Final Assignment 2008
3TU course Applied Statistics

- the final assignment may be performed in groups of 3 students. Individuals or groups of 2 students may only be formed after agreement with the instructor A. Di Bucchianico.
- every member of a group is responsible for all parts of the final assignment and is expected to be able to answer questions on all parts
- the final assignment consists of two tasks:
    1. study a journal paper with a practical application of Statistical Process Control (to be provided by the instructor A. Di Bucchianico). The presentation should include a critical check of the used methodology and a live data analysis.
    2. writing a report on the 3 problems on pages 2 and 3 of this document. The report should contain full details on the solutions of the problems (data analyses, conclusions, mathematical derivations, and `R` code).
- on May 19, the groups
    1. give a 15 minute group presentation on the journal paper and answer questions by the instructor
    2. get a small oral examination in which they explain with solutions of the problems. A paper copy of the report should be handed over during the oral examination.

       Both the group presentation and the oral examination are closed sessions with only the group members and the instructor A. Di Bucchianico.

       In case there are too many groups, Eindhoven based groups will have their presentation and oral examinations on another date in Eindhoven.
- a group grade will be assigned to each group within two weeks (in case of clear differences between group members, individual grades may be conferred).

- in case a group report is considered to be insufficient, then the group has only 1 opportunity to send in an improved report within 1 week after notice by the instructors of the insufficient grade.

The default software for the analyses is `R`. It is recommended to use the `R` package `qcc`, which contains several SPC procedures.

1. A pilot study of new process for making a jet aircraft engine yields observations of the vane diameter (measured in inches), which are recorded in the data file `vanes.txt`. The data consists of 20 rational subgroups, each of size 5. The customer requires that the diameters must be between 0.5005 and 0.5055 inches. The ideal value of the diameter is 0.5030 inches.

   (a) Check whether the process was in statistical control during the pilot study. Give a clear description of the tools that you use in your check. Indicate which observations may be removed after closer inspection.

   (b) Perform a capability analysis, including checking the assumptions. Give estimates for the capability indices C_p and C_{pk}. What is the estimated fall-out?

   (c) Compute a β-expectation tolerance interval that contains 99.73% of the underlying distribution of the data. Compare these intervals with the results of b).

   (d) It is required by customers that the production process has a C_{pk} of 2.0. Discuss whether the pilot study indicates that the process is sufficiently capable. If the current process is not sufficiently capable, discuss whether the pilot study indicates that with minor changes the process will be able to meet the capability requirements.

   (e) An engineer claims that it is not unlikely that during mass production the mean diameter of the vanes shifts by an amount of 0.0005 inches, while the variance keeps its current value. Calculate the impact on the capability indices of such a change.

   (f) Set up two-sided "improved data driven control charts" (with in case of the minimum chart $m = 3$), see *Albers, W. and Kallenberg, W.C.M. (2006). Improved data driven control charts. TW-Report No. 1791*, using the 100 observations as individual observations (ignore the group-structure) for the following situations:

      i. use corrections to reduce the bias

      ii. use corrections for the exceedance probability approach with $\varepsilon = 0.1$ and $\alpha = 0.2$.

   Take in any case $p = 0.002$. Clearly indicate how you arrived at the control limits.

(g) (Here we do not use the data file.) Investigate which building block (the normal, the parametric or the minimum chart with $m = 3$) is chosen by the improved data driven control chart. Consider only the upper control limit. Perform a simulation study with $n = 100$. Take 10000 simulations (each time 100 observations) from

  i. the standard normal distribution
  ii. the exponential distribution with mean 1.

Report for both situations how many times the normal, the parametric or the minimum chart is chosen and give comments on the results.

2. Suppose that a measurement of the hardness of a product has a normal distribution. Every hour a sample of n units is drawn and an $\overline{X}$-chart with control limits $\mu_0 \pm 3\sigma/\sqrt{n}$ is used.

   (a) Compute how many hours it should take on average to detect a shift in the process mean of size σ when $n = 5$.

   (b) What should be the smallest sample size n so that a shift in the mean of size σ would be on average detected in less than 3 hours? Same question if we would like to detect this change within 3 hours with 95% confidence.

   (c) Assume you have two options: to sample $n_1 = 5$ elements every hour or to sample $n_2 = 10$ elements every 2 hours. Which one would you choose? Clearly indicate the criterion of your choice and the calculations that you used for your choice.

   (d) Suppose that apart from the standard out-of-control criterion ("one observation outside the control limits") we also use the following runs rule: two successive points within $\sigma/\sqrt{n}$ of the centre line. Compute the $\mathrm{ARL}_{\text{in-control}}$ for this control chart.

3. Viscosity of reactants is an important intermediate quality characteristic of a certain chemical process. A monitoring procedure must be developed in order to ensure a good quality of the end-product. Since viscosity is not a characteristic of the final product, it is not essential that viscosity has a constant level. Instead, it is essential that viscosity stays below a certain threshold. In an in-control situation the viscosity follows a normal distribution with variance 1.4 and a mean not exceeding 30.

   (a) Derive a Generalized Likelihood Ratio statistic for on-line (Phase II) monitoring this situation.

   (b) Implement the procedure from a) in `R` and derive critical values such that your procedure satisfies $ARL_{\text{in-control}} = 200$.

   (c) Run your GLR control chart on the data set `viscosity.txt` and report whether your control chart detects an out-of-control situation.