ChE 253M

Experiment 4

Statistical Process Control

Background

Imagine the following hypothetical scenario: You graduate from Texas, and get your first job, getting paid \$80k a year! Your job involves working in a factory that makes metal ball bearings for skate board wheels. Ideally, the bearings are supposed to be 1.00 cm in diameter. However, as you discover when you first start working, it is impossible to consistently make bearings that are exactly 1.00 cm in diameter. Typical data may look like the following over time as the bearings come off the assembly line (for example):

Time	Diameter		
1	1.00		
2	1.05		
3	1.03		
4	0.97		
5	1.02		
6	0.95		
7	1.00		
8	1.01		
9	0.96		
10	1.01		
Avg:	1.00		



Cutaway of a ball bearing

As an engineer, you must ask yourself the following: Does a diameter of 1.02 cm indicate that there is something wrong with your process? Would your customers (skate board manufacturers) buy a ball bearing with a diameter of 1.02 cm?

As you know, random, statistical fluctuations in your process result in variations in the diameter, which is an unavoidable aspect of your process. However, processing errors (such as wear in equipment, a malfunctioning pump, or operator errors for instance) also results in variations (non-random, assignable variations) in the diameter. Your first assignment on the job is to create "control charts" that will allow you to distinguish between acceptable statistical variations (i.e. 1.00 +/- 0.06 for instance) and those that could indicate problems in your process, which would require intervention. This will allow your company to save money and time by identifying processing issues before they become a significant problem.

You also have to decide if this process meets the specs required by your customers (i.e. skaters). For instance, skateboard bearings may only work if they are 1.00 ± 0.02 cm.

Purpose

The primary purpose of this experiment is to learn how to determine whether a process (such as ball bearing manufacturing) is "in control" and how to make and interpret control charts. Furthermore, you will be using an estimation method to create the control charts, which will be compared against a standard deviation method (an estimation method used when there is insufficient data to calculate the standard deviation with confidence).

Specifically, you will be analyzing a continuous mixing process (mixing a stream of green dyed water with a stream of pure water), which is a generic process that you may encounter in industry. Based on your data, you will quantify the benefits of an in-line mixer in this process.

Statistical Process Control Background

SPC is used as an analytical tool for quality improvement programs, just like the hypothetical example of the ball bearing. It was invented in the 1920's by Walter Shewhart at Bell Research Labs. SPC really took off in the 1950's when Edward Deming introduced SPC in Japan to help in the rebuilding of the country after the war (which is widely pointed to as a reason for the quality of Japanese products). As an aside, other important techniques used by industry in quality improvement includes designed experiments to minimize research efforts in the creation of new products, Pareto and fishbone diagrams for determining cause and effect in problem solving, and "brainstorming" sessions among the people involved with a project.

As seen in the ball bearing example, <u>random</u> variation in a manufacturing process is ALWAYS present. A quality improvement program (such as the one you may be hired to do!) focuses on understanding and correcting the causes of variation due to <u>random fluctuations</u> and / or inconsistencies in equipment and materials. Understanding of the process variability comes through construction of charts showing the variation with time of the process. An example of raw data plotted below:



Performing statistical analysis on this data will help to create the actual "control charts". A control chart helps illustrate the state of statistical reproducibility or control of the process (i.e. is the process "in control"). When we say **in control**, we are saying that all variabilities in the process are results of unimportant and uncontrollable random fluctuations. Once the process is stable (i.e. operating with only random fluctuations), identification of assignable sources of variation can begin in order to improve product quality. On the other hand, a process may experience more serious types of variability in key performance measures. These sources may arise from one of several nonrandom "assignable causes", such as operator errors or improperly adjustments on the machine. A process operating in this state is considered **out of control**. In short, a <u>control chart is intended to detect the nonrandom or out-of-control states within a process</u>.

So in our example, let's say you determine through statistical analysis that random fluctuations result in 99.7% of the bearings being 1.00 ± 0.06 cm. Therefore, during future processing it is safe to assume that if you get a bearing of 1.09 cm, then there is probably something wrong with your process (i.e. it may be out of control).

As mentioned earlier, the statistical variability of the process can be used to calculate the process capability index, which predicts the level of product quality that can be expected from the process. So if your customers require bearings of 1.00 ± 0.08 cm, and your "in control" process makes bearings of 1.00 ± 0.06 cm, then you will have a high process capability index because essentially all your parts will meet your customer's specifications (as long as it stays in control).

