MAT2379, Introduction to biostatistics

Solutions to Assignment 2
Total = 100 marks

Part (I) (50 marks)

Problem 6.1 (10 marks)
(a) $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) = 0.2 + 0.1 + 0.1 = 0.4$.
(b) $P(X \leq 3) = P(X = 2) + P(X = 3) = 0.05 + 0.25 = 0.3$.
(c) $E(X) = 2(0.05) + 3(0.25) + 4(0.3) + 5(0.2) + 6(0.1) + 7(0.1) = 4.35$.
(d) $\mathrm{Var}(X) = 2^2(0.05) + 3^2(0.25) + 4^2(0.3) + 5^2(0.2) + 6^2(0.1) + 7^2(0.1) - (4.35)^2 = 1.8275$

Problem 6.5 (10 marks)
Let X be the number of patients that report a significant reduction in their pain among 20 patients. Assuming that the medication has no effect, then X has a binomial distribution with $n = 20$ and $p = 0.5$. The probability that at least 17 of 20 patients would report significant reduction in their pain is

$$\begin{aligned} P(X \geq 17) = & \binom{20}{17}(0.5)^{17}(0.5)^3 + \binom{20}{18}(0.5)^{18}(0.5)^2 \\ & + \binom{20}{19}(0.5)^{19}(0.5)^1 + \binom{20}{20}(0.5)^{20}(0.5)^0 \\ = & \, 0.00129. \end{aligned}$$

Problem 7.8 (15 marks) Let X be the weight of an adult male grizzly from the Alaska Penisula region. Then, $X \sim N(357, 21^2)$.
(a)

$$\begin{aligned} P(X > 420) &= 1 - P(Z < (420 - 357)/21) \\ &= 1 - \Phi(3.0) = 1 - 0.9987 = 0.0013 \end{aligned}$$

(b)

$$\begin{aligned} P(X > 300) &= 1 - P(Z < (300 - 357)/21) \\ &= 1 - \Phi(-2.71) = 1 - 0.0034 = 0.9966 \end{aligned}$$

(c) We want x_0 such that $P(Z < (x_0 - 357)/21) = 0.05$. Thus,

$$(x_0 - 357)/21 = -1.645 \;\Rightarrow\; x_0 = 357 - 21(1.645) = 322.455.$$

(d) We want x_0 such that $P(Z < (x_0 - 357)/21) = 0.75$. Thus,

$$(x_0 - 357)/21 = 0.675 \;\Rightarrow\; x_0 = 357 + 21(0.675) = 371.175.$$

(e) Let Y be the number of bears among the 6 selected bears that weigh less than 300 kg. Y has a binomial distribution with $n = 6$ and $p = 1 - 0.9966 = 0.0034$. We want

$$\begin{aligned} P(Y \leq 2) = & \binom{6}{0}(0.0034)^0(0.9966)^6 + \binom{6}{1}(0.0034)^1(0.9966)^5 \\ & + \binom{6}{2}(0.0034)^2(0.9966)^4 \\ = & \ 0.999999. \end{aligned}$$

Problem 9.4 (15 marks)

(a) We arrange the data in increasing order:

$$y_1 = y_2 = 4.2 \quad y_3 = 4.3 \quad y_4 = 4.4 \quad y_5 = 4.5 \quad y_6 = y_7 = 4.6 \quad y_8 = 4.7$$

$$y_9 = y_{10} = 4.9 \quad y_{11} = 5.1 \quad y_{12} = 5.6$$

Since $n = 12$ is even, the median is $Q_2 = (y_6 + y_7)/2 = 4.6$. The first quartile is

$$Q_1 = (0.75)y_3 + (0.25)y_4 = (0.75)(4.3) + (0.25)(4.4) = 4.325.$$

The third quartile is

$$Q_3 = (0.25)y_9 + (0.75)y_{10} = (0.25)(4.9) + (0.75)(4.9) = 4.9.$$

(b) $IQR = Q_3 - Q_1 = 4.9 - 4.325 = 0.575$. The fences are located at $Q_1 - 1.5(IQR) = 4.325 - 0.8625 = 3.4625$ and $Q_3 + 1.5(IQR) = 4.9 + 0.8625 = 5.7625$. There is no outlier.

(c) The sample mean is

$$\bar{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = 4.6667$$

The sample standard deviation is

$$s = \sqrt{\frac{1}{11}\sum_{i=1}^{12}(y_i - \bar{y})^2} = 0.4097$$

Part (II) (50 marks)

1. (5 marks) a) $P(X \geq 3) = 1 - P(X \leq 2) = 0.747$; b) $P(4 \leq X < 8) = P(X \leq 7) - P(X \leq 3) = 0.498$.

```
> 1-pbinom(2, size=12, prob=0.3)
[1] 0.7471847
> pbinom(7, size=12, prob=0.3) - pbinom(3, size=12, prob=0.3)
[1] 0.4979949
```

2. (15 marks) a) $P(X \le 9) = 0.655$; b) $P(1.8 < X < 7.8) = P(X < 7.8) - P(X < 1.8) = 0.462$.

```
> pnorm(9, mean=8, sd=2.5)
[1] 0.6554217
> pnorm(7.8, mean=8, sd=2.5) - pnorm(1.8, mean=8, sd=2.5)
[1] 0.4615495
```

c) It is equivalent to find a t such that $P(X < t) = 0.75$. Using `qnorm` function, we find $t = 9.686$.

```
> qnorm(0.75, mean=8, sd=2.5)
[1] 9.686224
```

d) Since $P(|X - 8| < t) = 0.8$, we have $P(8 - t < X < 8 + t) = 0.8$. Note that the PDF of X is symmetric about 8; this implies that

$$P(X < 8 - t) = P(X > 8 + t) = (1 - P(8 - t < X < 8 + t))/2 = (1 - 0.8)/2 = 0.1.$$

Using R, we find $P(X < 4.796) = 0.1$, which implies that $8 - t = 4.796$. So, we have

$$t = 8 - 4.796 = 3.204$$

```
> qnorm(0.1, mean=8, sd=2.5)
[1] 4.796121
```

3. (20 marks) a) In `R`, we assign the Gentoo penguins to variable x, and the Chinstrap penguins to variable y.

