Show all of your work, and *please* staple your assignment if you use more than one sheet. Write your name, the course number and the section on every sheet. Problems marked with * will be graded and one additional randomly chosen problem will be graded.

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

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

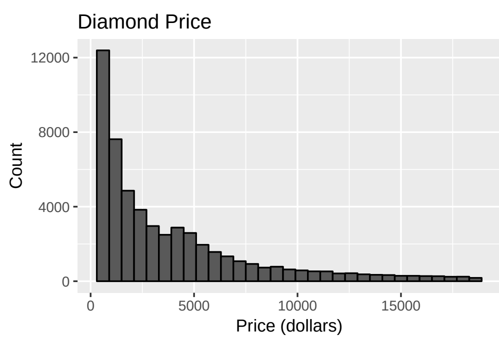
   (a) Compute the mean, median, IQR, and standard deviation

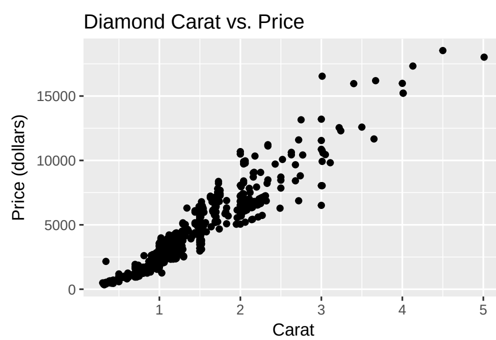
   (b) Check for outliers using the 1.5(IQR) rule, and indicate which data points are outliers.

   (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. Let $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$, where $\mathbb{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2$. The method of moments estimator for σ^2 is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, which is biased. However, it is more common to use $(n-1)$ in the denominator to get the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. The reason for using $(n-1)$ in the denominator is that it makes S^2 an unbiased estimator of σ^2. Finish the following proof to show that S^2 is unbiased:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$\Rightarrow \mathbb{E}(S^2) = \frac{1}{n-1}\left(\mathbb{E}\left(\sum X_i^2\right) - n\mathbb{E}(\bar{X}^2)\right) = \frac{1}{n-1}\left(n\mathbb{E}(X^2) - n\mathbb{E}(\bar{X}^2)\right) = \ldots = \sigma^2$$

   Fill in the dotted steps of the above proof. (Hint: for any random variable X, $\mathbb{E}(X^2) = Var(X) + (\mathbb{E}(X))^2$)

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

   (a) Derive the method of moments estimator for λ, i.e., it should be a function of the y_i.

   (b) Derive the maximum likelihood estimator for λ, i.e., it should be a function of the y_i.

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

6. 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)

7. 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 estimator of the parameter λ using the method of moments.

   (b) Find the maximum likelihood estimator of λ.