It is important to track both the variability of your process and the average value of the product. That is, you could have low variability, but produce a product at a value that is off spec, which is clearly undesirable.

In this experiment, you will produce means charts (\overline{X} charts), range variability plots (R-charts), and process capability index (Cp) to determine the quality of control of your process.

The Process for this Experiment

As previously mentioned, you will be mixing two streams in this experiment; a stream of green dye in water with a stream of pure water. The concentration of dye in the streams will be measured before and after an "in-line" mixer using a spectrometer (discussed later).



Figure 1. Process Flow Diagram

This process was chosen to mimic a process encountered in industry. To draw a comparison, you might imagine a similar process used to make LongHornade, where a freight train brings in a concentrate of colored sugar water that must be diluted before packaging. It is your job to ensure that the process is in control.

The "accumulator tank" will be used to calibrate the rotameter (by measuring volume versus time to get flow rate). Similarly, the accumulator tank will be used to measure the flow rate of the peristaltic pump.

The "detection" flow cells are discussed under "spectrometry" later in this hand out. Essentially, we will be using the absorbance of light by green dye to determine the dye concentration. The dye concentration will vary with time, which will be the basis for your control charts.

Do you expect more variability in green dye before or after the mixer? Do you expect the average dye concentration to be different or the same before and after the mixer?

Preparing Control Charts

Control chart techniques are differentiated between those for variables (e.g. ball bearing diameter measurements) and those for attributes (e.g. whether product is defective or not). In this experiment, charts for measurements will be constructed for the dye concentration made from blending a 0.20 wt% green dye solution with water (see Figure 1). The target concentration of your product stream will be 0.14 wt% once the water is mixed with the 0.20 wt% "stock" solution.

Data for control charts are best if arranged in a logical way to reflect the process variation. For this reason measurements are usually tabulated in <u>subgroups</u> to reflect these differences. For example, day-to-day or sample-to-sample variations may be important in your process. Subgroups should be chosen such that variation within the subgroup is minimized.

For example (continuing with the ball bearing example), suppose the factory you work in has operators that work in shifts of 8 hours, and samples are taken every hour. You decide that operator variation is something worth tracking (maybe operator Joe measures differently than operator Billy), so you divide your subgroups into these shifts (8 data points per subgroup):

Time	Value	Subgroup	X-bar	R	S
1	1.00				
2	1.05				
3	1.03	1	1.014	0.08	0.0305
4	0.97				
5	1.02				
6	0.95				
7	1.00				
8	1.01	2	0.986	0.06	0.0258
9	0.96				
10	1.01				

The first step in the analysis of your data is to calculate the average (X-bar or \overline{X}), range (R), and standard deviation (S) of each subgroup of data, where:

- $\overline{\mathbf{X}}$ = numeric average of the measurements within each subgroup
- R = absolute value of the difference between largest and smallest values in subgroup
- S = standard deviation of the measurements within each subgroup

The X-bar, range, and standard deviation values are then plotted as a function of time. Also

plotted on each chart is a centerline that is the average of all the X-bars (X) and the average of all the subgroup ranges or standard deviations. Upper and Lower Control Limits for both $\overline{x}\,$ and R are also shown on the chart as horizontal lines.

UCL and LCL are essential for determining if the process is in-control, as they give an indication of whether the process is considered in control or not. For example, see the hypothetical figure below:



Figure 2. Example Control Chart

As you can see from Figure 2, all of the measurements vary about the X line, and stay within the UCL and LCL. Therefore, the process is considered in control. If points fell outside the UCL

or LCL lines, then the process would be considered out of control (technically, there is a 100-99.7 = 0.3% chance that a point will fall outside the control limits and still be in control). The center line represents the average value, and generally reflects the ideal value of the measured parameter.

How do I calculate the UCL and LCL?

Control charts are based on statistics, and therefore the variation is reflected in the standard deviation in a random system. The control limits are simply $\mu \pm 3\sigma_x$, where:

$$\sigma_{x} = \frac{\sigma}{\sqrt{n}}$$

 σ = the standard deviation of each subgroup of measurements,

n = the number of measurements within each subgroup,

 μ = the process average

Using ± 3 standard deviations is a generally accepted philosophy on control charts, but you could always pick a tighter control limit if operating out of control is costly for your process.

