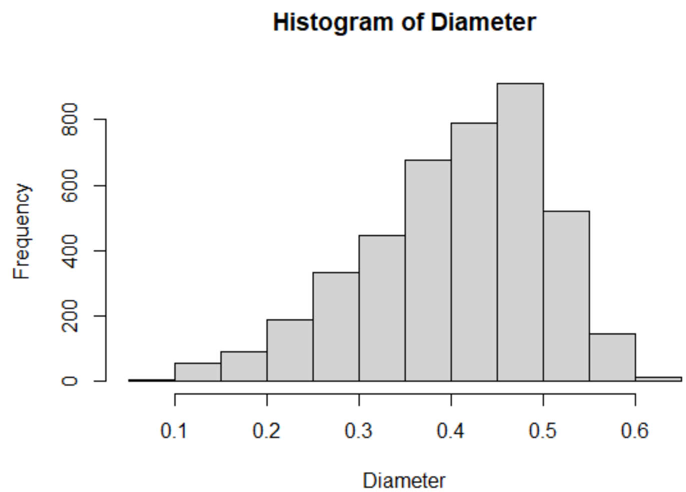
Written Assignment-1 Solutions

Question 1

a) (*3 marks*: 1 mark for the graph, 1 mark for giving suitable title and naming axes, 1 mark for the description.)

The histogram is unimodal and left-skewed.



b) (*2 marks*: 1 mark for five number summary, 1 mark for the standard deviation.)

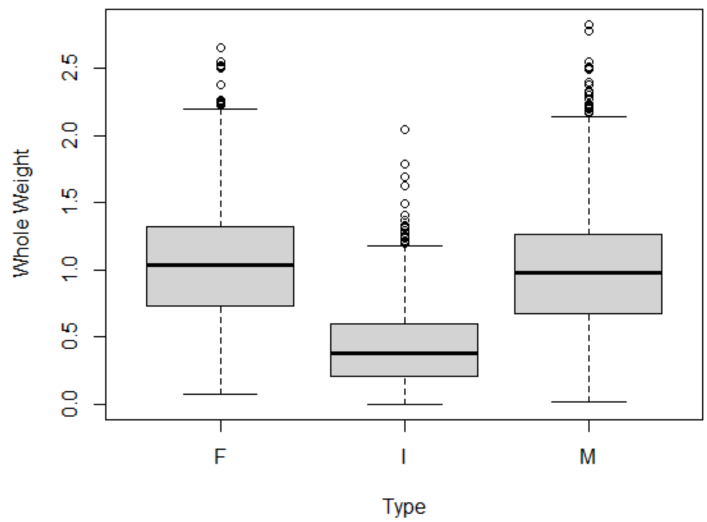
The five number summary gives the following results when we use the R code summary(). It will produce slightly different results if you use fivenum() in R.

- Min.: 0.0550
- 1st Qu.: 0.3500
- Median: 0.4250
- 3rd Qu.: 0.4800
- Max.: 0.6500

The standard deviation of diameter is 0.0992.

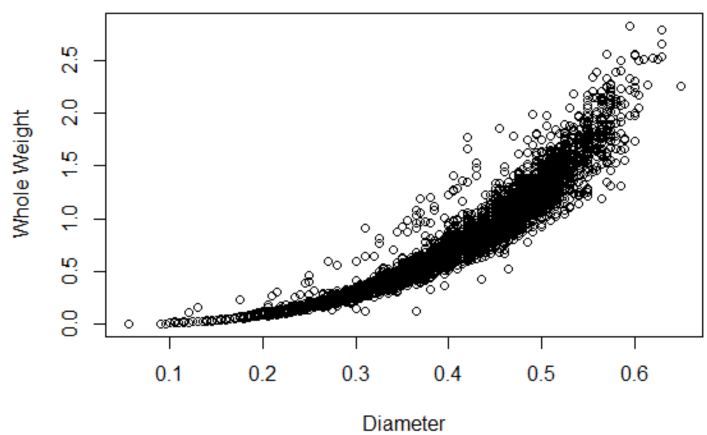
c) *(3 marks* :1 mark for the graph, 1 mark for naming axes, 1 mark for the comment.)

The plot does not suggest a difference in the whole weight between male and female abalones as the boxplots for male and female are nearly the same.



d) *(2 marks* : 1 mark for the plot, 1 mark for the description.)

There seems to be a positive association between "diameter" and "whole weight"; however, this relationship does not appear to be linear.



