Statistical Process Control and Optimization **GIIE, NTU**

Midterm Exam (Due: Friday 5:30pm, May 3, 2024) **Prof. Argon Chen**

Note:

(1) Submit the answers in Word or PDF AND the corresponding Excel file to the COOL system by 5:30pm, 3rd of May

(2) You can use any information on the internet to help you solve the problems.

(3) You are not allowed to discuss with any human on earth (絕對不允許與任何人討論)

(5) If you are not certain about the exact meanings of the problems, please make and clearly state your own assumptions and solve the problems under the assumptions. (如果你不確定題目的意思，請自行假設並清楚說明你解題的假設)

1. (35%) The following problems arise from Big Lotto (大樂透). Big Lotto is a lottery-type game. You must choose any 6 numbers from 01 to 49 to place your bet. During the lottery draw, six prize numbers (orange) plus a special number (red) are randomly drawn, which are the winning numbers of the lottery. If more than three of your six selections match the prize numbers and the special number drawn in the same period, you will win a prize. The winning methods for all prizes are as follows: https://www.taiwanlottery.com.tw/Lotto649/index.asp. Translate the web page into English if necessary.

    (i) What would be the probability of winning the fifth prizes (中伍獎機率 p_0)? You have to show the details of your calculation procedure, rationale and assumptions.

    (ii) Based on https://www.taiwanlottery.com.tw/Lotto/Lotto649/history.aspx, use the information of total sales amount (每期銷售金額) in NT$ (note: price of each bet is NT$50) to estimate the total number of bets sold (每期總銷售注數). You may also find the number of winning bets of winning the fifth prize for each draw (每期總中獎注數) on the same website. Using the data of the latest 50 draws as of April 23 (含 4 月 23 日的最近 50 期), what is the Maximum Likelihood Estimator (MLE) of the probability to win the fifth prize. What is the standard error of the MLE? Compare the MLE with the p_0 calculated in (i).

    (iii) Develop a hypothesis test to test the probability of winning the fifth prize 中伍獎機率 (p) is equal to p_0, i.e., H_0: $p=p_0$. Use the results of the latest 50 draws to perform the test with $\alpha=0.1$.

    (iv) Construct a p chart to monitor the probability of winning the fifth prize in the latest 50 draws. What is the Type I error probability and ARL_0 for this chart?

    (v) Given that you have chosen one of the six numbers that matches the special number, what would be the probability (p_1) for you to win the fifth prize.

    (vi) What is the Type II error probability for the hypothesis test developed in (iii) if the probability to win the fifth prize is changed to p_1 calculated in (v)?

    (vii) What are the Type II error probability and ARL_1 for the p chart constructed in (iv) if the probability to win the fifth prize is changed to p_1.

2. (20%) Based on the accident death data from 24Spring-HW2.xls,

    (i) Construct a χ^2 proportion test to test whether or not the number of accident death per month from Jan. of 2011 to Dec. of 2017 follows a Poisson distribution with $\lambda=10$.

    (ii) Let $X_1, X_2, \ldots, X_i, \ldots, X_n$ be numbers of accident death of month 1, 2,…, n and are assumed to be iid Poisson random variables with parameter λ. Develop a hypothesis test ($\alpha=0.05$) to test H_0: $\lambda=\lambda_0$. Test ($\alpha=0.05$) whether the number of accident deaths per month in Taoyuan city equals to $\lambda_0=10$ after Jan. of 2018.

    (iii) Suppose λ has been changed from $\lambda_0=10$ to 13 after Jan. of 2018, what would be the type II error probability for the test developed in (ii)?

(iv) Use the accident death data from Jan. 2011 to Dec. 2017 to construct a *c* chart and then use the control chart to monitor the number of accident deaths from Jan. 2018 to March 2022. What is the type I error probability and ARL_0 of the chart?

3. (35%) Use the thickness data of **the bottom zone** in 24Spring-HW4.xls.

(i) Construct and test the hypothesis H_0 (null hypothesis): $\mu_{right}=\mu_{left}$ and H_1: $\mu_{right}\neq\mu_{left}$ with $\alpha=0.02$. *p*-value?