X-bar Control Limits:

Upper Control Limit: UCL_{x-bar} = μ +3 σ_x Lower Control Limit: LCL x-bar = μ -3 σ_x

In an ideal world, you would take an infinite amount of measurements, and calculate a standard deviation based on your data. Unfortunately, in the real world you are limited in the number of samples taken due to associated measurement cost (for example: analysis labor cost, lost product, down time, etc.). As a result, often you have to estimate the variation based on a small sample population. In this lab, we will use both standard deviation and the estimation techniques and compare the results.

The estimation technique requires at least 20 subgroups of samples taken to be meaningful. An

 \overline{X} is calculated for each subgroup, and these are averaged to give X, which is assumed to be the mean (µ) for the process (i.e. the centerline on the \overline{X} control chart). The standard deviation is then estimated based on the ranges of the subgroups. \overline{R} is the average of the ranges for all subgroups.

An estimate of $\,\sigma\,$ (which we will call $\,\widehat{\sigma}\,$) is obtained by

$$\widehat{\sigma} = \frac{R}{d_2}$$

where d_2 is a constant depending on the sample size. This estimation method is rooted in statistical theory, and explained on page 631 of your book. It is important to remember the assumption of normality, which is implicit in the x-bar chart.

Therefore

$$UCL_x = \overline{\overline{X}} + 3 \frac{\overline{R}}{d_2 \sqrt{n}}$$
 $\overline{\overline{X}}$ = centerline $LCL_x = \overline{\overline{X}} - 3 \frac{\overline{R}}{d_2 \sqrt{n}}$

Or to simplify, the variable A₂, is defined such that:

$$UCL_{x} = \overline{\overline{X}} + A_{2}\overline{R} \qquad \qquad LCL_{x} = \overline{\overline{X}} - A_{2}\overline{R}$$

R-bar charts:

X-bar charts generally give an analyst an idea of the center of location of product (i.e. how the average value varies). R-bar charts are equally important in that they give an indication of process variability. On an R-bar chart, the centerline is given by the average of the ranges R for each subgroup, noted as \overline{R} . Again, an estimation method is used to calculate the control limits (see book for more detail).

 $UCL_R = D_4 \overline{R}$

 $LCL_R = D_3 \overline{R}$

The values for Ai and Di are given in the appendix.

Process Capability Index

Since all manufacturing processes exhibit randomness, there are limits to the quality of material that can be produced. This can be expressed as the ratio of the specification requirements to the statistical variability of the process, where the spec requirements are usually set by marketing, your customers, the consumers, etc.

As you know, for truly random fluctuations, 99.7% of the data falls within \pm -3 standard deviations based on statistical analysis (i.e. we are dealing with "normal populations", with 99.7% confidence limits, which is just another way of saying the data falls within \pm 3 standard deviations).

With that in mind, the Process Capability can be calculated using:

$$\mathsf{C}\rho = \frac{USL - LSL}{6s}$$

Cp = Process Capability IndexUSL = Upper Specification Limit LSL = Lower Specification Limit s = standard deviation

At the same time, the standard deviation can be estimated from R-bar if there are not enough data points for an accurate calculations of standard deviation:

 $s = \overline{R}/d_2$

Going back to the ball bearings example, due to random fluctuations, your process may produce 99.7% of product between 1.00 \pm 0.06 cm (i.e. \pm 3 σ). However, your customers may only need bearings with a spec of 1.00 +/- 0.08 cm. Therefore, your Cp will be:

$$Cp = \frac{1.08 - 0.92}{6 \times 0.02} = \frac{1.08 - 0.92}{1.06 - 0.94} = \frac{0.16}{0.12} = 1.33$$

Lastly, you want a Cp of at least 1.3 for a process to be considered <u>capable</u>. Why is a high Cp desirable?

Interpretation of Control Charts

When using control charts to monitor a process, they must be kept up to date and the limits should be recalculated any time that an improvement or change of any type is made. Limits should be also recalculated if the charts show a distinctly new pattern for an extended time. The reason for this is that the randomness of a process, which is statistical control, is due to chance alone. A change made to the equipment or the operation will cause the pattern of randomness to change and will be reflected in the charts. For a process in statistical control the pattern of points on the charts should be random with no trend lines or out of control points (i.e. 99.7% of the points should be within the control limits).

