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:

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

Answer: The ordered data is: 37, 43, 45, 45, 50, 51, 52, 58, 58, 130

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

Answer:

- mean $= \frac{1}{10}\sum_{i=1}^{10} x_i = 56.9$
- median $= \frac{50+51}{2} = 50.5$
- $Q_1 = \frac{43+45}{2} = 44$; $Q_3 = \frac{58+58}{2} = 58$
  $\rightarrow IQR = Q_3 - Q_1 = 58 - 44 = 14$
- variance $= s^2 = \frac{1}{10-1}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 702.7667$
  $\rightarrow$ standard deviation $= s = \sqrt{s^2} = \sqrt{702.7667} = 26.5097$

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

Answer:

- $Q_1 - 1.5(IQR) = 44 - 1.5(14) = 23$
- $Q_3 + 1.5(IQR) = 58 + 1.5(14) = 79$
- Any values less than 23 or greater than 79 are outliers.
- Outlier: 130

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

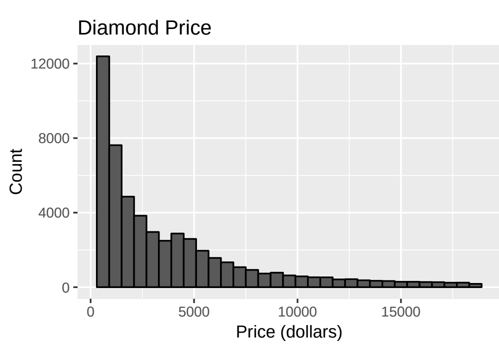
Answer: The new ordered data: 37, 43, 45, 45, 50, 51, 52, 58, 58

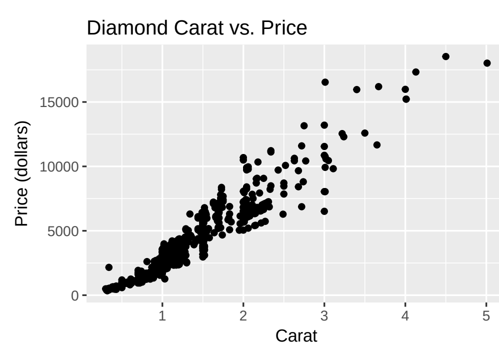
- mean $= \frac{1}{9}\sum_{i=1}^{9} x_i = 48.78$
- median $= 50$
- $Q_1 = \frac{43+45}{2} = 44$; $Q_3 = \frac{52+58}{2} = 55$
  $\rightarrow IQR = Q_3 - Q_1 = 55 - 44 = 11$
- variance $= s^2 = \frac{1}{9-1}\sum_{i=1}^{9}(x_i - \bar{x})^2 = 48.4444$
  $\rightarrow$ standard deviation $= s = \sqrt{s^2} = \sqrt{48.4444} = 6.9602$

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

Answer: The outlier increased the mean and variance. The median and IQR *slightly* increased with the outlier but not by much. Thus, the mean and variance seem to be affected greatly by outlier, but the median and IQR were not affected much by the outlier (robust)

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?)

Answer: The majority of diamond prices are in the lower end of the price spectrum between about 0 and $5,000$ dollars. After about $5,000$ dollar, the number of diamonds drop off.

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

Answer: Since the histogram of our sample has a similar shape to an exponential distribution, an exponential distribution is a reasonable choice for the population distribution.

(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?)

Answer: In general, as number of carats increases, the price of the diamond also increases. The data points are more compact for lower (< 2) carats indicating lower variability in prices for lower carats. Conversely, the data points are more spread out for higher (> 2) carats indicating higher variability in prices for larger carats.

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:

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

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

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

Answer: $\mathbb{E}(X^2) = \sigma^2 + \mu^2$ and $\mathbb{E}(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$

$$\begin{aligned}
\mathbb{E}(S^2) &= \frac{1}{n-1}(n\mathbb{E}(X^2) - n\mathbb{E}(\bar{X}^2)) \\
&= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) \\
&= \frac{1}{n-1}(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) \\
&= \frac{1}{n-1}(\sigma^2(n-1)) \\
&= \sigma^2
\end{aligned}$$

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 $\text{MSE}(\hat{p}_1)$ and $\text{MSE}(\hat{p}_2)$ (Recall that if an estimator $\hat{\theta}$ is unbiased, $\text{MSE}(\hat{\theta}) = Var(\hat{\theta})$).

Answer:

(a)

$$\mathbb{E}(\hat{p}_1) = \frac{1}{4}\mathbb{E}(X_1 + X_2 + X_3 + X_4) = \frac{1}{4}4\mathbb{E}(X) = p \implies Bias(\hat{p}_1) = E(\hat{p}_1 - p) = E(\hat{p}_1) - p = p - p = 0$$
$$\mathbb{E}(\hat{p}_2) = \frac{1}{4}\mathbb{E}(X_1 + 2X_2 + X_3) = \frac{1}{4}(p + 2p + p) = p \implies Bias(\hat{p}_2) = E(\hat{p}_2 - p) = E(\hat{p}_2) - p = p - p = 0$$

So both estimators are unbiased for p.