(ii) Construct and test the hypothesis H_0 (null hypothesis): $\sigma^2_{middle}=\sigma^2_{right}$ and H_1: $\sigma^2_{middle}<\sigma^2_{right}$ with $\alpha=0.02$. What is the *p*-value?

(iii) Plot a histogram for thickness of left position. Assuming a Gamma distribution, plot the Q-Q plot and perform the χ^2 proportion test for the thickness of the left position.

(iv) Let $\overline{X}$ be the average of the thickness of five positions on the same wafer. Use the first 45 wafers from the bottom zone to estimate the $\sigma^2_{\overline{X}}$ by the sample variance $s^2_{\overline{X}}$ and construct the Shewhart $\overline{X}$-R control chart to monitor the last 40 wafers.

(v) Assuming $\mu_0=350$ and $\mu_1=\mu_0-1.6\sigma_{\overline{X}}$, use the sequential likelihood ratio test with a chart like the one on Slide 32 of SPCO2.2 to test the $\overline{X}$ of the last 40 wafers with $\alpha=0.003$ and $\beta=0.2$. Assume $\mu_0=350$ and construct a "graphical" Tabular CUSUM charts (like the one on slide 36 of SPCO2.2.pdf) for the wafer thickness average $\overline{X}$ with the (K, H)=($0.8\sigma_{\overline{X}}$, $6\sigma_{\overline{X}}$) and ($0.5\sigma_{\overline{X}}$, $5\sigma_{\overline{X}}$) to monitor the last 40 wafers. Compare the sequential likelihood ration test and the two CUSUM control charts.

(vi) Design an optimal Tabular CUSUM chart with $\delta=1.6\sigma_{\overline{X}}$ and $ARL_0=500$. What would be the ARL_1 if the mean is shifted by $0.5\sigma_{\overline{X}}$, $1.0\sigma_{\overline{X}}$, $1.5\sigma_{\overline{X}}$, or $2.0\sigma_{\overline{X}}$. Use $\sigma_{\overline{X}}$ estimated by the pooled moving sample variances s^2_p of the first 50 wafers and construct the optimal CUSUM chart to monitor the last 40 wafers. Estimate the change point and the new shifted process mean if there are out-of-control signals.

(vii) Design an optimal EWMA chart with $\delta=1.6\sigma_{\overline{X}}$ and $ARL_0=500$. What would be the ARL_1 if the mean is shifted by$0.5\sigma_{\overline{X}}$, $1.0\sigma_{\overline{X}}$, $1.5\sigma_{\overline{X}}$, or $2.0\sigma_{\overline{X}}$. Compare the performance of this EWMA chart and the CUSUM chart designed in (iii). Use $\sigma_{\overline{X}}$ estimated by the sample variance $s^2_{\overline{X}}$ of the first 45 wafers and construct the optimal EWMA chart to monitor the last 40 wafers and compare it to the CUSUM constructed in (iii).

4. (10%) Use the thickness data in 24Spring-HW4.xls to perform the following analysis.

(i) Suppose the engineering target is 350 and the specification window is (335, 360). Use all 425 thickness readings of **the bottom zone** to estimate C_p, C_{pk}, C_{pm}, and C^*_{pm} for the overall process capability of **the bottom zone**. Assuming normal distribution of the CD, calculate the overall out-of-specification probability. How would you propose to improve the process to achieve overall $C_{pk}=2.0$? How would you propose to improve the process to achieve overall $C_{pm}=2.0$

(ii) With the engineering specifications (335, 360), what are the "within-wafer" C_p, C_{pk}, C_{pm}, and C^*_{pm} of **the bottom zone** (Note: "within-wafer" σ and $\tilde{\sigma}$ have to be estimated by calculating the pooled among-position sample variance and pooled among-site sample squared deviation from target). How would you propose to improve the process to achieve within-wafer $C_{pk}=2.0$ and within-wafer $C_{pm}=2.0$?