Getting Students to Account for Variation in their Analysis of Real ChE Processes

Milo D. Koretsky

Department of Chemical Engineering
Oregon State University
Corvallis, OR 97331-2702

Abstract
As educators are well aware, the customary educational setting in which students develop problem solving skills is one where the numerical values presented are specific and absolute. The deterministic nature of the end-of-chapter type problems is imbedded in their minds well before students even matriculate. However, as practicing engineers, they will confront the variation associated with measured data in the real world. A course in introductory statistics can force students to attend to the concept of variation. Statistics can be defined as the science of how to collect, analyze, interpret and present data with the purpose of understanding variation in a system. A key objective of introductory engineering statistics is to have students recognize variation is inevitable, and teach them skills to quantify the variation and make engineering decisions which account for it. The importance of statistics is well recognized in the chemical engineering community. For example, several recent articles in *Chemical Engineering Progress* have focused on applied statistics. Indeed, many chemical engineering programs have incorporated statistics into their curriculum. This paper describes efforts to infuse statistics into the curriculum at Oregon State University (OSU). The approach is primarily at two levels. A sophomore/junior level introductory statistics course, *Chemical Process Statistics*, has been developed. Concepts are introduced through case studies using industrial data, whenever possible. Statistical analysis of the data is discussed in terms of the physical process. In this way, the statistics and the science are coupled. However, these concepts are best synthesized when integrated with hands-on application of these concepts. To this end, statistical concepts are reinforced in senior lab. The content and structure of the introductory statistics course and efforts to integrate these concepts into senior lab will be discussed.

1. Introduction
Undergraduate chemical engineering education emphasizes *analysis* and then *design*. In the typical curriculum, the majority of the technical credit hours are devoted to fundamental science (e.g., general chemistry, physics, physical chemistry, and organic chemistry) and engineering sciences (e.g., mass and energy balances, thermodynamics, transport processes, reaction engineering, process dynamics and control). The student is then asked to *synthesize* this material in unit operations and then the capstone design course. However, the majority of graduates are hired as Process Engineers whose main focus is on *production*. Topics such as measurement system analysis (MSA), statistical process control (SPC), and design of experiments (DOE) are

essential to manufacture quality products at reduced costs.[1] In fact, upon accepting their first job offer, many entry level engineers, enroll in in-house statistics related courses such as *Practical Data Analysis, Statistical Process Control,* and *Design of Experiments.*[2]

The importance of statistics is well recognized in the chemical engineering community. For example, several recent articles in *Chemical Engineering Progress* have focused on applied statistics[1,3-6]. Many chemical engineering programs have incorporated statistics into their curriculum[2]. Two ChE specific courses in applied statistics have been recently reported[7,8]. Indeed, a survey of our alumni who graduated prior to implementation of the program described below found that statistics presented the largest discrepancy between preparation at the university relative to the importance in employment[9]. Given the curricular constraints of the program, statistics in the chemical engineering department at Oregon State University (OSU) is addressed at two levels. (1) a required introductory statistics course, *Chemical Process Statistics*, is offered in the sophomore/junior year, and (2) these concepts are reinforced in the senior unit operations laboratory. To facilitate this connection, it has been found effective to have the statistics instructor give two "refresher" lectures to the lab class.

In this paper, some educational opportunities for a statistics course to address are first anecdotally illustrated with a couple of examples pulled from student work. An overview of the chemical process statistics class at OSU is then presented. This overview includes the course goals, the course learning objectives, the industrial case studies which form the heart of the class, and the assessment of the class. Finally comments are made towards the effectiveness of integration into senior lab.

2. Educational Opportunities in Statistics from Student Work

As educators are well aware, the customary educational setting in which students develop problem solving skills is one where the numerical values presented are specific and absolute. The deterministic nature of the end-of-chapter type problems is imbedded in their minds well before students even matriculate. However, as practicing engineers, they will confront the variation associated with measured data in the real world.

An example which illustrates this mindset follows. It comes from student analysis of a Ta etch process described in the Case Study II presented later in the paper. In Figure 1, the student presents box plots of wafer thickness vs. wafer number for 10 wafers measured, in order, from 11 different lots. These data represent the normal variation associated with a stable chemical process. Of particular interest is the curve the student drew below the box plots. The student reports,

> There appears to be a sinusoidal trend in the minimums, maximums and upper quartiles. The box plot graph shows only the minimum trend for simplicity. The sinusoid has an amplitude of 0.3 microns. While small, this might be attributed to special causes (such) as cycle contaminates.

This student is desperately looking for structure in the common cause variations associated with real processes.