Many rules for interpretation of control charts can be found in literature. In general the R chart should be examined first since it is a measure of variability of the process. If the R chart is out of control most likely there is a problem that must be resolved. The X-bar chart gives information about the level at which the process is operating. Process cycling, stratification of material (causes sampling from different populations), and autocorrelations (sampling from large back mixed tanks) can results in nonrandom data patterns. What are the issues in this process that may cause non-random fluctuations?

If your process is out of control, you will want to check if the data is both **normal AND random**. What is the probability of getting 5 points in a row above the mean for a truly random process (1/2 raised to the fifth power!) Therefore, trends in the data (i.e. a slope), or a significant number of points in a row above (or below) the mean usually indicate a problem in the process. The two biggest sources of non-random data are (1) something malfunctioning in your process OR (2) correlation in your data (correlation means that each point is non-independent from the previous measurement; you can see this kind of data in unemployment over time. Correlation occurs if you sample too frequently). As an engineer, it should be your top priority to identify problems in the process and fix them as quickly as possible.

Spectrophotometry

How will you actually measure the concentration of green dye in the mixed stream? Spectrophotometry refers to the use of light to measure chemical concentrations.

To understand the basis for spectrometry, imagine taking a glass of water and adding green dye to it. As you add more dye, the contents of the glass get darker. This is due to the fact that the dye is absorbing light. In other words, absorbance is a function of concentration. Along the same lines, if you have a graduated cylinder of green dyed water, it will be easier to see through the cylinder horizontally than vertically. In other words, absorbance is a function of path length (i.e. how much of the green dye does the light have to travel through?). Spectrophotometry

takes advantage of this phenomenon to measure concentration. The light in this experiment comes from a tungsten halogen lamp, which emits a spectrum from 300-1000 nm wavelength (for the absorbance of green dye by UV light, see the appendix). The light is guided using fiber optics to a "flow cell", which is best illustrated by a picture.



Figure 3. Flow Cell Diagram

As the light enters the "flow cell", it crosses through ~ 0.475 cm of the green water. The dye in the water absorbs certain wavelengths of light. As a result, only the transmitted light will arrive at the detector. Transmittance, T, is defined as the fraction of the original light that passes through the sample, where I_0 is the intensity of a particular wavelength of light incident on the sample and I is the light intensity of the same wavelength emerging from the other side of the flow cell.

$$\Gamma = \frac{I}{I_0}$$

The detector is called a spectrophotometer, or spectrometer for short. As the light enters the spectrometer, it is diffracted into a spectrum of its wavelengths (like a prism!). Each wavelength is directed onto a separate photodiode (there is an array of photodiodes), which triggers an electrical signal that is dependent on light intensity. This electrical signal is processed by the computer, which will measure the intensity at each wavelength, I. To get I_0 , the spectrometer must be "zeroed" by taking a measurement of pure water such that any change in incident light (I_0) will be associated with green dye. On a similar note, the spectrometer must also be zeroed separately by taking a "dark measurement", which is just the read out given by the spectrometer when no light reaches the detector. Inevitably, there will be some noise and random signal due to thermal effects and small errors in the detector. This is eliminated to an extent with the dark measurement.

Absorbance is related to transmittance by Beers Law:

Absorbance = A = -log
$$\left(\frac{I}{I_0}\right) = \varepsilon c I$$

where A is absorbance (unitless), *l* is transmitted light, I_0 is the reference light intensity (i.e. the intensity of light when no dye is present), *c* is concentration, *l* is path length, and \mathcal{E} is the extinction coefficient, which gives an indication of the "absorbtivity" of the dye. The extinction coefficient can be determined from the attached plot of absorbance vs. wt.% dye, which was taken with a 1 cm path length spectrometer.

Procedure

The overall goal of this experiment is to learn about control charts. However, in doing so, your initial goal will be to create a "product" stream of 0.14 wt% green dye. Refer to process flow diagram to clarify any points throughout the procedure. You will need to do several things before you can actually acquire data.