Question 2

a) *(3 marks* : 1 for defining events, 1.5 for the method, 0.5 for the answer)

Let D_A denote the event that detector A locates a whale and D_B the event that detector B locates a whale. The probability that a whale is located is given by

$$\begin{aligned}
P(D_A \cup D_B) &= P(D_A) + P(D_B) - P(D_A \cap D_B) \\
&= P(D_A) + P(D_B) - P(D_A)P(D_B) \quad \text{independence of } D_A \text{ and } D_B \\
&= 0.9 + 0.7 - 0.9 \times 0.7 \\
&= 0.97
\end{aligned}$$

b) *(3 marks* : 2 for the method, 1 for the answer)

the probability that a whale is located by only one detector is

$$\begin{aligned}
P((D_A \cap D_B^c) \cup (D_A^c \cap D_B)) &= P(D_A \cap D_B^c) + P(D_A^c \cap D_B) - P((D_A \cap D_B^c) \cap (D_A^c \cap D_B)) \\
&= P(D_A \cap D_B^c) + P(D_A^c \cap D_B) + 0 \\
&= P(D_A)P(D_B^c) + P(D_A^c)P(D_B) \\
&\quad \text{; since } D_A \text{ and } D_B \text{ are independent,} \\
&\quad \text{(i) } D_A \text{ and } D_B^c \text{ and (ii) } D_A^c \text{ and } D_B \text{ are also independent} \\
&= 0.9 \times (1 - 0.7) + (1 - 0.9) \times 0.7 \\
&= 0.34
\end{aligned}$$

or $P(D_A \cup D_B) - P(D_A \cap D_B) = 0.97 - 0.9 \times 0.7 = 0.34$

c) *(4 marks* : 0.5 for defining event, 2.5 for the mehod, 1 for the answer)

Let L denote the event that a whale's speed is accurately recorded. The probability that a whale's speed is accurately recorded is

$$\begin{aligned} P(L) &= P(L|D_A \cap D_B)P(D_A \cap D_B) + P(L|D_A \cap D_B^c)P(D_A \cap D_B^c) + P(L|D_A^c \cap D_B)P(D_A^c \cap D_B) \\ &= P(L|D_A \cap D_B)P(D_A)P(D_B) + P(L|D_A \cap D_B^c)P(D_A)P(D_B^c) \\ &\qquad + P(L|D_A^c \cap D_B)P(D_A^c)P(D_B) \quad \text{; by independence} \\ &= 1 \times 0.9 \times 0.7 + 0.6 \times 0.9 \times (1 - 0.7) + 0.8 \times (1 - 0.9) \times 0.7 \\ &= 0.63 + 0.162 + 0.056 \\ &= 0.848 \end{aligned}$$

Question 3

a) *(4 marks* : 2 marks for the pmf,1 mark for expected value, 1 mark for variance)

The possible value of X are 0, 1, 2, 3. Due to the independence of stopping, the pmf of X is given as:

$$P(X = x) = \begin{cases} p^0(1-p)^3 = 0.027, & x = 0 \\ 3 \times p^1(1-p)^2 = 0.189, & x = 1 \\ 3 \times p^2(1-p)^1 = 0.441, & x = 2 \\ p^3(1-p)^0 = 0.343, & x = 3 \end{cases}$$

By definition, we have

$$E(X) = \Sigma_{x=0}^{3} x \times P(X = x) = 0 + 0.189 + 2 \times 0.441 + 3 \times 0.343 = 2.1$$

$$\begin{aligned} Var(X) &= \Sigma_{x=0}^{3} (x - E(X))^2 \times P(X = x) \\ &= (0 - 2.1)^2 \times 0.027 + (1 - 2.1)^2 \times 0.189 + \\ &\qquad (2 - 2.1)^2 \times 0.441 + (3 - 2.1)^2 \times 0.343 \\ &= 0.63 \end{aligned}$$

b) *(4 marks* : 2 marks for the pmf,1 mark for expected value, 1 mark for variance)

$Stopat1$	$Stopat2$	$Stopat3$	Probability	Total commute time (y)
No	No	No	0.027	10
Yes	No	No	0.063	11
No	Yes	No	0.063	12
No	No	Yes	0.063	13
Yes	Yes	No	0.147	13
Yes	No	Yes	0.147	14
No	Yes	Yes	0.147	15
Yes	Yes	Yes	0.343	16

y	$P(Y=y)$
10	0.027
11	0.063
12	0.063
13	0.210
14	0.147
15	0.147
16	0.343

$$\begin{aligned} E(Y) =& \sum_y yP(Y=y) \\ =&(10 \times 0.027) + (11 \times 0.063) + (12 \times 0.063) + (13 \times 0.210) + \\ & (14 \times 0.147) + (15 \times 0.147) + (16 \times 0.343) = \quad 14.2 \end{aligned}$$

$$\begin{aligned} E(Y^2) =& \sum_y y^2P(Y=y) \\ =&(10^2 \times 0.027) + (11^2 \times 0.063) + (12^2 \times 0.063) + (13^2 \times 0.210) + \\ & (14^2 \times 0.147) + (15^2 \times 0.147) + (16^2 \times 0.343) = \quad 204.58 \end{aligned}$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = 204.58 - (14.2)^2 = 2.94$$