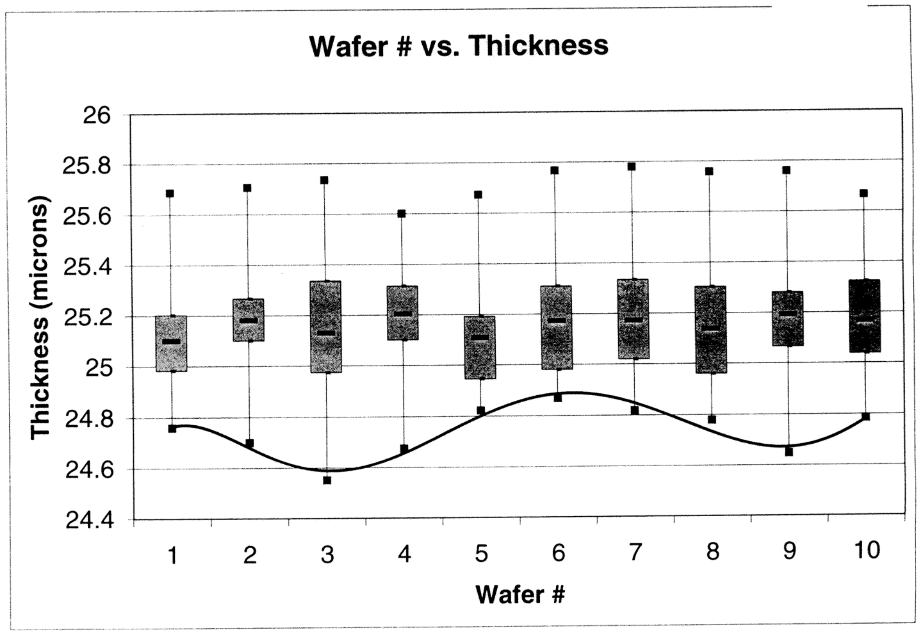


Figure 1. Example of student looking for structure in common cause variation

A second example comes from the heat transfer unit operation in senior lab. In this lab, the student group was tasked with comparing the overall heat transfer coefficient in two heat exchangers, the Armfield heat exchanger and the pilot scale heat exchanger, and fitting the data to a power law form. They ran the experiment in co-current and counter current configurations. The results are presented in Figure 2. When discussing the Armfield heat exchanger for co-current flow the group reports, " with an R^2 = 07566. Statistically, this is a poor fit for the data..." On the other hand, for co-current operation in the pilot scale they report, "R^2 = 0.9398. Statistically this is an acceptable correlation." This group is blindly applying the value of the correlation coefficient to draw a conclusion without realizing it is much easier to fit four data points to a power law correlation than seventeen. The logical conclusion that they would draw is that the correlation given by the four points for the pilot heat exchanger presented on the right of Figure 2 are statistically more reliable than the fit for the Armfield heat exchange on the left!

These examples are but two of many; you may even have your own stories. They are provided to illustrate the conceptual areas in designing an introductory statistics course and integrating statistics into the curriculum. The first example shows that students need to conceptually recognize the variation in real measurements. This task is challenging in light of the deterministic nature that science and engineering is typically taught. The second example illustrates two points. First, the statistical methodology should be understood well enough that proper interpretation is given to the statistics used, such as to R^2 in that example. There is a second, subtler, lesson as well. In the context of the second example, after learning statistics, one would hope the students can recognize to ask, "Is there any difference between the co-current and counter current configurations in the Armfield heat exchanger?" Moreover, to realize that statistics can be used to answer such a question and that it may be more appropriate to fit one expression to all the data rather than separate expressions for co-current and counter current configurations.

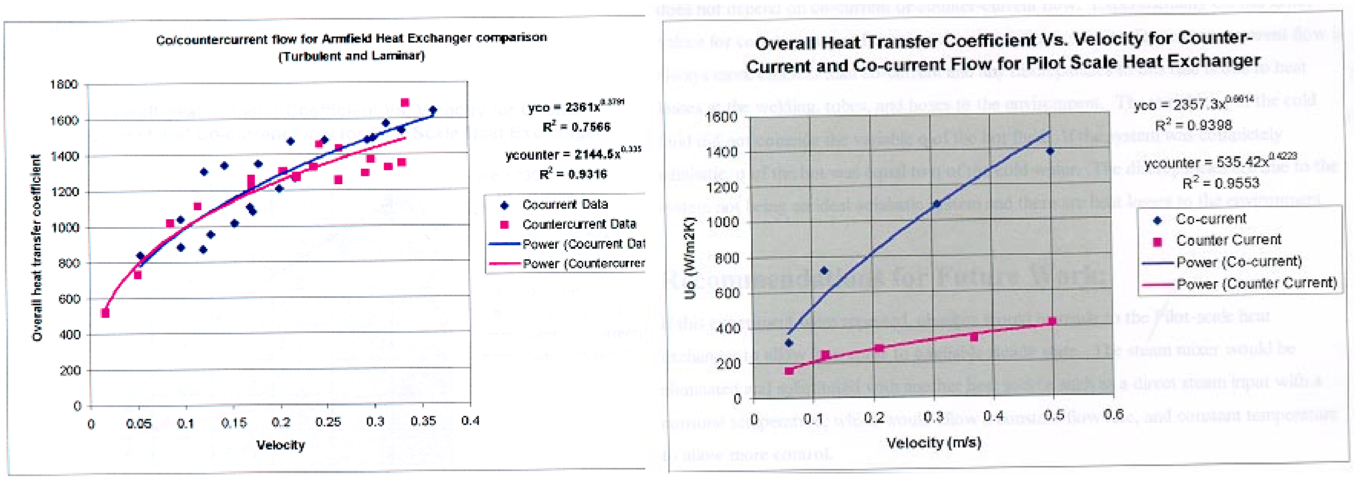


Figure 2. Correlation of heat transfer coefficients in senior lab.

3. ChE 302: Chemical Process Statistics

3.1 Course overview

Chemical Process Statistics (ChE 302) was developed to provide students exposure to statistics, in the context of the educational challenges discussed in Section 2. It especially focuses on those topics that will be useful for work in industry. The Course Goals and Course Learning Objectives are presented in Figures 3 and 4, respectively. An outline of the topics is presented in Table 1.

The course content reflects, for the most part, topics covered in many engineering statistics textbooks[10-19]. However, one topic not commonly covered is measurement system analysis. measurement system analysis evaluates the instruments used to measure a process in order to determine their accuracy and estimate the sources of variation and their extent. The process of evaluating a particular set of measurement instruments is often called a gauge study. For example, a gauge study is the first step in the formal process and equipment qualification plan developed by SEMATECH, a government supported consortium of major US semiconductor manufacturers[20,21]. In fact, many interns and recent college graduates are tasked with executing

Course Goals:

1. Develop an awareness of the utility of statistics in assessing experimental data and operating industrial chemical processes.
2. Describe the basic concepts and nomenclature associated with applied statistics, Measurement System Analysis, Statistical Process Control, and Design of Experiments.
3. Work through real industrial examples (case studies) in the field of chemical engineering to gain experience with these tools.
4. Utilize computer software (Microsoft Excel, StatGraphics) to aid in statistical analysis.

