Ahmed AlSaleh, 300192312
MAT 2379, Introduction to Biostatistics
Assignment 2 :

Question 4.4 :

$E(X) = 0.63$

$\sum_{x=0}^{2} x \cdot P(X=x) = 0.63$

$= 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2)$

$= P_1 + 2(P_2) = 0.63$

$P(X \geq 1) = 0.40$

$= P(X=1) + P(X=2) = 0.40$

$P_1 + 2P_2 - (P_1 + P_2) = 0.63 - 0.40$

$P_2 = 0.23 = 23\%$

$P_0 + P_1 + P_2 - (P_1 + P_2) = 1 - 0.40 = 0.6$

$P_0 = 0.6 = 60\%$

$P_1 = 1 - P_0 - P_2 = 0.17 = 17\%$

Question 4.8

a) dbinom (6,10,0.3) gives $P(X=6)$ where $X \sim$ Binomial (10,0.3) : true

b) pbinom (20,7,0.5) gives $P(X \leq 7)$ where $X \sim$ Binomial (20,0.5) : False

c) pbinom (5,13,0.8) - pbinom (2,13,0.8) gives $P(2 \leq X \leq 5)$ where $X \sim$ Binomial (13,0.8) : false

d) pbinom (20,12,0.8) - pbinom (2,8,0.8) gives $P(9 \leq X \leq 12)$ where $X \sim$ Binomial (20,0.8) : false

e) dbinom (7,9,0.8) + pbinom (6,9,0.8) gives $P(X \leq 7)$ where $X \sim$ Binomial (9,0.8) : true

f) pbinom (2,6,0.4) + pbinom (3,6,0.4) + pbinom (4,6,0.4) gives $P(2 \leq X \leq 4)$ where $X \sim$ Binomial (6,0.4) : false

Question 5.10

a) $P(X>2) = P(\ln X > \ln 2) = P\left(\frac{\ln X - E(\ln x)}{\sqrt{Var(\ln x)}} > \frac{\ln 2 - E(\ln x)}{\sqrt{Var(\ln x)}}\right)$

$Z = \frac{\ln X - E(\ln x)}{\sqrt{Var(\ln x)}} \sim N(0,1)$

→ other page for continued calculation

5.10 a) continued

$$= P\left(Z > \frac{\ln 2 - 0.68}{0.22}\right)$$

$$= P\left(Z > \frac{0.693 - 0.68}{0.22}\right)$$

$$= P(Z > 0.0591)$$

$$\boxed{P(X > 2) = 0.4764}$$

$$P(X \le 1) = P(\ln X \le \ln 1)$$
$$= P(\ln X \le 0)$$

$$= P\left(\frac{\ln X - E(\ln X)}{\sqrt{Var(\ln X)}} \le \frac{0 - E(\ln X)}{\sqrt{Var(\ln X)}}\right)$$

$$= P\left(Z \le \frac{0 - 0.68}{0.22}\right)$$

$$= P(Z \le -3.091)$$

$$= 0.001$$

$$\boxed{P(X \le 1) = 0.001}$$

b) $P(X \le X_0) = 0.5$

$$P(X \le X_0) = P(\ln X \le \ln X_0)$$

$$= P\left(\frac{\ln X - E(\ln X)}{\sqrt{Var(\ln X)}} \le \frac{\ln X_0 - E(\ln X_0)}{\sqrt{Var(\ln X_0)}}\right)$$

$$= P\left(Z \le \frac{\ln X_0 - 0.68}{0.22}\right)$$

$$\rightarrow \frac{\ln X_0 - 0.68}{0.22} = 0$$

$$\ln X_0 = 0.68$$

$$X_0 = e^{0.68}$$

$$\boxed{X_0 = 1.974}$$

Question 7.4

a) mean $= \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $n = 18$ and $\Sigma x_i = 35.9$

$$\bar{x} = \frac{35.9}{18} = 1.994$$

$\therefore$ mean is 1.994

Standard deviation:

$$s^2 = \frac{1}{17}\left[(1.1-1.994)^2 + \ldots + (3.1-1.994)^2\right]$$

$$s^2 = 0.437$$

$$s = 0.6611$$

$\therefore$ standard devation is 0.6611

b) median $= \frac{2+2.1}{2} = 2.05$

Q_1 = median for first 9 values $= 1.5$

Q_3 = median for last 9 values $= 2.4$

$IQR = Q_3 - Q_1 = 2.4 - 1.5 = 0.9$

$\therefore$ IQR is 0.9

c) There are no outliers in this data set.

