Laboratory 5 – Univariate Data
Assigned Week of October 29, 2012
Due Week of November 5, 2012

I — Introduction:

Engineering measurements and the statistical nature of measured data are basic to all engineering disciplines. Univariate (single parameter) measurement is the focus of this laboratory exercise.

Technological problems can often be solved by people with experience. For example, most of the building of single-family low-rise homes is the responsibility of practitioners with considerable experience in the industry. Generally, engineers are not directly involved in home building because it is not necessary. Engineers have written building codes that allow low-rise structures to be constructed safely by contractors and builders.

However, when a building exceeds three stories, experience in not enough. Society does not trust practitioners who only have experience to build high-rise buildings. In fact, it is the law that "professional engineers" with a "certificate of authorization" in the structural area must design any building over three stories. So, what is it about engineering that warrants this kind of special trust and responsibility? It must be that engineers provide skills other than experience to deal such situations. The general understanding is that engineers use applied science to address problems of this kind. That understanding is valid as far as it goes—but there is a great deal more to it than that.

Engineers model technological problems to find practical solutions. Applied science is basic to the model approach to be sure, but generally applied science alone is not enough to assure the safety of a high-rise building. In addition, measured data are essential inputs used to validate science-based mathematical high-rise building models. For example, if it is a steel structure, the strength of the steel beams used must be known from measurements. But even this is not enough. By and large, there are many approximations and assumptions that have to be incorporated into any real-life engineering model. The significance of such approximations and assumptions is evaluated on the basis of engineering judgment, which in turn is based, at least in part, on laboratory and field measurements.

While a "civil engineering" example has been used to illustrate the need for an understanding of measurements in this discussion, there are equally compelling examples in all other engineering fields. It is, therefore, essential to develop an understanding of the statistical nature of measurements. We begin by considering univariate (single parameter) measurements.

II — Problem Statement:

Consider that a plastic knob is produced for an electrical device by a plastic molding machine in your company. The plastic knob will eventually contain a metal insert purchased from an outside supplier. Your company has specified an acceptable size of a particular dimension on the knob to be (0.140±0.003) inches: i.e., the technical drawing says that the dimension must be in the range of 137 thou to 143 thou to be acceptable, where *thou* is one thousandth of an inch.

$$0.137" \leq \text{acceptable} \leq 0.143"$$

Note the less-than-or-equal signs defining the acceptable range. This range is to accommodate the size of the metal insert, so it will fit into the knob.

You are the quality control engineer for the company. You are asked to report on the probability that the production process will produce knobs with acceptable dimensions, *i.e.,* in the range of 137 thou to 143 thou. You decide to sample the production during one day. The results of the sampling are recorded in the accompanying *"Production Record Sheet",* which you must complete and include in your report. You are asked to comment on whether you think the process used to produce the knob is efficient.

III — Steps and Calculations:

1. For the sample data presented in the Production Record Sheet, construct a relative frequency **table** and then plot a relative frequency **histogram** (refer to your textbook for help using Excel). Use **9** classes or bins.

   Note: Excel histograms associate the label of the current bin with the number of times data fall within the range defined as greater than the value of the previous bin, up to and including the value of the current bin: $\text{previous bin} < \#\,\text{data} \leq \text{current bin}$. The first bin includes the number of measurements with values from zero up to and including the value used to label the first bin. As an example, if the data are: {1, 1, 2, 2.1, 2.5, 3} and the bins are 1, 2, 3, 4, then the frequency histogram will be {2 ,1, 3, 0}. That is, there are 2 data points between zero up to and including 1, which is the first bin; there is one data point in the range greater than 1, but less than or equal to 2, which is the second bin; there are 3 data points greater than 2 and less than or equal to 3, which is the third bin; and there are zero points greater than 3, but less than or equal to 4.

   The first bin of your histogram should include the number of measurements that lie in the range less than or equal to 0.133"; the second bin should include the number of measurements greater than 0.133", but less than or equal to 0.135", and so on. Use an appropriate title for the histogram.

2. Describe the shape of the obtained distribution. Is it normal? Is it skewed? Does the sample have any obvious outliers?

3. What would you say if on another day you were told the dimension of one sample had been recorded as 0.114"? How would you explain this measurement?

4. For the production data, determine the following statistics using the equations provided in class and Excel, as described below:
   (a) The sample mean;
   (b) The sample variance and standard deviation;
   (c) The standard error;
   (d) The 90% confidence interval for the population mean.