Figure 3. Course Goals for ChE 302, *Chemical Process Statistics*

Course Learning Objectives

By the end of the course, you will be able to:

1. Define major terms used in applied statistics including those on the assessment matrix.
2. Define the typical steps in analysis of data. Apply these steps when you treat measured data.
3. Calculate measures of central tendency and dispersion based on data and frequency distributions, and make the following plots: Box Plot, histogram, run chart.
4. Define variation. Identify different factors which contribute to variation. Estimate whether a source of variation is due to a common cause or a special cause.
5. List the major factors that affect measurement system analysis. Calculate the repeatability and reproducibility of a gauge based on measured data. Calculate the precision to tolerance ratio.
6. Use sampling distributions to estimate confidence intervals based on measured data.
7. Set up a hypothesis test to statistically make a decision.
8. Define when a process is statistically in control. Given data from a process, calculate control limits and capability (C_P and C_{pk}).
9. Perform least squares regression analysis to fit experimental data to an empirical model equation. Calculate the value for the regression coefficient.
10. Quantify the effect of (i) a single factor and (ii) two factors on a process by applying Analysis of Variance (ANOVA).
11. In the context of Design of Experiments (DOE), (i) set up a balanced design array, (ii) create a marginal means plots and/or an interaction plot from the experimental response, and (iii) develop an empirical model equation.

Figure 4. Course Learning Objectives for ChE 302, *Chemical Process Statistics*

gauge studies. This topic not only gives process engineers a useful tool for immediate practice but also provides a useful platform to learn about variation and variance. A more detailed description of this topic is presented elsewhere[22]. The fundamental concepts are introduced in class and reinforced in homework as in a standard lecture class. However, intertwined with this presentation are case studies using industrial data where statistical analysis of the data is discussed in terms of the physical process. In this way, the statistics and the science are coupled. The case studies covered in ChE 302 are described briefly below.

Throughout the term, two software programs are used to perform calculations on larger data sets in homework or case studies. A standard spreadsheet program, Excel, is used since it is readily available. Additionally, StatGraphics is used as an example of a statistics specific software application. StatGraphics is the university licensed software available to students. There are many common statistics packages used including: Minitab, Statistical Applications System — JMP, and Statistica. However, the objective is not to train students on a specific software package, but rather so that they get exposure to the general structure of a computer software packages. Two computer labs provide hands-on introduction to these programs.

Table 1. Chemical Process Statistics: Course Outline

Topic
1. Typical steps in analysis of data: AIChE Salary Survey
2. Measurements of central tendencies and dispersion
3. Graphical treatment of data (e.g. Box Plots and Pareto diagrams)
4. Probability
5. The normal distribution and other probability distributions
6. Variation: common causes and special causes
7. Measurement System Analysis
8. Sematech Qualification Plan
Case Study: Gauge capability of video micrometer
9. Sampling from populations - Student t distribution
Case Study: HP Ta etch linewidth batch process
10. Confidence Intervals and Hypothesis Testing
Case Study: Ta etch batch vs. continuous processes
11. Statistical Process Control
Case Study: Control of Cu Etching
Case Study: CD control in photolithography
Case Study: Ta batch etch control limits
12. Curve Fitting — linear regression
13. Analysis of Variance - ANOVA
14. Design of Experiments
Case study: Design to improve uniformity in plasma etch

3.2 Industrial Case Studies

The heart of *Chemical Process Statistics* is applying the concepts listed in Table 1 to real manufacturing data from chemical processes. Statistical analysis of the data is discussed in terms of the physical process. In this way, the statistics and the science are coupled. Moreover, it allows students to experience how these concepts often need to be extended when applied to the complexities in a manufacturing environment. Most of the case studies are taken from the microelectronics industry since this is the instructor's area of expertise; moreover, the majority of OSU BS ChEs have been placed in this industry. However, the principles can be applies to any chemical process. The case studies will be briefly described below. For more details, see references 22-25.

Case Study I: Measurement system analysis (Gauge R&R Study)

Experimental or process data are obtained through a measurement system. Values of variables such as temperature, pressure, flow rate, concentration, thickness, etc. are needed to analyze and control processes. If we are not able to adequately make measurements, we cannot hope to make useful decisions. The first step in assessing and analyzing data should be to characterize the measurement system through a gauge study. The gauge study introduced in *Chemical Process Statistics* is based on data collected by an OSU ChE interning at Merix Corporation, a printed circuit board manufacturer[a]. This study was performed to evaluate the capability of a

[a] While the data from Merix were used, the analysis used in class differs significantly.

video micrometer in use and to assess if newer instruments needed to be purchased. While the experimental design is an important component to this process, that methodology is covered later in the course and the design that was used is simply presented.

In this analysis, the class examines components of variation due to repeatability and reproducibility. The repeatability measures the variability inherent in the micrometer, itself, while the reproducibility in this gauge study is characterized by the variation in values between operators. Reproducibility can also be assigned to different environments, different gauges, etc. From these values, a precision to tolerance ratio is determined. Calculations are performed using both nested and non-nested designs, concluding the nested design is better. The students then repeat the analysis using StatGraphics. They discover that the program's default is a non-nested design. They are then shown the less straight-forward way to do the calculation for a nested design with StatGraphics. In additional to learning the details of calculating the precision of a measurement gauge, this case study introduces the class to a methodical account of different sources of variation and shows them the pitfalls of blindly using computer programs for analysis.

