**MAT 2379, Introduction to biostatistics**

**Assignment 2 - solutions**

**Part I:**

[6] **Question 1:** Problem 4.8 from the textbook.

(a) true;
(b) false; the probability given by the `R` command is $P(X \leq 20)$ where $X \sim \text{Binomial}(7, 0.5)$.
(c) false: the probability given by the `R` command is $P(X \leq 5) - P(X \leq 2) = P(3 \leq X \leq 5)$;
(d) false; the probability given by the `R` command is $P(X \leq 20) - P(Y \leq 2)$, where $X \sim \text{Binomial}(12, 0.8)$ and $Y \sim \text{Binomial}(8, 0.8)$
(e) true;
(f) false: the probability given by the `R` command is $P(X \leq 2) + P(X \leq 3) + P(X \leq 4)$ where $X \sim \text{Binomial}(6, 0.4)$.

[5] **Question 2:** Problem 4.10 from the textbook.

(a) Since the total probability is equal to 1, then

$$1 = \frac{c}{18} + \frac{c}{18} + \frac{2c}{9} + \frac{c}{6} + \frac{c}{3} + \frac{c}{3} + \frac{c}{6} = \frac{4\,c}{3}.$$

Thus, $c = 3/4$.

(b) The probability mass function is

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $f(x)$ | 1/24 | 1/24 | 1/6 | 1/8 | 1/4 | 1/4 | 1/8 |

The expected number of properly stained cells is

$$E(X) = \sum x\, P(X = x) = 0\,(1/24) + 1\,(1/24) + \cdots + 6\,(1/8) = 3.75.$$

(c)(i) $P(|X - 2| \leq 1) = P(X = 1) + P(X = 2) + P(X = 3) = 1/3 = 0.3333.$

(ii) Using $\mu = E(X) = 3.75$, we compute

$$\begin{aligned} \sigma &= \sqrt{V[X]} = \sqrt{\left(\sum x^2\, P(X = x)\right) - \mu^2} \\ &= \sqrt{0^2\,(1/24) + 1^2\,(1/24) + \cdots + 6^2\,(1/8) - (3.75)^2} = 1.5877. \end{aligned}$$

Thus,

$$P(X > \mu + \sigma) = P(X > 5.3377) = P(X = 6) = \frac{1}{8} = 0.125.$$

[4] **Question 3:** Problem 5.2 from the textbook.

Let $X$ be the temperature on January 15. Let us compute

$$\begin{aligned} P(-20 < X < -8) &= P\left(\frac{-20-(-12)}{5} < \frac{X-(-12)}{5} < \frac{-8-(-12)}{5}\right) \\ &= P(-1.6 < Z < 0.8) = \Phi(0.8) - \Phi(-1.6) = 0.7881 - 0.0548 = 0.7333 \end{aligned}$$

In a sequence of 5 years, let $Y$ be the number of years that the grapes can be picked on January 15. $Y$ has a binomial distribution with $n = 5$ trials and $p = 0.7333$ probability of success. We want the following probability

$$P(Y = 5) = \binom{5}{0}(0.7333)^5(1-0.7333)^0 = (0.7333)^5 = 0.2120$$

[3] **Question 4:** Problem 5.8 from the textbook.

Let $X$ be the weight of a randomly chosen watermelon. We have to find $x_0$ such that

$$0.25 = P(X < x_0) = P\left(\frac{X-1.5}{0.7} < \frac{x_0-1.5}{0.7}\right) = P\left(Z < \frac{x_0-1.5}{0.7}\right)$$

In the table for the cdf of the standard normal, we search a value $z_0$ such that $\Phi(z_0) = 0.25$. We find $P(Z < -0.67) = 0.2514$ and $P(Z < -0.68) = 0.2483$. We take the midpoint between -0.67 and -0.68, i.e. $z_0 = -0.675$. We conclude that:

$$\frac{x_0-1.5}{0.7} = -0.675$$

Hence

$$x_0 = 1.5 - (0.7)(0.675) = 1.5 - 0.4725 = 1.0275 \approx 1.03$$

[3] **Question 5:** Problem 6.12 from the textbook.