- 1. Make up at least 40L of a 0.20 wt% solution of green dye in the bulk tank (see flow diagram). Once the solution is made up, use the circulation pump to mix the dye.
- 2. Calibrate the rotameter by measuring volume versus time in the accumulator tank (i.e. flow rate as a function of rotameter setting). Take care to be consistent when reading the rotameter. The rotameter must be calibrated so that you can figure out the flow rate of the pure water stream.
- 3. Calibrate the spectrophotometer software by taking a Dark Measurement and a Reference Measurement; so the software will be able to automatically convert intensity to absorbance. The TA will explain how to use the spectrometer software.
- 4. Measure the flow rate of the peristaltic pump once the dye solution is made up. This will give you a flow rate for the dye stream. Flow rate will be measured by taking volume versus time measurements in the accumulator. <u>Do not adjust the flow rate of the pump, we will leave this constant</u> (i.e. we will only adjust the water flow rate to control the final dye concentration).
- 5. Measure the concentration of the bulk dye solution using the spectrometer. Take at least 500 data points to monitor how the system performs in the absence of water and to measure your bulk solution concentration.
- 6. At this point, you know the flow rate and concentration of the dye stream, plus have a calibration of the rotameter. This will allow you to calculate and select a rotameter setting such that the final concentration should be 0.14%.
- 7. Once both the green dye and fresh water streams are flowing for a few minutes, take at least 1000 data points using the spectrometer software. Absorbance versus time data will be taken by the computer. Absorbance can then be converted to dye concentration using Beer's Law. You will assume that the path length of the post-mixer flow cell is 0.475 cm, and then calculate the path length of the pre-mixer flow cell based on your measurements.
- 8. Insert your data into JMP and create control charts. Determine if your process is in control and determine causes for being out of control.

Note: Attached is a calibration chart for the dye used in this experiment. We will only be following the absorption of light at 630 nm, although green dye has two absorption peaks (see appendix).

By the end of the experiment, you should have the following data:

- 1. Rotameter flow rate calibration
- 2. Peristaltic pump flow rate
- 3. Spectral data from the stock dye solution, before and after mixer
- 4. Calculated rotameter setting for 0.14 wt% dye concentration
- 5. Spectro data of dye water mixture, before and after mixer

Hints for Report Write-up

The purpose of this experiment includes:

- 1. Learn about control charts (use JMP to create the charts, make it <u>clear</u> that you know how JMP works in either your results or sample calcs)
- 2. Use control charts to analyze process and any sources of variations.
- 3. Use control charts and supplemental calculations to analyze the benefits of the in-line mixer.

Control charts

For the purpose of this experiment, we will use **subgroup size of 10**. In other words, for those 1000 point measurements, lump them in groups of 10 in order to calculate X-bar and R values, and the charts should end up with 100 points plotted.

Within the lab report, there should be at least 10 graphs:

Stock dye solution (4 graphs):

- 1. Pre mixer X-bar and R charts, est method
- 2. Post mixer X-bar and R charts, est method

Dye + water Mixture (6 graphs)

- 1. Pre-mixer X-bar and R charts, est method
- 2. Post-mixer X-bar and R charts, est method
- 3. Pre-mixer X-bar chart, std dev method
- 4. Post-mixer X-bar chart, std dev method

Some questions to consider during control charts setup and analysis:

- 1. Should your X-bar-bar be the same for pre vs. post mixer flow cell?
- 2. Are your charts in control? If not, why?
- 3. Why do you see variation in the bulk solution even though there is no mixing going on?
- 4. Why would you use one method over the other? Compare the results of the two methods.
- 5. For the post mixer control charts, what is the primary source of variability? What is the primary source of non-randomness in the X-bar data?

Mixer benefit analysis

There are <u>at least</u> four ways to show the benefits of the mixer.

- 1. Graph "raw" concentration data on the same scale.
- 2. Compare R-bar, should R-bar be greater before or after mixer?
- 3. Compare Cp values.
- 4. Reynolds Number and how it related to mixing.

Contradiction?

At this point, you may find that the mixer provides benefits, but the post mixer flow cell may indicate the process to be statistically out of control. You need to understand why it is out of control before you can truly recommend the mixer. Remember, an in-control process has only

normal AND random data. So to analyze why the mixer is out of control, you need to see if it is normal and random.