Question 4

a) *(3 marks* : 0.5 for correct notations, defining random variables, 1 for mean, 1 for variance, 0.5 identifying normal distribution for total)

Let X_1, X_2 and X_3 denotes the scores of a student in Assignment 1, Assignment 2, and Assignment 3 respectively and T is the total assignment grade. Then,

$$\begin{aligned} X_1 &\sim N(65, 15^2) \\ X_2 &\sim N(55, 20^2) \\ X_3 &\sim N(60, 17^2) \\ T &= 0.25X_1 + 0.35X_2 + 0.40X_3 \end{aligned}$$

The expectation (or mean) of Total assignment grade is

$$\begin{gathered} E(T) = E(0.25X_1 + 0.35X_2 + 0.40X_3) \\ = (0.25 \times 65) + (0.35 \times 55) + (0.40 \times 60) \\ = 59.5 \end{gathered}$$

The variance of the Total assignment grade is

$$\begin{gathered} Var(T) = Var(0.25X_1 + 0.35X_2 + 0.40X_3) \\ = (0.25^2 \times 15^2) + (0.35^2 \times 20^2) + (0.40^2 \times 17^2) \\ = 109.30 \end{gathered}$$

Since X_1, X_2 and X_3 are independently and normally distributed, their linear combination, total grade also follows normal distribution with mean 59.5 and standard deviation $\sqrt{109.30}$. That is the distribution of total assignment grade of a randomly selected student is, $T \sim N(59.5, 109.30)$.

b) *(3 marks* : 2 for the method, 1 for the answers)

From part a) the distribution of total assignment grade, T is $T \sim N(59.5, 109.30)$. We require c (the number of total grade) such that,

$$P(T < c) = 0.87$$
$$P\left(\frac{T - 59.5}{\sqrt{109.3}} < \frac{c - 59.5}{\sqrt{109.3}}\right) = 0.87$$
$$P\left(Z < \frac{c - 59.5}{\sqrt{109.3}}\right) = 0.87$$

This implies,

$$\frac{c - 59.5}{\sqrt{109.30}} = 1.13$$
$$c = 59.5 + (1.13 \times \sqrt{109.30}) = 71.314$$

c) *(4 marks*: 1 for mean of $\overline{T}$, 1 for variance of $\overline{T}$, 0.5 for identifying $\overline{T}$ follows a normal distribution, 1 for method for finding the required probability, 0.5 for the answer)

Let T_i denote the total grade in assignments of student i for i = 1, 2,...,40. The average of their total grades is denoted by

$$\overline{T} = \frac{1}{40}\ (T_1 + T_2 + ... + T_{40})$$

Thus, the average and variance of the total grade is

$$E(\overline{T}) = E\left[\frac{1}{40}(T_1 + T_2 + ... + T_{50})\right]$$
$$= 59.5$$
$$Var(\overline{T}) = Var\left[\frac{1}{40}(T_1 + T_2 + ... + T_{50})\right]$$
$$= \frac{109.30}{40} = 2.7325$$

In other way, we know that if $X \sim N(\mu, \sigma^2)$ then the distribution of mean, $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ Hence, the distribution of $\overline{T}$ is $\overline{T} \sim N(59.5, 2.7325)$

The probability that averages of their total grade is between 55-60 is

$$\begin{aligned}
P(55 < \overline{T} < 60) &= P\left(\frac{55-59.5}{\sqrt{2.7325}}\right) < Z < \left(\frac{60-59.5}{\sqrt{2.7325}}\right) \\
&= P(-2.72 < Z < 0.30) \\
&= P(Z < 0.30) - P(Z < -2.72) \\
&= P(Z < 0.30) - (1 - P(Z < 2.72)) \\
&= 0.6179 - 1 + 0.9967 \\
&= 0.6146
\end{aligned}$$

Hence, The probability of average assignment grade between 55 and 60 of 40 students is 0.6146 or 61.46%.