QUALITY DATA ANALYSIS

17/07/2024

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
- **For multichance students only: Exam duration is 2h 30min**

Exercise 1 (15 points)

A cold rolling process of a sheet of steel is monitored with two sensors mounted in parallel that measure the deviation from the pre-defined thickness every minute. A collection of 100 measurements from each sensor ('x1', 'x2') is stored in `cold_rolling.csv` file.

1. *(All students)* Inspect the dataset and verify the assumptions. Discuss the results.
2. *(All students)* Using appropriate statistical tests on the original variables 'x1' and 'x2', verify if the average difference between the two sensors in the last 9 minutes of production is equal to 1 mm at 99% confidence level. Compute the corresponding confidence interval.
3. *(All students)* Suggest one or more suitable models for the data. Clearly report the model coefficients and p-values.
4. *(10 CFUs / 8 CFUs / 5 CFUs math.eng.)* Design one or more control charts to monitor the process. Clearly report the chosen control chart, the formulas and the values of the control limits (alpha_fam=0.003). Discuss the results. *Note: in case of violations of control limits, assume that no assignable cause was found and skip any transformation after designing the control chart.*

Exercise 2 (14 points)

A company specialized in the production of electrical switches for the electronic sector is implementing a control chart to monitor the stability over time of a performance metric measured during sample verification tests. Data are randomly collected every 1.2 hours with a sample size n = 6. The process runs 24/7 without interruption. Phase 1 data for control chart design are stored in the XS_PHASE1.csv file.

1. *(10 CFUs / 8 CFUs / 5 CFUs math.eng.)* Design an $\bar{X} - S$ control chart for monitoring the sample mean and variability of the performance metric such that the average time to signal (ATS) is equal to 500 h. In case of violations of control limits assume that assignable causes have been found.
2. *(10 CFUs / 8 CFUs / 5 CFUs math.eng.)* A new set of data is acquired and stored in XS2_PHASE2.csv. Using the control chart design in 1) determine if the process is in control or not.
3. *(10 CFUs / 8 CFUs / 5 CFUs math.eng.)* The head of the quality department decides to tune the settings of the testing equipment. After such operation, a massive data collection is performed, and the performance metric results to be normally and independently distributed with $\mu = 3$ and $\sigma = 1.1$. For process monitoring purposes, an $\bar{X} - S$ control chart is still used with the same settings used in the past (n = 6, target ATS = 500 h and a time between two consecutive sample collections equal to 1.2 h). Compute the new control limits for the S chart.
4. *(10 CFUs / 8 CFUs)* Based on the result in 3) compute the real value of the ATS reminding that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, and discuss the result.

5. *(10 CFUs)* Based on the result in 3) determine the probability that the S chart does not signal an inflation of the performance metric variability equal to $\sigma_I = 2.5\sigma$ at the first sample after the change.
6. *(10 CFUs)* The head of the quality department would like to increase the sample size from n = 6 to n = 12. How does the answer to points 4) and 5) change in the presence of the increased sample size? Discuss the result.

Exercise 3 (14 points)

Using the cold rolling data of the first exercise:

1. *(5 CFUs others)* Perform a PCA analysis on the original data. Discuss and compare the results obtained when using the covariance and correlation matrix (report the eigenvectors, eigenvalues and cumulative explained variance ratio).

2. *(5 CFUs others)* Reconstruct the original dataset using the first principal component obtained from the correlation matrix. Report the first 5 values of each variable. Compare with the original data in time series plots and discuss the results.

Exercise 4 (4 points)

In the following questions select one of the four possible choices as your answer and provide a short justification of your choice. Answers **without** justification will **not** receive any credit.

Question 1 (2 points) *(All students except 5 CFUs math.eng.)*
We estimate the order k autocovariance (γ_k) and autocorrelation (ρ_k) for a random variable X, from a stationary process. Which of the following is **not** valid?
a. The order $k = 0$ autocorrelation (ρ_0) is always +1.
b. The order $k = 0$ autocovariance (γ_0) is the variance of the random variable.
c. For $k \neq 0$ we have that $\rho_k = \rho_{-k}$
d. If for some $k \neq 0$ we have autocorrelation, $\rho_k = 0$, then X_{t-k} and X_t will be independent.

Question 2 (2 points) *(All students)*
In which of the following scenarios the Bonferroni correction should **not** be used?
a. When we aim to control the overall type I error in multiple hypothesis testing.
b. When we examine several lags in a test for autocorrelation.
c. When we test the significance of a linear model with k predictors in the ANOVA table.
d. When we combine several univariate control charts.