QUALITY DATA ANALYSIS

06/06/2024

General recommendations:

- Write the solutions in CLEAR and READABLE way on paper and show (qualitatively) all the relevant plots.
- Avoid (if not required) theoretical introductions or explanations covered during the course.
- Always state the assumptions and report all relevant steps/discussion/formulas/expression to present and motivate your solution.
- When using hypothesis tests provide the numerical value of the test statistic and the test conclusion in terms of p-value.
- Exam duration: 2h
- **Multichance students should skip: point b) in Exercise 1, point a) in Exercise 2**

Exercise 1 (15 points)

The concentration of a contaminant (measured in ppm) in the production of synthetic rubber is monitored over time. '230609_ex1.csv' contains the measurements collected in 50 consecutive samples.

a) Being known that a negative value is the result of a temporary miscalibration of the measuring device, fit a suitable model to these data;
b) Based on the result of point a), estimate the 95% prediction interval for the contaminant concentration in the next sample.
c) Based on the result of point a), design an appropriate control chart for these data with $ARL_0 = 250$.
d) From historical data, it is known that the most appropriate model for this process yielded a standard deviation of residuals equal to $\sigma_\varepsilon = 2.5$. Determine, with a statistical test, if the model fitted at point a) is such that the standard deviation of residuals is greater than this value (report also the p-value of the test). Discuss the result.

Exercise 2 (15 points)

A company produces aluminum laminates. The quality control department has recently introduced a statistical monitoring tool to keep under control the planarity of the laminates. It consists of an $\bar{X}$ control chart designed such that the number of samples before a false alarm is equal to 250.

a) Estimate and draw the curves of ARL_1 as a function of the mean shift δ expressed in standard deviation units with a sample size $n = 4$ and $n = 8$, respectively (show the two curves for $\delta \in [0\ 2]$ and report the ARL_1 values for $\delta = 1$ and $\delta = 2$).
b) Estimate and draw the curves of ARL_1 as a function of the sample size n for two values of the shift, $\delta = 1$ and $\delta = 2$, where δ is expressed in standard deviation units (show the two curves for $n \in [2\ 20]$ and report the ARL_1 values for $n = 3$ and $n = 6$).
c) The head of the quality control department is interested in selecting an optimal sample size n to minimize the lack of quality costs in the presence of a mean shift equal to $\delta = 2$ standard deviation units. Knowing that samples are gathered every 4 hours, the cost of planarity measurements for each laminate is $C_1 = 2$ € and an extra cost equal to $C_2 = 15$ € is due for each hour spent in the out-of-control state, determine the optimal sample size that minimizes the overall expected costs (assume the cost of the process in its in-control state as a reference baseline). Discuss the results.

Exercise 3 (3 points)

A quality characteristic X_t follows a stationary AR(1) model $X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with positive autocorrelation coefficient and known σ_ε^2. Let $E(X_t) = \mu$ and $V(X_t) = \sigma^2$. Compute the expressions of ξ and ϕ_1 as functions of μ, σ^2 and σ_ε^2.