# Homework 6

Please complete the problems on a separate sheet of paper with your name at the top. Make sure to show your work and/or provide an explanation for each problem. Please be clear in your work. Partial credit will be given when merited. The total credit is 7 points with one bonus point.

**Problem 1.** The female students in an undergraduate engineering core course at ASU self-reported their heights to the nearest inch. The data are 60, 62, 63, 70, 63, 70, 61, 65, 64, 64, 67, 63, 70. **Calculate the sample mean (0.5 point), sample standard deviation (0.5 point), and the sample median of height (0.5 point) Construct a Boxplot (0.5 point) and a histogram (0.5 point) for the height data**. If you use R to draw the graphs or calculate the summaries, make sure to include the R codes in your solutions. (2.5 points total)

**Problem 2.** Let $X_1, \ldots, X_n$ be a random sample from exponential distribution with rate parameter $\lambda$, i.e. $f_X(x) = f_X(x) = \frac{1}{\lambda}\exp(-x/\lambda), x > 0$.

(a) Derive the maximum likelihood estimator of $\lambda$. Hint. The joint likelihood of observing a realization $x_1, x_2, \ldots, x_n$ is $L(\lambda) = \prod_{i=1}^{n}\{\frac{1}{\lambda}\exp(-x_i/\lambda)\} = \frac{1}{\lambda^n}\exp(-\sum_{i=1}^{n} x_i/\lambda)$. The log of the likelihood is $\log(L(\lambda)) = -n\log(\lambda) - \sum_{i=1}^{n} x_i/\lambda$. Take the derivative of $\log(L(\lambda))$ with respect to $\lambda$ and set that derivative to zero. Next solve for $\lambda$. The answer should be $\sum_{i=1}^{n} x_i/n$ and hence the estimator is $\sum_{i=1}^{n} X_i/n$. Check the second derivative of $\log(L(\lambda))$ to make sure it is less than zero or provide a condition so that the second derivative is less than zero. (0.5 point)

(b) Bonus question: Is the estimator biased? Derive the expected value of the estimator in part (a) using property $E(c_1X_1+c_2X_2+\cdots+c_nX_n) = c_1E(X_1)+c_2E(X_2)+\cdots+c_nE(X_n)$. Since $X_1, \ldots, X_n$ all follow the same distribution, you just need to find out the mean of the distribution. If it equals $\lambda$, then the estimator is unbiased. Otherwise, it is biased. The answer should be that the estimator we derived in part (b) be unbiased. (0.5 point)

(c) Bonus question: Calculate the Mean squared error of the estimator. Mean squared error is defined as $E(\hat{\theta} - \theta)^2$ where in our case it becomes $E(\sum_{i=1}^{n} X_i/n - \lambda)^2$. Since we show the estimator $\sum_{i=1}^{n} X_i/n$ is unbiased in part (c), which means $\lambda$ is the expected value of it, $E(\sum_{i=1}^{n} X_i/n - \lambda)^2$ becomes the variance of $\sum_{i=1}^{n} X_i/n$. Using the property that if $X_1, \ldots, X_n$ are mutually independent, then $Var(\sum_{i=1}^{n} X_i/n) = Var(X)/n$ where $X$ follow the distribution $f_X(x) = f_X(x) = \frac{1}{\lambda}\exp(-x/\lambda), x > 0$. (0.5 point)

**Problem 3.** Past experience has indicated that the breaking strength of yarn used in manufacturing drapery material is normally distributed and that $\sigma = 2$ psi. A random sample of 9 specimens is tested, and the average breaking strength is found to be 98 psi.

(a) Give the formula for finding a two-sided confidence interval for a population mean. Explain the notations. (0.5 point)

(b) Find a 95% confidence integral for the true mean breaking strength. Interpret the confidence interval. (0.5 point)

(c) What would be confidence interval if the level of confidence is 99%? (0.5 point)

(d) What sample size is needed to be 95% confident that the error in estimating the true breaking strength is less than 1psi? (0.5 point)

**Problem 4.** Of 1000 randomly selected cases of lung cancer, 823 resulted in death within 10 years.

(a) Give the formula for finding a two-sided confidence interval for a population proportion. Explain the notations. (0.5 point)

(b) Calculate a 95% confidence interval for the death rate (within 10 years) of lung cancer. (0.5 point)

(c) Using the point estimate of $p$ obtained from this sample, what sample size is needed to be 95% confident that the error in estimating the true value of $p$ is less than 0.03? (0.5 point)

(d) Using 0.5 for $p$, what sample size is needed to be 95% confident that the error in estimating the true value of $p$ is less than 0.03? (0.5 point)