Case Study II: Comparing variation in two processes: batch vs. semi-continuous Tantalum Etch process

Data from two process alternatives in the manufacture of ink jet printer pen heads at Hewlett-Packard Corporation are compared. A schematic of these process alternatives is shown in Figure 1. The data represent the measurement of the width (µm) of a portion of thin-film tantalum that is defined through a wet chemical etch. The original process (Figure 1a) is a batch process in which 24 wafers (1 lot) are all etched at once in a bath of etchant. The new process (Figure 1b) is a continuous process in which each wafer is processed individually in a tool that is constantly bleeding old etchant solution while gaining fresh solution. There are six processing chambers in the tool, so the first wafer goes into chamber 1, the second to chamber 2 etc. It is not unreasonable to assume that all six chambers do not perform exactly alike, so the tool can be thought of as really six separate tools. The data was taken as follows: 11 lots of wafers were measured, 10 wafers measured from each lot, 5 "sites" were measured on each wafer.

Several studies were performed with these data. Students were first asked to examine data from the batch process. They performed summary statistics, such as lot to lot, wafer to wafer and site to site means and standard deviation, and to put the data in graphical form. From this analysis, they were asked to identify opportunities for improvement. Next, students are asked to determine if the continuous process represents an improvement over the batch process. The box plots presented in Figure 1 represent data from the continuous Ta etch process. To accomplish this task, they needed to identify that variation reduction is the primary issue, since they could easily change the centering of the distribution as needed. Thus, they needed to identify, and compare the Chi-Square distribution for each process. In Case Study V they are asked to construct control charts from these data, as will be discussed later.

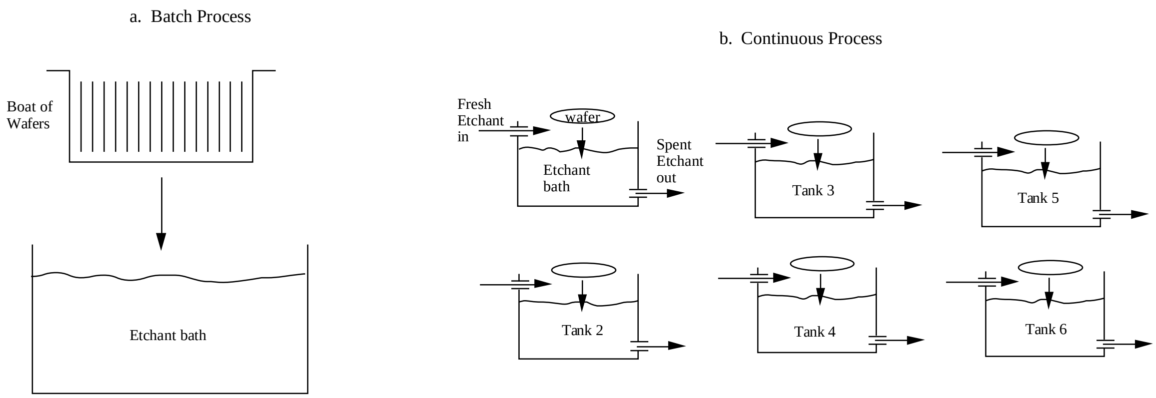


Figure 5. Comparison of the a) batch and b) continuous Ta etch processes from Hewlett-Packard

Case Study III: Statistical Process Control: Copper Etching in Printed Circuit Board Manufacturing

A principle unit process in the manufacture of printed circuit boards is the patterned copper etch. This process defines the "wiring" of the electronic circuit which connects the components which are mounted on the board. Copper is etched by the following disproportionation reaction in which cupric ion (Cu^{2+}) reacts with metallic copper to form cuprous ion (Cu^{+}):

$$Cu + Cu^{2+} \rightarrow 2Cu^{+} \tag{1}$$

Etchant is then regenerated as follows:

$$Cu^{+} + HCl + \tfrac{1}{2}H_2O_2 \rightarrow Cu^{2+} + Cl^{-} + H_2O \tag{2}$$

As time goes on:

1. Cu^{2+} accumulates in the bath as the copper clad is etched.
2. HCl is consumed by reaction 2
3. H_2O_2 is consumed by reaction 2

To control this process we must

1. Determine how much H_2O to add to dilute the cupric ion accumulation.
2. Determine how much HCl to add to account for how much is diluted by control 1 above and by consumption in reaction 2.
3. Determine how much H_2O_2 to add to account for how much is diluted by control 1 above and by consumption in reaction 2.

Attention was focused on items 1 and 2. One particular educational benefit of this system is the intricacy of the control. It is desired to measure and respond to the concentrations of Cu^{2+} and HCl. However, they cannot be measured on the production line. Therefore, specific gravity and conductivity are tracked and correlated to concentrations via laboratory quantitative analysis. When the conductivity falls below a "low" setpoint, concentrated hydrochloric acid is added to the chemical sumps until the conductivity reaches an upper setpoint. Specific gravity is monitored by the position of hydrometer float in relation to an inductive sensor. When the density of the solution increases, the hydrometer rises. When the float rises above the sensor position, the inductive contact opens. Water is added until the density decreases and the float re-makes the sensor contact. The conductivity controller setpoints are numerically entered in the PLC program and can be changed. The density controller sensor position can be adjusted. This system also includes an adjustable "stop" which prevents the float from falling below the sensor. If the stop were not there, a decrease in density would cause the float to drop, opening the contact and adding water, thus further decreasing density.