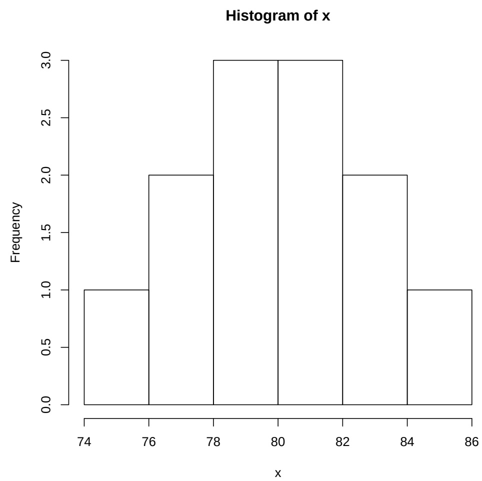
```
> x <- c(76.9, 84.2, 78.0, 78.1, 80.8, 78.2, 81.8, 82.1, 80.8, 75.7, 83.5, 78.6)
> y <- c(76.2, 76.3, 75.5, 75.9, 76.5, 73.1, 73.8, 76.3, 74.9, 76.3, 76.8, 75.7)
```

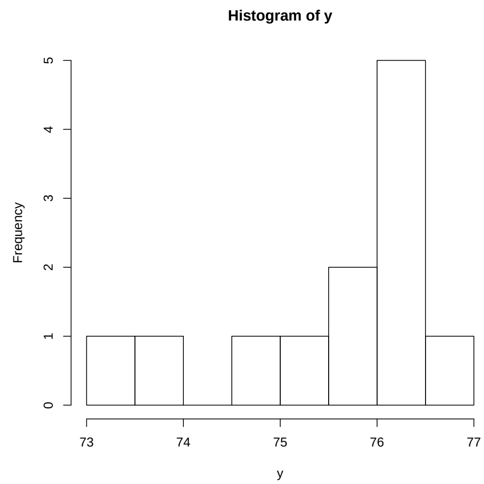
The five-number summaries of x and y are given as follows.

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  75.70   78.07   79.70   79.89   81.88   84.20 
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  73.10   75.35   76.05   75.61   76.30   76.80 
```

b) The histograms of x and y are as follows.

```
> hist(x, breaks=6)
> hist(y, breaks=6)
```

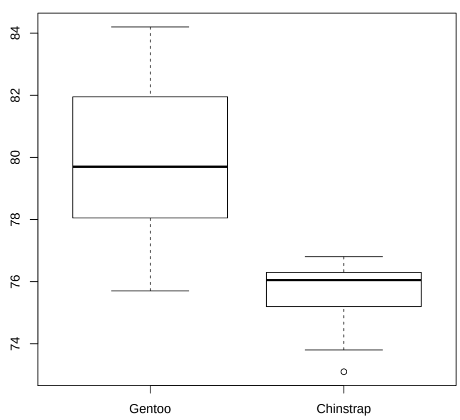




The distribution of x (i.e. body length of Gentoo penguins) is symmetric. The distribution of y (i.e. body length of Chinstrap penguins) is skewed to the left.

c) The side-by-side boxplots of x and y are as follows.

```
> boxplot(x, y, names=c("Gentoo", "Chinstrap"))
```



We observe that the body lengths of Gentoo penguins have a higher central tendency than that of Chinstrap penguins. The length distribution of Gentoo penguins is nearly symmetric, while the length distribution of Chinstrap penguins is skewed to the left. There is more variability in the lengths of Gentoo penguins; this is indicated by the (relatively) large IQR. There is one extreme value (outlier) in the Chinstrap penguin group.

4 (10 marks)

Note that Fahrenheit and Kelvin have a positive linear relationship:

$$F = K \times 9/5 - 459.67.$$

Thus, the descriptive statistics for the temperatures in degrees Fahrenheit are given as follows.

- The minimum temperature in degrees Fahrenheit is

$$\min(F) = \min(K) \times 9/5 - 459.67 = 309.1 \times 9/5 - 459.67 = 96.71.$$

- The maximum temperature in degrees Fahrenheit is

$$\max(F) = \max(K) \times 9/5 - 459.67 = 312.3 \times 9/5 - 459.67 = 102.47.$$

- The first quartile in degrees Fahrenheit is

$$Q_1(F) = Q_1(K) \times 9/5 - 459.67 = 310.2 \times 9/5 - 459.67 = 98.69.$$

- The median in degrees Fahrenheit is

$$Q_2(F) = Q_2(K) \times 9/5 - 459.67 = 310.7 \times 9/5 - 459.67 = 99.59.$$

- The third quartile in degrees Fahrenheit is

$$Q_3(F) = Q_3(K) \times 9/5 - 459.67 = 311.5 \times 9/5 - 459.67 = 101.03.$$

- The IQR in degrees Fahrenheit is

$$IQR_F = Q_3(F) - Q_1(F) = 101.03 - 98.69 = 2.34.$$

- The sample mean in degrees Fahrenheit is

$$\overline{F} = \overline{K} \times 9/5 - 459.67 = 310.8 \times 9/5 - 459.67 = 99.77.$$

- The sample standard deviation in degrees Fahrenheit is

$$s_F = (9/5)s_K = (9/5) \times (1.04) \approx 1.872.$$