# SOEN384 – Fall Term 2015

Homework 2 on measurement data analysis

**Individual work. Weight: 2%**

Due on Sept. 26th by midnight, submit electronically (EAS) as "Proposal 2"

**Background knowledge:**

In homework 1 all SOEN384 students wrote programs to calculate the ***distance between two gps coordinates using latitude and longitude in radians*** in a programming language of their personal choice. They recorded the effort in person-minutes it took to develop their programs and calculated the length of code (the rules for the counting of the lines of code were documented).

**Goal:** In this exercise you will apply three useful simple data analysis techniques for exploring software measurements regardless of the data distribution.

**To-Do:**

**Task 1 (1%):** apply the box-plot analysis technique seen in class to the LOC data collected in Homework 1, to be provided in an excel file <H1-results-summary.xlsx> Interpret the results.

**Task 2 (1%):** apply Bar chart analysis technique to Effort data provided in the excel file <H1-results-summary.xlsx> Interpret the results.

**Task 3 (1%):** calculate the productivity of each student. Apply Control limits chart to the collected Productivity data. Interpret the results (Hint: use the analysis model and decision criteria provided in the ISO 15939 productivity measurement model, Appendix 2, page 25).

**Submit a report containing a cover page (including your name and ID) and three sections, one section per task.**

Each section should be maximum one page long and should contain the following:

1. the corresponding charts ( box-plot for Task 1, bar chart for Task 2, Control limits chart for task 3);
2. interpretation of the results and decision-making:
   1. Task 1: acceptable range of values, brief review range of values, outliers. Possible reasons for the outliers? (if any)
   2. Task 2: reason for the variations? Can they be explained in terms of programming in a particular programming language, or student experience, or inaccurate data collection, or LOC counting rules’ variation?
   3. Task 3: How well, or not, the students were performing? Are there suggestions for large departures from the norm?
3. **Help material for task 3: What are Control Limits?**

Let us understand what you are looking at.

**Mean:** A statistically calculated number that defines the average amount of variation in your productivity data.

**UCL (Upper Control Limit):** A statistically calculated number that defines the higher limit of variation in your measurement data.

**LCL (Lower Control Limit):** A statistically calculated number that defines the lower limit of variation in your measurement data.

**How do you compute the Control Limits (UCL & LCL)?**

The general rule of thumb for calculating control limits is:

(Average Value) +/- (**2** x (Standard Deviation))

Control limits are calculated 2 standard deviations above or below the mean of your productivity data values. Anything within the control limits should be viewed as expected variation. Anything outside of control limits warrants investigation.

**Discussion on how to use the control limits in software management:**

In a software project world where there are tons of measurement data, control limits are extremely helpful in leveraging the power of statistics to be the first filter of when you should dig deeper or look for a cause of deviation:

If you measurement data and trends have variations from milestone to millstone this is a great way to isolate what is “normal” and what is “abnormal” in the data trend. Not only that but if a series of data points fall outside the control limits then it is a bigger red flag in terms of something highly impactful going wrong.