One of the problems with this control system is interactions between the measured variables, i.e., their relation to the chemical constituents. Adding concentrated HCl not only increases conductivity, but also specific gravity. Cu^{2+} also does the same. It is not hard to imagine that a control system can be configured that adds HCl when the density is to low and adds water when the conductivity is too high will also work in this application (and is, in fact, used in other systems). Hence the system can take a long time to stabilize. An adjustment for Cu^{2+} will impact HCl and vice versa. The system can also stabilize at three different points. For instance, at a particular setting, one can have "good" Cu^{2+} and HCl readings or "high" Cu^{2+} and "low" HCl or "low" Cu^{2+} and "high" HCl. Lab analysis is used to avoid these situations. Students are given sets of lab data of Cu^{2+} and HCl concentrations. They are asked to determine control limits and identify special causes. They then calculate the process capability indices C_p and C_{pk}. They discover that there exists a high number of special causes, so criteria must be applied judiciously. They also can compare the line during start-up and when it is more mature. The sluggishness of stabilization is also demonstrated.

Case Study IV: Statistical Process Control: CD control in photolithography

This case study involves implementation of SPC at Digital Semiconductor[24]. The process examined is during photolithography and etch in integrated circuit manufacturing where wafers are batch produced in lots of 25. After etch, the width of the remaining line is a very important parameter in device operation and is called the critical dimension, CD. Students are provided with a run charts of the average and range of CD for a set a batches. The problem with implementing SPC in this case is that an individual die on a wafer is more likely to be defective if its neighbors are defective.

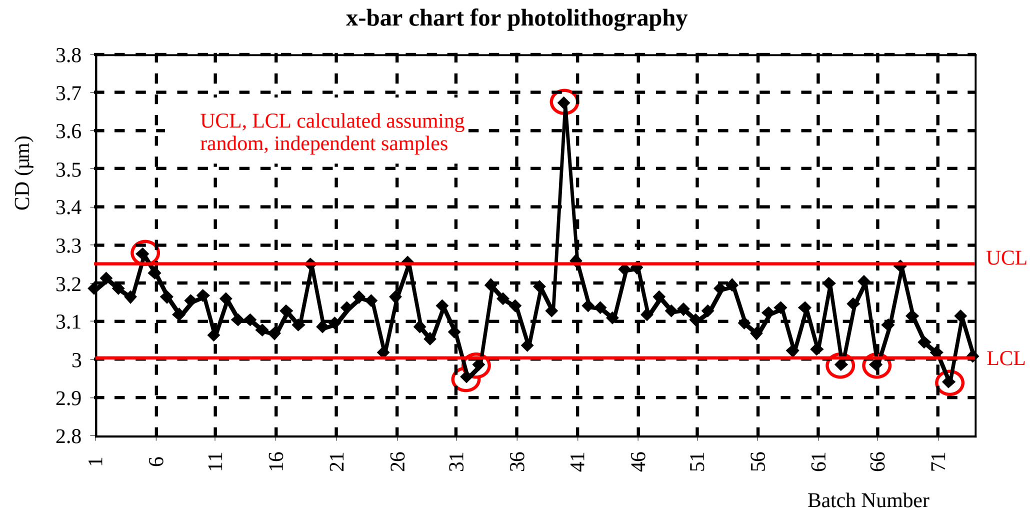
Students are asked to calculate the control limits for the photolithography data and plot control lines on the run charts. The limits they calculate when they use algorithms learned in class are shown in the top of Figure 6. They identify six points beyond the control limits, and many others just within the limits. This number is unusually high for a stable process. The possible causes for this result are discussed in class[b]. In the discussion, the *assumption* of random and independent samples is scrutinized. An alternative solution is presented, where the control limits are calculated based on an *effective sample size*, n_{eff}, as follows. Instead of assuming that the site readings form n independent, random samples, the question is posed in reverse. If the data were hypothetically selected from this process, with mean $\bar{x}$ and standard deviation, s, how many independent readings would it represent? This approach leads to control limits plotted on the

---

[b] In fact, the failure of this algorithmic approach led to resistance at Digital towards implementing SPC, and motivated the solution reported in this case study.

bottom of the figure. The process appears under control, with only the run in batch 40 showing a special cause.

In the homework assigned following this case study, students are asked to form control charts from the batch process in the Case Study II. If they use the same principle as the Digital study, they get reasonable limits for this batch process.



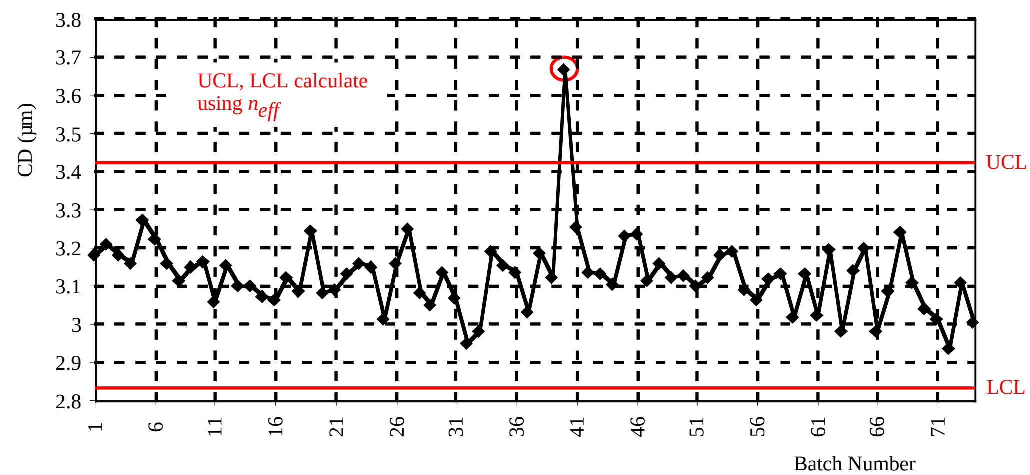


Figure 6. Upper control limit (UCL) and Lower control limit (LCL) of critical dimension after etch. The top plot shows limits calculated for random, independent samples while the bottom plot uses n_{eff}. Data modified from (24).

