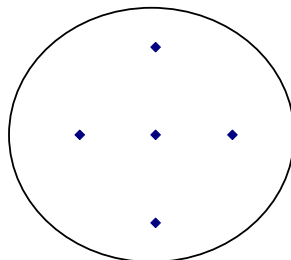


## 2019 Spring –SPCO HW#4

1. In the attached excel file (19Spring-HW4.xls), we collect a sample of waiting times for abdomen ultrasound test appointments. Among many factors that affect patients' satisfaction rating for the hospital service, the waiting time is ranked as the most important factor. The hospital has decided to closely monitor the waiting time and wants to understand more about the waiting time.

- (1) Calculate the sample mean and sample variance of the waiting times.
- (2) Assuming a Normal distribution, use sample mean and sample variance as the estimate of the mean and variance to plot the distribution of the waiting time using Excel.
- (3) Assuming Exponential distribution, use sample mean and sample variance as the estimate of the mean and variance to plot the distribution of the waiting time using Excel.
- (4) Assuming Gamma distribution, use sample mean and sample variance as the estimate of the mean and variance to plot the distribution of the waiting time using Excel.
- (5) Plot the histogram (with various range settings) of the waiting time.
- (6) Calculate the minimum, 1<sup>st</sup> quartile, median, 3<sup>rd</sup> quartile and maximum and plot the Box Plot.
- (7) Compare the histogram and box plot in (5) and (6) with the distribution shapes obtained in (2), (3) and (4). Which distribution model, normal, exponential or Gamma, can better describe the randomness of the waiting time?

2. In the attached excel file (19Spring-HW4.xls), we collect 5 readings of critical dimension (CD) from 100 wafers in semiconductor manufacturing. The 5 CD readings (CD\_site1, CD\_site2, CD\_site3, CD\_site4, CD\_site5) are from 5 positions, as shown below, on each wafer.



- a) Use the data collected from 100 wafers to calculate the sample mean and the sample standard deviations for CD\_site1, CD\_site2, CD\_site3, CD\_site4, and CD\_site5. Plot scatter plots for (CD\_site1, CD\_site3), (CD\_site2, CD\_site3), (CD\_site4, CD\_site3) and (CD\_site5, CD\_site3). Calculate the

- sample covariance and correlation coefficients for the 4 pairs.
- Plot the histograms for CD\_site1, CD\_site2, CD\_site3, CD\_site4, and CD\_site5.
  - Plot the box plots for CD\_site1, CD\_site2, CD\_site3, CD\_site4, and CD\_site5
  - Based on (a), (b) and (c), do you think the CD's on the five sites follow the same probability distribution model?
  - Calculate  $\bar{X}$  (average),  $\tilde{X}$  (median),  $\bar{X}_e$ ,  $\bar{X}_{tr(10\%)}$ , Sample Standard Deviation, Sample Range and IQR for CD\_site1, CD\_site2, CD\_site3, CD\_site4, and CD\_site5. Discuss the differences among the 4 measures of location and 3 measures of dispersion of the 5 sites.

3. Let

$$\frac{\sum_i (x_i - \bar{x})^2}{n}$$

be the variance estimator. Show that it is a biased estimate and find the bias.

4. Given the *iid* random sample  $x_1, x_2, \dots, x_n$ , taken from a Poisson distribution with parameter  $\lambda$ , let  $\hat{\lambda} = \bar{x}$ . Is  $\hat{\lambda}$  an unbiased estimate of  $\lambda$ . What is the standard error of this estimate? When would the standard error reach the smallest? How to estimate the standard error?