the final.

A Note about Populations and Probability Models

We have come across **populations of interest**, **samples from those populations** (sometimes random samples), **probability distributions** that describe populations of numbers, **sample frequency distributions** that summarize the random samples, and even **distributions of sample summaries**. All these things sound a bit similar and are a source of a lot of confusion. I want to review these items, and try to explain the differences with reference to some of the examples we have discussed over the course so far.

Populations of interest: for example, the population of voter choices (candidates 1,2,3, or 4 say) is a population of interest, as is the population of possible collections of 6 integers in the range 1 to 49 (in Lotto 6/49), the population of gay men in seven large US cities (as in the Young Men's Survey), the population of registered subscribers to the Butterball Thanksgiving webpage.

Usually we cannot have all the data we would like from a population of interest, and we must rely on partial data, such as a random sample of data from the population.

Samples from the populations of interest: voter sampling for political opinion is based on samples that use random sampling but are not simple random samples; in the case of 6/49, we want to be assured that the numbers chosen to identify prize winners are selected like a random sample from the integers {1,2,…49}; the population of gay men was sampled by first taking a random sample of venues; and the butterball sample was divided up into 432 sub-samples that were to be compared.

The method of inferring something about the population of interest from the samples from these populations always depends critically on the way the sample was obtained from the population.

Probability Distributions:

Sometimes a population is hypothetical, such as "a Normal population with mean m and SD s". There is no attempt to describe the long (infinite) list that represents the population of numbers in this case – we simply describe the population as having all possible numerical values from $-\infty$ to $+\infty$, and with relative frequencies defined by the bell curve with mean m and SD s. So the specification of a probability distribution actually is also the specification of a population, and possibly of interest. Often the "parameters" m and s are the items of interest.

Another example of a probability that defines a population is the very small set{0,1} or {H,T}, and we might specify the probability of each choice as (½, ½) or (.6,.4). As long as we sample with replacement, this describes an infinite population.

Frequency Distributions of Samples: Sample data can be summarized by a histogram, a dotplot, or even a table of frequencies. Two random samples (of size 25 say) usually produce quite different sample frequency distributions, even when they are selected from the same population. This is why we like to know that the sample was selected randomly, because then we can predict how much variability to expect from sample to sample, and, more importantly, from population to sample. A histogram of a sample of 20 from (0,1) would usually not have ten 0s and 10 1s. The dotplots and histograms I have shown you are usually illustrating the frequency distribution of sample data.

Distributions of Sample Summaries: The most common summaries we have used involved means and SDs, although a histogram or a dotplot is also a good summary. But sample means and sample SDs are called, in stats jargon, “statistics”. Yes, it is confusing to have one word mean several things, but this is the tradition! Any summary value from a sample is a “statistic”. “Statistics” can also have distributions – usually we talk about the theoretical distribution of statistics, since to have more than one statistic you need more than one sample, and one sample (of 25 values say) is the usual. For example, the sample mean has a theoretical distribution that we have discussed. In fact when the population is Normal, the sample mean also has a Normal distribution (exactly) with mean and SD the same as the population sampled. The surprising result is that the sample mean has this same distribution, approximately, no matter what the population distribution is. (This last statement is the “Central Limit Theorem”.)

I have illustrated sample distributions of means, but you should realize this is quite an artificial situation, since usually we only have one sample and one mean to work with. The way I get several samples is with simulation, and in this situation I suppose that I have a situation in which the population is known exactly (in the form of a probability distribution) and can generate as many samples as I want. But you need to conceive of the distribution of a sample mean because it is the key to understanding how a single sample mean can estimate a population mean, and moreover the single sample can provide information about the precision of the estimate.

Please try to keep these five things (bolded, above) separate in your mind.

(KLW 2010/03/30)