Case Study V: Design of Experiments: Uniformity in oxide etching

The final case study is presented after the class has looked at some straight-forward, textbook examples of design of experiments. It is the most complex study of the term and is intended to have the seminar-like quality of illustrating and extending the concepts they have been learning. This study looks at complexities in the design of experiments that arise from confounding of the interactions. The goal of this study was to improve the uniformity obtained in etching of SiO_2. It was conducted at SEMATECH[25].

First a six factor two-level design using 16 runs is analyzed. This original fractional factorial design leads to 3 significant factors and nine interactions. Thus, not enough information is available to resolve the confounding. Instead of performing an additional 16 runs, semifolding on one factor allows most of the interactions to be de-convoluted in eight runs. Thus, experimental time is reduced. Physical arguments allow the remaining confounded interactions to be determined. A confirmation run shows greater uniformity and validates the analysis. This example pushes the understanding of the class. However, in this way, it appears a large group are imparted with an understanding of interactions and design arrays on a deeper level.

3.3 Course Assessment

Assessment of *Chemical Process Statistics,* ChE 302, was performed to evaluate if the learning objectives shown in Figure 4 were achieved by the students. The assessment period covers three course offerings from 1999-2002. Feedback has been obtained from the ChE department's industrial advisory board, from student interns and from a multi-faceted in class assessment process.

The chair of the industrial advisory board wrote:

> Degreed chemical engineers are an important resource for our company as well as other related microelectronics companies. My historical experience in interviewing graduates of chemical engineering departments from a number of institutions is that there seems to be a lack of good understanding of the concepts which are addressed in *Chemical Process Statistics*. In short, this course is meeting our needs.

In the department sponsored internship program, MECOP, students give an oral presentation after each of their two, six-month industrial internships. They are asked to list the classes they found particularly useful on their internship. *Chemical Process Statistics* was by far the most common class cited, appearing in 78% of the student presentations over the last three years. The next highest class was cited, *Technical Writing*, appeared in 46%. These results should be interpreted with caution since students have not yet taken some ChE core classes before their internships. Moreover, other classes may develop an overall thinking processes rather than specific skills, and their value may not be so readily apparent to the student. None the less, clearly ChE 302 is of value.

The in-class assessment in 1999 differed from the later two years as the process was modified to conform to the departmental approach to ABET 2000. In the first year, the students qualitatively assessed the coverage and comprehension of the seven learning objectives at the time. The results, modified to correspond to those objectives presented in Figure 4, are shown in Table 2. Exam questions also were recorded according to their corresponding learning objective. The average percentage that the class obtained as well as the standard deviation is also reported in Table 2. There is good correlation between the Exam results and the students' self-assessment. This process indicates, overall, the students have achieved the stated learning objectives. The topic picked to improve in Fall 2000 was ANOVA analysis.

Table 2: Assessment Summary from Fall 1999

Learning Objective #	Exam Scores Midterm or Final	Student Assessment	
		Coverage	Comprehension
1	85–17%		excellent
3	68–21%	good	excellent
5	Case study	good	good
7	68–20%	may need more	good
8	76–19%	good	excellent
10	60–28%	may need more	good
11	72–26%	may need more	good

In Fall 2000 and Spring 2002, the following assessment tools were used:

(i) **Homework Scores (HW)**. There were eight homework assignments (approximately 30 problems total) assigned in ChE 302. Each problem was recorded according to its corresponding learning objective. The average percentage that the class obtained as well as the standard deviation is reported in Table 3.

(ii) **Midterm and Final Exam Scores**. Exam questions were recorded according to their corresponding learning objective. The average (avg) percentage that the class obtained as well as the standard deviation (stdev) is reported in Table 3.

(iii) **Lab Scores** Two in class computer based lab exercises were conducted. The average percentage that the class obtained as well as the standard deviation is reported in Table 3.

(iv) **Learning Objective Self Assessment (Self).** The survey was administered on the last day of class as part of course evaluation. The percentage mastery perceived by students along with the standard deviation is presented in Table I (4.0 being 100%).

(v) **The Language of Statistics Assessment.** The survey and the results for a sample year, 2002, are shown in the Appendix. This survey was administered both before and after the course was taught. Before the course the average of all words during the three years was **2.14** with a range of 0.11, indicating the students' belief: "I have heard this word used, but I am not sure what it means " After the course, the average was **3.66**, with a range of 0.18 indicating "I can define this word."

The homework average vs. learning objective varied from 76% to 96%. The exam average varied from 59% to 94%. The lab averages varied from 93% to 95% and the self evaluation from 66% to 88%. As indicated from the table above as well as the language of assessment survey, learning objectives 7-9 appear to be lower than 1-6. This results from several possible factors - (i) more difficult material, (ii) learning in progress when surveys were conducted, (iii) not enough coverage of material. More effort will be made next year to go through LO 1-6 more quickly to leave more time for 7 and 8.

Table 3: Measurement of Learning Objectives Fall 2000 and Spring 2003

LO	HW		Exams		Labs		Self	
	avg	stdev	avg	stdev	avg	stdev	avg	stdev
1			86%	10%			77%	24%
2							83%	14%
3	84%	13%	65%	32%			88%	11%
4			94%	3%			83%	18%
5	88%	10%	67%	22%	93%	6%	73%	19%
6	87%	12%					81%	21%
7	82%	14%					79%	27%
8	76%	10%	74%	18%			80%	23%
9	96%	6%			95%	3%	71%	25%
10	76%	21%	59%	29%			71%	31%
11	76%	21%	59%	29%			66%	30%

4. Integration of Statistics into Senior Unit Operations Lab

One measure of the effectiveness of the statistics class is the degree to which students then apply these concepts to the real measurements they are taking in senior lab. In fact, the lab problem statements typically contain statements such as,

> Your final, written report should include as a minimum:
>
> 3. ...
> 4. Use of appropriate statistical methodologies for data analysis and interpretation.

