Stat 330: Homework 5 (Module 5)

Show all of your work, and upload this homework to Canvas.

1. The following data set represents the number of new computer accounts registered during ten consecutive days:

$$43,\ 37,\ 50,\ 51,\ 58,\ 52,\ 45,\ 45,\ 58,\ 130$$

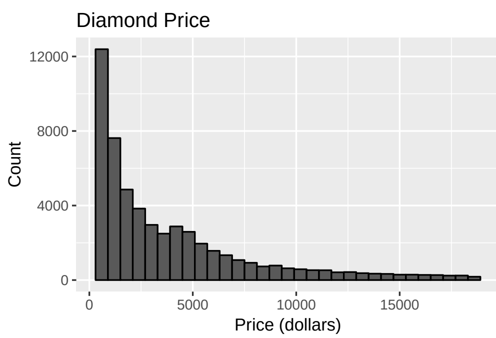
   (a) Compute the mean, median, IQR, and standard deviation

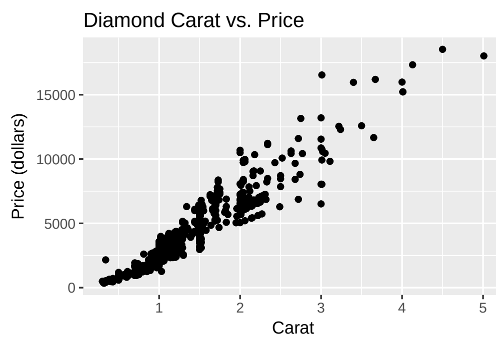
   (b) Check for outliers using the 1.5(IQR) rule, and indicate which data points are outliers.

   (c) Remove the detected outliers and compute the new mean, median, IQR, and standard deviation.

   (d) Make a conclusion about the effect of outliers on the basic descriptive statistics from (a) and (c).

2. A histogram of the price of diamonds, and a scatterplot of carat vs. price of diamonds are given below.





   (a) Describe the shape of the histogram of price of diamonds. (Where are the majority of diamond prices located? Where are the minority of diamond prices located?)

   (b) Are exponential, normal, or uniform distributions reasonable as the population distribution for the price of diamonds? Justify your answer.

   (c) Describe the relationship between carat and price of diamonds. (What happens to price as number of carats increases? What happens to the variability as number of carats increases?)

3. Suppose $X_i \overset{iid}{\sim} \text{Unif}(0, \theta)$ for $i = 1, \ldots, n.$. Suppose we propose an estimator for θ as $\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i$.

   (a) Is $\hat{\theta}$ an unbiased estimator for θ?

   (b) Calculate $se(\hat{\theta})$ (Recall the standard error of an estimator is the square root of the variance of an estimator).

4. Let $X_1, \ldots, X_4 \overset{iid}{\sim} \text{Bern}(p)$. Suppose we propose two estimators for p:

$\hat{p}_1 = \frac{X_1+X_2+X_3+X_4}{4}$

$\hat{p}_2 = \frac{X_1+2X_2+X_3}{4}$

   (a) Show that both estimators are unbiased estimators of p.

   (b) Which estimator is "best" in terms of having a smaller MSE? Calculate MSE($\hat{p}_1$) and MSE($\hat{p}_2$) (Recall that if an estimator $\hat{\theta}$ is unbiased, MSE($\hat{\theta}$) $= Var(\hat{\theta})$).

5. Suppose $X_i \overset{iid}{\sim} p_X(x)$, where $p_X(x) = \frac{1}{N}$ for $x \in \{1, \ldots, N\}$. Here, N is the parameter. Derive the method of moments estimator for N.

6. Suppose $X_i \overset{iid}{\sim} Pois(\lambda)$ for $i = 1, \ldots, n$.

   (a) Give the method of moments estimator for λ.

(b) Next we will give the maximum likelihood estimator by going through the steps:

   i. Write down the likelihood function.
   ii. Give the log-likelihood function.
   iii. Give the derivative of the log-likelihood function with respect to λ.
   iv. Set the derivative equal to zero and solve for λ in terms of the data.
   v. Report the maximum likelihood estimator for λ.

(c) If we observe the data, 7, 6, 7, 2, and 4, what are the numerical estimates of the method of moments and maximum likelihood for λ?

7. A sample of 3 observations of waiting time to access an internet server is $x_1 = 0.4, x_2 = 0.7, x_3 = 0.9$ seconds. It is believed that the waiting time has the continuous distribution

$$f(t) = \begin{cases} \theta t^{\theta-1}, & 0 < t < 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) Find an estimate of the parameter θ using the method of moments. (Give a numerical value)
   (b) Find the maximum likelihood estimate of θ. (Give a numerical value)

8. Let $X_1, \ldots, X_n$ be a random sample from the Gamma distribution with $\alpha = 3$. The pdf is shown as follows.

$$f(x) = \frac{\lambda^3}{2} x^2 e^{-\lambda x}$$

   for $x \geq 0$.

   (a) Find an estimate of the parameter λ using the method of moments.
   (b) Find the maximum likelihood estimate of λ.