Determine the quantities (a) to (d) above in the following two ways:

First: Use the appropriate equations and the "z" tables from the class notes and Chapters 18 and 19 of your textbook. You may, if you wish, use a spreadsheet program (e.g., Excel) to assist with the computations.

Second: Determine these same statistics, (a) to (d), with the Excel tool: "Data Analysis" (Descriptive Statistics).

Attach all calculations to the report (Typed or handwritten).
The results determined with these two methods (formulas and Excel) will be included in a summary table (see Table 1 below).

5. As the quality engineer, complete the Production Record Sheet by hand. The order number on the sheet is the **last 3 digits** of your student number. **Attach the handwritten completed Production Record Sheet to your assignment in the appendix.**

6. Provide two estimates for the fraction of knobs that will be within the acceptable range: *i.e.,* 137 thou to 143 thou, or $0.137" \leq \text{acceptable} \leq 0.143"$ (Note the greater-than-or-equal signs).

- First, provide an estimate of the proportion of acceptable knobs, along with a 95% confidence range, based on how many measurements in the sample are within the acceptable range.
- For the second estimate, assume the distribution of measurements is a normal Gaussian distribution and calculate the fraction using the z-tables. Provide a 99% confidence range.

7. Use the sample mean and the sample standard deviation calculated in Step 4 to generate a random set of normally distributed data (45 numbers) with Excel's "Random Number Generation" function. Select "normal" for distribution and **use the last three numbers of your student number as the random seed variable**. For your random data, determine the parameters (a) to (d) in step 4 – you can choose whether to use the formulas or the Excel data analysis tool, but remember to record your choice.

For your report you will generate a summary table (Table 1) of the results of steps 4 and 7, and then compare these results. In addition, you will discuss any differences you observe. The summary table will look like Table 1 below. Be sure to include units for the various results.

Table 1. Summary Table: [REPORT ALL NUMBERS IN THE TABLE TO FOUR SIGNIFICANT FIGURES]

Sample Statistic	Result via equations	Result via MS Excel	Random Normal Distribution using Excel with seed = ***"insert last 3 numbers of your student number"***
Mean			
Sample Variance			
Sample Standard Deviation			
Standard Error			
Population Mean with 90% Confidence Interval			

IV — Report Requirements and Deliverables:

• Using the guidelines presented in Laboratory 1, produce a formal laboratory report that summarizes your findings.

• In the "Results and Discussion" of the report:

- Report the mean, standard deviation and standard error for the measurements with the proper number of significant figures.
- Comment on the measurement reported in Step 3: 0.114".
- Comment on any differences in the summary table between the statistics determined for the sample data with the two methods in Step 4 (using the equations and Excel's data tools, *i.e.*, the numbers in the second and third columns).
- Comment on any differences between the statistics determined for the data that were randomly generated, and the statistics determined for the knob measurements?
- Comment on the proportion of knobs that you estimate will be fit-for-purpose.
- Do you think the production process is efficient?

What are your conclusions? Did you fulfill the objective? Be as specific as possible.

• Include your calculations along with any formulae (MS Equation Editor) you used. In addition, include your generated sample of random data in the appendix(ces).

• Keep in mind that the entire written report is to be limited to a title page, a one-page report and appendices for figures, tables and graphs, etc. All pages are to be well organized. **Use the file template provided on WebCT to prepare your assignment.**

Deliverables Summary   ***The lab assignment includes the following:***
1. Title page
2. One-page report
3. Handwritten completed "Production Record Sheet"
4. A Histogram of the recorded data accompanied by a frequency table
5. Calculations and answers for Steps 1-7 in Section III as appendices
6. IMPORTANT: submit as electronic version of your assignment to the specific folder*

***File name: "Lab Session_Student number.doc OR .docx" (e.g. "C3_100812345.doc": it is for C3 Lab session)**

VI — Submission and Timing:

Your report is to be submitted to the Teaching Assistant within the first 30 minutes of your next laboratory period. **<u>LATE SUBMISSIONS WILL NOT BE ACCEPTED.</u>**

VII — Marking:

Laboratory submissions will be marked on a 10-point scale: 9-10 (excellent); 7-8 (good); 5-6 (marginal); less than 5 (poor). **Be sure that you are familiar with the University's policy on plagiarism and academic integrity. Your instructors are obligated to report all suspected violations to the Associate Dean's office for investigation (see also chapter 14 at www4.carleton.ca/calendars//ugrad/current_ /regulations).**