- 1. Is it Normal (show this using histogram)?
- 2. Is it Random (you can either talk about trends in the data, or regions where many points are on one side of X-double bar), or, if you are really slick, you can RANDOMIZE your data and show that R-bar changes. The ONLY conclusion that can be drawn by a change in R-bar is that your data isn't truly random. This does **NOT** magically make your process in control. The logic is the following: If your process is random to begin with, and you randomize it, then nothing should change!
- 3. If criteria 1 OR 2 is not met, then explain why! As an engineer, it should be of utmost importance to determine why your data is non-normal or non-random. It may help to look at the pure dye bulk data. Remember, in that part of the process no water was running, so error associated with the process (pumps, water, etc) will be eliminated (i.e. you are only looking at detector error essentially). Also, keep in mind the way the methods are calculated; the estimation method is very "sensitive" to non-random data, but the requirement for normality is not as stringent.

Recommendations

Now that you have done all this analysis, hopefully you can offer a solid argument as to why we should or should not use the mixer. Also, based on your analysis of the sources of out-of-control behavior, provide some recommendations that may make the process run in control. Would you recommend the mixer for a factory? What changes would you recommend?

Finally

The more you quantify, the better off you will be grade-wise, but also more scientifically sound! For example, rather than saying the "standard deviation method and estimation method are similar" actually QUANTIFY it by saying, "the control limits (UCL-LCL) for the std dev method are 3% larger than the estimation method, which is an insignificant difference".

Please answer these questions in the appendix section your report.

- 1. Assuming that the post-mixer flow cell has a path length of 0.475 cm, what is the path length of the pre-mixer flow cell? Use the data from the dye concentration measurement.
- 2. Why does the green dye have two absorption peaks? Many materials absorb at multiple wavelengths, but green dye absorbs at two peaks for a specific reason.
- 3. Is the data truly random? What are the implications of your findings? (As a hint, you can use JMP to plot the distribution of the data. Make a histogram of the data. Does the distribution appear normal?)
- 4. What are the pros and cons of using a spectrometer for this experiment?
- 5. Calculate the Cp for your experiment. Your marketing department has requested a product with a spec of your mean +/- 0.003wt%. (i.e. your mean is ideally 0.14 wt%, but just use whatever mean you measure). There will be a Cp before and after the mixer for each experiment. Discuss the values you get and their implications. Is the Cp value acceptable?
- 6. Calculate the average velocity in the ¼ " tubing (i.e. is sampling frequency an issue? Is the flow turbulent?)
- 7. Is the pre-mixer or post-mixer more "sensitive" (i.e. tighter control limits) to small nonrandom fluctuations in your process? Is this a good thing?

References

Johnson, Richard A. Miller & Freund's Probability and Statistics for Engineers, Sixth Edition, Prentice Hall: Upper Saddle River, NJ, p. 500-520.

John, Peter W. M. Statistical Methods in Engineering and Quality Assurance, Wiley: New York, 1990.

TABLE OF CONSTANTS FOR CONTROL CHART CALCULATIONS

Subgroup Size (n)	D ₃	D_4	A ₂	d ₂
2	0.00	3.27	1.88	1.12
3	0.00	2.57	1.02	1.69
4	0.00	2.28.	0.73	2.05
5	0.00	2.11	0.58	2.32
6	0.00	2.00	0.48	2.53
7	0.08	1.92	0.42	2.70
8	0.14	1.86	0.37	2.84
9	0.18	1.82	0.34	2.97
10	0.22	1.78	0.31	3.07
11	0.26	1.74	0.29	3.17
12	0.28	1.72	0.27	3.25
13	0.31	1.69	0.25	3.33
14	0.33	1.67	0.24	3.40
15	0.35	1.65	0.22	3.47
16	0.36	1.64	0.21	3.53
17	0.38	1.62	0.20	3.58
18	0.39	1.61	0.19	3.64
19	0.40	1.60	0.19	3.68
20	0.41	1.59	0.18	3.73



Note: Data taken with a spectrometer at wavelength 630 nm. The path length was 1 cm for this data.



An example of the absorbance of green dye taken over the UV spectrum. For higher concentrations of dye, the absorbance would be higher, and vice versa. Notice that there are two absorption peaks. We are only tracking the peak at 630nm. Why are there two peaks?