upper fence $= Q_3 + 1.5(IQR) = 2.4 + 1.5(0.9)$
$= 3.75$

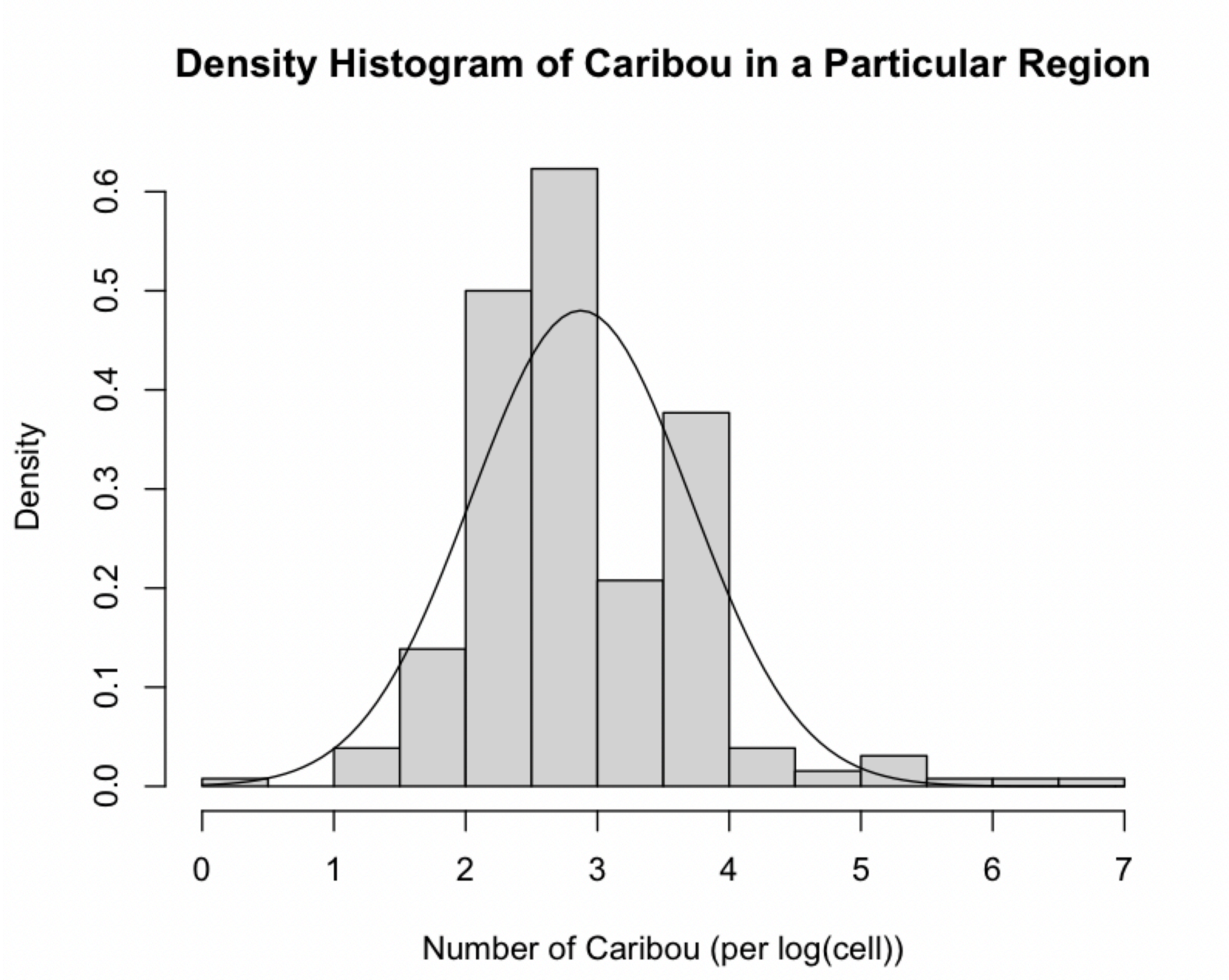
lower fence $= Q_1 - 1.5(IQR) = 1.5 - 1.5(0.9)$
$= 0.15$