(b)

$$Var(\hat{p}_1) = \frac{1}{16}Var(X_1 + X_2 + X_3 + X_4) = \frac{4}{16}Var(X) = \frac{4p(1-p)}{16} = \frac{p(1-p)}{4}$$
$$Var(\hat{p}_1) = \frac{1}{16}Var(X_1 + 2X_2 + X_3) = \frac{1}{16}\left[Var(X_1) + 4Var(X_2) + Var(X_3)\right] = \frac{6Var(X)}{16} = \frac{6p(1-p)}{16}$$
$$MSE(\hat{p}_1) = Bias^2(\hat{p}_1) + Var(\hat{p}_1) = 0^2 + \frac{p(1-p)}{4} = \frac{p(1-p)}{4}$$
$$MSE(\hat{p}_2) = Bias^2(\hat{p}_2) + Var(\hat{p}_2) = 0^2 + \frac{6p(1-p)}{16} = \frac{6p(1-p)}{16}$$

We see that $MSE(\hat{p}_1) < MSE(\hat{p}_2)$, so $\hat{p}_1$ is the better estimator.

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?

Answer:

(a) The first population moment is $E[Y_i] = \lambda$ and the first sample moment is $\bar{y}$. Thus the method of moments estimator is $\hat{\lambda}_{MOM} = \bar{y}$.

(b)

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda}\lambda^{\sum y_i}}{\prod_{i=1}^{n} y_i!}$$
$$\ell(\lambda) = logL(\lambda) = -n\lambda + \sum y_i \log(\lambda) - \sum_{i=1}^{n} \log(y_i!)$$
$$\frac{\partial}{\partial\lambda}\ell(\lambda) = -n + \frac{\sum y_i}{\lambda} \overset{set}{=} 0$$

Setting the above equal to zero and solving for λ, we obtain $\hat{\lambda}_{MLE} = \frac{\sum y_i}{n} = \bar{y}$.
2^{nd} derivative test:
$\frac{\partial^2}{\partial\lambda^2}\ell(\lambda)|_{\lambda=\hat{\lambda}} = \frac{\partial}{\partial\lambda}\left(-n + \frac{\sum y_i}{\lambda}\right)|_{\lambda=\hat{\lambda}} = \frac{-\sum y_i}{\lambda^2}|_{\lambda=\hat{\lambda}} = \frac{-\sum y_i}{\hat{\lambda}^2} < 0$. We have a maximum at $\hat{\lambda}_{MLE}$.

(c) The average of 7, 6, 7, 2, and 4 is 5.2. Thus $\hat{\lambda}_{MOM} = \hat{\lambda}_{MLE} = 5.2$.

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)

Answer:

(a) Using method of moments,

$$\mu_1 = E(X) = \int_{\mathbb{R}} x f(x) dx = \int_0^1 x\theta x^{\theta-1} dx = \frac{\theta}{\theta+1}$$

and $m_1 = \bar{x}$ so that

$$\frac{\hat{\theta}}{\hat{\theta}+1} = \bar{x} \implies \hat{\theta}_{MoM} = \frac{\bar{x}}{1-\bar{x}}.$$

For the given data, $\bar{x} = (0.4 + 0.7 + 0.9)/3 = 2/3$ so that

$$\hat{\theta}_{MoM} = \frac{2/3}{1-2/3} = 2$$

(b) Using MLE,

$$L(\theta) = \prod_{i=1}^{3} \theta t_i^{\theta-1} = \theta^3 (\prod_{i=1}^{3} t_i)^{\theta-1}$$

$$\ell(\theta) = 3\log\theta + (\theta-1)\sum_{i=1}^{3}\log t_i$$

$$\frac{\partial \ell(\theta)}{\partial\theta} = \frac{3}{\theta} + \sum_{i=1}^{3}\log t_i \overset{set}{=} 0$$

$$\hat{\theta}_{MLE} = -3/\sum_{i=1}^{3}\log t_i = -3/-1.378326 = 2.17655$$

2^{nd} derivative test:
$\frac{\partial^2}{\partial\theta^2}\ell(\theta)|_{\theta=\hat{\theta}} = \frac{\partial}{\partial\theta}\left(\frac{3}{\theta} + \sum_{i=1}^{3}\log t_i\right)|_{\theta=\hat{\theta}} = \frac{-3}{\theta^2}|_{\theta=\hat{\theta}} = \frac{-3}{\hat{\theta}^2} < 0$. We have a maximum at $\hat{\theta}_{MLE}$.

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.

Answer:

(a) Since it is the gamma distribution, we have

$$E(X) = \frac{\alpha}{\lambda} = \frac{3}{\lambda}$$

Using method of moments,

$$\mu_1 = E(X) = \frac{3}{\lambda}$$

and $m_1 = \bar{x}$ so that

$$\hat{\lambda} = \frac{3}{\bar{x}}.$$

(b)

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \frac{\lambda^{3n}}{2^n} \left(\prod_{i=1}^{n} x_i^2 \right) e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ell(\lambda) = \log L(\lambda) = 3n \log \lambda - n \log 2 + 2 \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{3n}{\lambda} - \sum_{i=1}^{n} x_i \overset{set}{=} 0$$

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{3n} = \frac{3}{\bar{x}}$$

2^{nd} derivative test:
$\frac{\partial^2}{\partial \lambda^2} \ell(\lambda)|_{\lambda=\hat{\lambda}} = \frac{\partial}{\partial \lambda} \left(\frac{3n}{\lambda} - \sum_{i=1}^{n} x_i \right) |_{\lambda=\hat{\lambda}} = \frac{-3n}{\lambda^2}|_{\lambda=\hat{\lambda}} = \frac{-3n}{\hat{\lambda}^2} < 0$. We have a maximum at $\hat{\lambda}_{MLE}$.