While a systematic analysis has not been undertaken, some historical evidence can be presented. It is hoped a more careful analysis from the 2003 lab can be presented at the meeting. The first quarter of the two-quarter lab sequence (ChE 414) is highly structured and focuses on the students completing 3 unit operation experiments. This second quarter of the senior lab course (ChE 415) builds on the work done in UO Lab 1. The focus is on working independently, developing a project proposal, completing experimental work and writing a final technical memorandum that includes recommendations for future work.

During 2001, in serving as a technical consultant for the microelectronics related labs in ChE 415, and discussing the statistical analysis of the students throughout the year with the lab instructor, it was determined that the use of statistical methods needed to be enhanced. Review of the written reports confirmed this belief. Thus, two "refresher" lectures of ChE 302 were included in ChE 414 during W 2002. One of the three experiments asked students to compare if different levels of several factors (velocity, flow orientation and inclusion of a steam trap) had an effect on the performance of a heat exchanger. This study is amenable to ANOVA, and was the most sophisticated statistical analysis needed for the three labs. Therefore, the first refresher lecture, focused largely on a review of ANOVA. The good news is groups correctly used ANOVA in analysis, and many groups demonstrated impressive statistical analysis for the heat exchanger experiment. The extent of statistical analysis rose notably throughout the course. The bad news is that some groups used ANOVA even when it was not needed and in fact, relied on its conclusions even when they contradicted common sense. This symptom seemed more evident in the open-ended assignments in the second term. It seems the pendulum had swung too far! This year there will be added emphasis on using judgment when applying statistical methods.

5. Summary

The approach to incorporating statistics in the ChE department at OSU has been discussed. Most of the content is delivered in a one term required class, *Chemical Process Statistics.* The use of industrial case studies reinforces learning. Assessment from the industrial advisory board, MECOP interns and a multi-faceted in class process demonstrate this class is effective. Review of the material in senior lab helps reinforce the use of statistical concepts. However, care must be taken to make sure students use solid engineering judgment in applying statistical methods.

6. Acknowledgments

Without the help, support and commitment of industry this course could never have been developed. The author is especially grateful to Michael Matthews who integrated the class into the Merix statistical process control software. Information from, support of, and technical discussions with Adrian Kriz and Lori Thompson of Hewlett-Packard and Mike McMaster, Eric Warninghoff, and Ian Reuf of Merix have also been of great value. The author is grateful for partial support provided by the Intel Faculty Fellowship Program.

7. References

1. Shunta, Joseph P., "Use Statistics to Identify Process Control Opportunities," Chem. Eng Progress p. 47 October 1996.
2. Roger E. Eckert, "Applied Statistics: Are Educators Meeting the Challenge," Chem. Eng. Ed. p. 122 spring 1996.
3. Wheeler, James M., "Getting Started: Six-Sigma Control of Chemical Operations," Chem. Eng Progress p. 76 June 2002.
4. Trivedi, Yogesh B., Applying 6 Sigma," Chem. Eng Progress p. 76 July 2002.
5. Deshpande, Pradeep B., Sohan L. Makker and Mark Goldstein, "Boost Competitiveness via Six Sigma," Chem. Eng Progress p. 65 September 1999.
4. Anderson, Mark J. and Patrick J. Whitcomb, "Optimize Your Process-Optimization Efforts," Chem. Eng Progress p. 51 December 1996.
7. Fahidy, Thomas Z., "An Undergraduate Course in Applied Probability and Statistics," Chem. Eng. Ed. p. 170 spring 2002.
8. Dianne Dorland and K. Karen Yin, "Teaching Statistics to ChE Students," Chem. Eng. Ed. 170 summer 1997.
9. McConica, C, *Self-Study Report for Chemical Engineering at Oregon State University* (2002).
10. Ayyub, B.M., and R.H. McCuen, *Probability, Statistics, and Reliability for Engineers and Scientists*, Chapman & Hall/CRC, Boca Raton, FL. (2003).
11. Hayter, A.J., *Probability and Statistics for Engineers and Scientists*, 2nd Ed., Duxbury Press, Pacific Grove, CA (2002).
12. Johnson, R.H., *Miller and Freund's Probability and Statistics for Engineers*, 6th. Ed, Prentice Hall, Upper Saddle River, NJ (2000).
13. L.L. Lapin, *Modern Engineering Statistics*, Duxbury Press; Belmont, CA (1997).
14. Levine, D.M., P.P. Ramsey, and R.K. Smidt, *Applied Statistics For Engineers and Scientists Using Microsoft Excel and MINITAB*, Prentice Hall, Upper Saddle River, NJ (2001).
15. Montgomery, D.C., G.C. Runger, and N.F. Hubele, *Engineering Statistics*, John Wiley & Sons, New York, NY (1998).
16. Vining, G.G., *Statistical Methods for Engineers*, Duxbury Press, Pacific Grove, CA (1998).
17. Walpole, R.E., R.H. Myers, S.L. Myers, and K. Yee, *Probability and Statistics for Engineers and Scientists*, 7th Ed., Prentice Hall, Upper Saddle River, NJ (2002).
18. Scheaffer, R.L. and J.T. McClave, *Probability and Statistics for Engineers and Scientists*, 4th Ed., Duxbury Press, Belmont, CA (2002).