There are no values greater than 3.75 or less than 0.15, therefore, there are no outliers.

R Problem

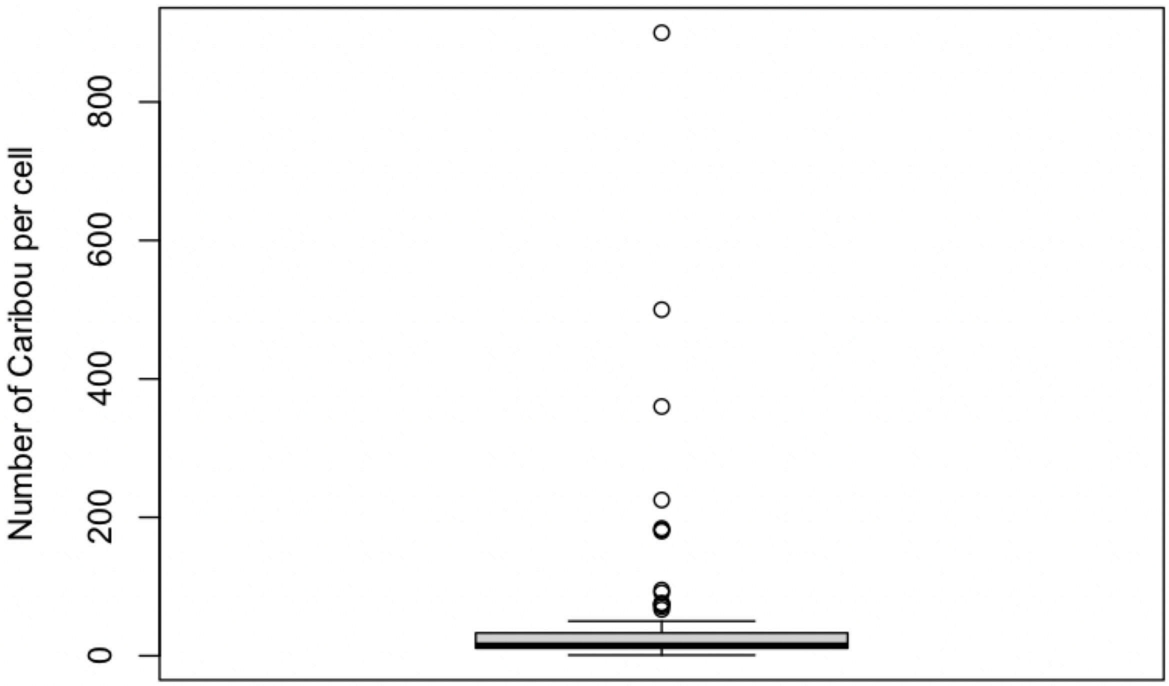
Question 14.14

a)



The shape of the distribution represents a normal distribution.

b) I recommend using the median to describe the center of distribution since it requires finding the middle value in the distribution which is the median. The mean finds the average value in a data which is not necessarily the middle value in a data set.
c) The center of distribution is 2.71 which shows that the middle value of the density histogram is 2.71 caribous per log(cell).
d)



This is a narrow boxplot since the number of caribous per cell are concentrated between 10-100. Thus, all the other values outside of the boxplot fences shows the values that are outliers. In this boxplot, there are multiple outliers outside of the boxplot's upper fence.