2019 Spring – SPC&O HW#5

1. To estimate the probability of success (p), n samples are taken: (x_1, n_1), (x_2, n_2),…, (x_n, n_n) where n_i is the number of trials of the ith sample and x_i is the number of successes observed in n_i trials.

   a) Assuming binomial distribution, derive with a detailed procedure the MLE of p.

   b) Let the other estimator of p be:

   $$\hat{p}_2 = \frac{\sum_i (x_i/n_i)}{n}$$

   c) Evaluate and compare this estimator and the MLE in terms of their biases and standard errors

2. Use the CD data provided in 19Spring-HW4.xls,

   a) Assuming normal distributions, use Excel to plot the (Q-Q) plots for CD_site1, CD_site2, CD_site3, CD_site4, and CD_site5.

   b) Assuming Gamma probability distribution, use the moment estimator to estimate the parameters of the Gamma distributions for CD_site1, CD_site2, CD_site3, CD_site4, and CD_site5. With the parameter estimates for Gamma probability distributions, use Excel to plot the (Q-Q) plots for CD_site1, CD_site2, CD_site3, CD_site4, and CD_site5.

   c) Which distribution, Gamma or Normal, can better describe the random natures of the CD's.

3. In Q-Q plot, [F^{-1}(estimated cumulated probability of the ith smallest observation.), the ith smallest observed value] are plotted on the scatter chart. Another plot is called P-P plot in which [F(the ith smallest observed value), the estimated cumulated probability of the ith smallest observation] are plotted on the scatter chart. With the waiting time data in 19Spring-HW4.xls,

   a) assuming that the waiting time is a discrete random variable possibly following the Geometric, Negative Binomial or Poisson distribution, plot the P-P plots for the three possible distributions.

   b) which distribution model fits best the waiting time observations?

4. Using 19Spring-HW4.xls,

   a) Calculate 75% confidence intervals for CD_Site1 mean and CD_Site3 mean using z-statistics with σ estimated by s.

b) Calculate 75% confidence intervals (CI) for CD_Site1 mean and CD_Site3 mean using t-statistic. Compare the CI using t-statistic with the CI using z-statistic.

c) From the CI found in b), do you think the two means are equal? How confident are you with your judgment?

d) Construct and test the hypothesis H_0 (null hypothesis): $\mu_{CD\_site5}=60$ and H_1: $\mu_{CD\_site3}<60$ with an acceptable risk $\alpha=0.25$.

e) Construct and test the hypothesis H_0 (null hypothesis): $\mu_{CD\_site1}=60$ and H_1: $\mu_{CD\_site1}>60$ with $\alpha=0.25$.

f) Based on d) and e), do you think the CD means of site 3 and site 1 are the same? How confidence are you?

6. Collect your body temperature with repeated measurements (more than 3 repeated measurements every time) several times a day as frequently as possible (e.g. before and after bed time, meals, exercises, etc.) starting from 3/25 to 4/05. Make a note of your measurement methods, i.e., types of thermometer and body location of measurement.