19. Devore, J. and N. Farnum, *Applied Statistics For Engineers and Scientist*, Duxbury Press, Pacific Grove, CA (1999).
20. Czitrom, Veronica and Karen Horrell, "SEMATECH Qualification Plan" in *Statistical Case Studies for Industrial Process Improvement*, Society for Industrial and Applied Mathematics, Philadelphia, PA; American Statistical Association, Alexandria, VA (1997).
21. International 300 mm Initiative, "Metrology Tool Gauge Study Procedure for the International 300 MM Initiative (I300I)" http://www.sematech.org/public/docubase/document/3295axfr.pdf (1997).
22. Koretsky, M.D., "Applied Statistics: Introducing Analysis of Variance through Measurement System Analysis" Chem. Eng. Ed., *submitted*.
23. Koretsky, M.D., "Applied *Chemical Process Statistics* - Bringing Industrial Data to the Classroom," Proc. ASEE, Session 2213 (1998).
24. Joshi, M. and K Sprague "Obtaining and Using Statistical Process Control Limits in the Semiconductor Industry," in *Statistical Case Studies for Industrial Process Improvement*, V. Czitrom and P. D. Spagon Eds, ASA (1997).
25 Barnett, J., V. Czitrom, P.W.M. John and R.V. Leon, "Using Fewer Wafers to Resolve Confounding in Screening Experiments," in *Statistical Case Studies for Industrial Process Improvement*, V. Czitrom and P. D. Spagon Eds, ASA (1997).

Milo D. Koretsky is an Associate Professor of Chemical Engineering at OSU. He received his BS and MS degrees from UCSD and Ph D from UC Berkeley, all in chemical engineering. Professor Koretsky's research interests are in thin film materials processing including: plasma etching, chemical vapor deposition, electrochemical processes and chemical process statistics. His book, *Engineering and Chemical Thermodynamics*, is due out in December 2003.

Appendix. The Language of Statistics Survey

The Language of Statistics

ChE 399
Chemical Process Statistics
Prof. Milo Koretsky

1 = I have no idea what whis word means in the context of applied statistics
2 = I have heard this word used, but I am not sure what it means
3 = I can understand this word in context, but cannot define it
4 = I can define this word

.= After Class

	Before			After			
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation	Δmean
mean	3.64	4	0.62	4.00	4	0.00	0.36
median	3.39	4	0.74	4.00	4	0.00	0.61
mode	2.89	3	1.15	3.82	4	0.50	0.93
central tendency	2.04	2	1.04	3.77	4	0.53	1.74
standard deviation	3.15	3	0.82	3.95	4	0.21	0.81
range	3.46	4	0.84	4.00	4	0.00	0.54
dispersion	2.44	3	1.01	3.50	4	0.67	1.06
population	3.32	4	0.82	3.95	4	0.21	0.63
sample	3.46	4	0.69	3.95	4	0.21	0.49
attribute variable	1.64	1.5	0.73	3.36	4	0.85	1.72
continuous variable	1.61	1.5	0.69	3.41	4	0.80	1.80
box plot	1.93	2	1.00	3.95	4	0.21	2.03
histogram	2.43	2.5	1.07	4.00	4	0.00	1.57
distribution	3.00	3	0.98	3.95	4	0.21	0.95
normal distribution	2.54	3	1.04	4.00	4	0.00	1.46
binomial distribution	1.68	1	0.90	3.73	4	0.55	2.05
Poisson distribution	1.32	1	0.55	2.50	2	0.96	1.18
student t distribution	1.29	1	0.60	3.50	4	0.60	2.21
variation	3.00	3	0.86	4.00	4	0.00	1.00
common causes	1.71	2	0.71	4.00	4	0.00	2.29
special causes	1.68	2	0.72	4.00	4	0.00	2.32
true value	1.86	1.5	0.97	3.45	4	0.91	1.60
accuracy	3.14	3	0.89	3.86	4	0.35	0.72
precision	3.07	3	0.90	3.86	4	0.35	0.79
repeatability	2.93	3	0.86	3.86	4	0.47	0.94
reproducibility	2.89	3	0.83	3.86	4	0.47	0.97
stability	2.64	3	0.95	3.45	4	0.80	0.81
precision to tolerance ratio	1.57	1	0.84	3.82	4	0.50	2.25
confidence interval	1.50	1	0.75	4.00	4	0.00	2.50
hypothesis testing	2.04	2	1.07	2.95	3	1.05	0.92
null hypothesis	1.57	1	0.84	2.73	3	1.03	1.16
alternative hypothesis	1.61	1	0.92	2.55	3	1.06	0.94
control chart	1.71	1	0.94	3.86	4	0.35	2.15
control limit	1.68	1	0.86	3.95	4	0.21	2.28
specification limit	1.61	1	0.96	3.95	4	0.21	2.35
out of control	2.14	2	1.11	3.91	4	0.29	1.77
run	2.19	2	0.88	3.90	4	0.30	1.72
trend	2.59	3	1.05	3.86	4	0.36	1.26
capability	1.96	2	0.76	3.67	4	0.58	1.70
C_P	1.30	1	0.67	3.76	4	0.44	2.47
C_{PK}	1.11	1	0.32	3.76	4	0.44	2.65
Analysis of Variance (ANOVA	1.26	1	0.59	3.81	4	0.40	2.55
F Test	1.11	1	0.42	3.57	4	0.60	2.46
Design of Experiment (DOE)	1.74	1	0.94	3.86	4	0.36	2.12
factor	1.81	1	1.04	3.71	4	0.56	1.90
level	1.67	1	0.92	3.67	4	0.58	2.00
full factorial design	1.22	1	0.64	3.81	4	0.40	2.59
fractional factorial design	1.22	1	0.64	3.76	4	0.44	2.54
Screening design	1.19	1	0.48	2.43	3	1.03	1.24
interaction	1.56	1	0.85	3.71	4	0.46	2.16
Total	104.36	102.5	25.11	182.41	186.5	17.30	78.05
	2.11			3.69			1.58