Let $X$ be the number of non-infected geese among 5 geese. $X$ has a binomial distribution with $n = 5$ and 0.95.
(a) $E(X) = n\,p = 5(0.95) = 4.75$
(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - (0.05)^5 = 1.0000$

**Question 6:**

(a) Let $p_m$ and $p_f$ be the proportion of males and the proportion of females, respectively, in the population. We have $p_m/p_f = 0.83$ and $p_m + p_f = 1$. Isolating $p_m$ from the first equation and substituting it in the second equation gives

$$p_f = 1 - p_m = 1 - 0.83\,p_f \quad \Rightarrow \quad p_f = 1/(1.83) = 0.5464.$$

So 54.64% of the individuals in this population are female.

(b) Let $M$, $F$, and $D$ be the events that the individual is male, the individual is female, and the individual suffers from depression. We know that

$$P(D|M) = 0.08, \quad P(D|F) = 0.16, \quad \text{and} \quad P(F) = 0.5464.$$

We want

$$\begin{aligned} P(D) &= P(D|M)P(M) + P(D|F)P(F) \\ &= (0.08)(1 - 0.5464) + (0.16)(.5464) = 0.1237. \end{aligned}$$

(c) We know that $P(D|M) = 0.08$ and $P(D) = 0.1237$. This means that $P(D|M) \neq P(D)$. Thus, the events $M$ and $D$ are dependent. So no, depression is **not independent** of gender.

(d) We want

$$P(F|D) = \frac{P(F \cap D)}{P(D)} = \frac{P(D|F)P(F)}{P(D)} = \frac{(0.16)(.5464)}{0.1237} = 0.7067.$$

**Part (II)**

[4] **Question 7:**

**a)** $P(X \leq 90) = F_X(90) = 0.9981$
**b)** $P(X \geq 84) = 1 - P(X \leq 83) = 1 - F_X(83) = 0.0591$;
**c)** $P(X < 80) = P(X \leq 79) = F_X(79) = 0.7937$
**d)** $P(60 \leq X \leq 80) = P(X \leq 80) - P(X \leq 59) = F_X(80) - F_X(59) = 0.8398$.

**With `R`:**

```
> pbinom(90,125,.6)
[1] 0.9980733
> 1-pbinom(83,125,.6)
[1] 0.0591082
> pbinom(79,125,.6)
[1] 0.793727
> pbinom(80,125,.6)-pbinom(59,125,.6)
[1] 0.8398265
```

**Question 8:**

**a)** $P(X < 30) = F_X(30) = 0.7666$
**b)** $P(30 < X < 50) = F_X(50) - F_X(30) = 0.8469$
**c)** $P(X > 45) = 1 - F_X(45) = 0.2375$
**d)** $x_0 = F_X^{-1}(0.95) = 51.514$.
**e)** We want a value such that $0.8 = P(X \leq x_0)$. Thus, $x_0 = F_X^{-1}(0.8) = 45.89135$
**With `R`:**

```
> pnorm(30,40,7)
[1] 0.07656373
> pnorm(50,40,7)-pnorm(30,40,7)
[1] 0.8468725
> 1-pnorm(45,40,7)
[1] 0.2375253
> qnorm(0.95,40,7)
[1] 51.51398
> qnorm(0.8,40,7)
[1] 45.89135
```

[ /**25**]

**Marking Scheme**

**Question 1:** 1 pt per part for properly identifying if it is true or false.

**Question 2:** a) 1/2 pt for noting that total prob is 1; b) 1/2 point for $c = 3/4$. b) 1 pt for properly computing $E[X]$
c) (i) 1/2 pt for noting that we want $P(X = 1) + P(X = 2) + P(X = 3)$; 1/2 pt for correct final answer
c) (ii) 1/2 pt for noting $\mu = E[X]$, 1/2 pt for properly computing $\sigma$, 1/2 for replacing the value of $\mu$ and $\sigma$ to get $P(X > P(X > 5.3377)$. c) (ii) 1/2 pt for noting for the correct final answer.

**Question 3:** 1pt for getting the standardized values $z = 0.8$ and $z = -1.6$; 1pt for properly using the table and giving $\Phi(0.8) - \Phi(-1.6) = 0.7881 - 0.0548$. 1 pt for defining a binomial random variable. 1 pt for properly computing the probability with the binomial.

**Question 4:** 1 pt for getting the quantile $z = -0.675$. 1 pt for equating $(x - 1.5)/0.7 = -0.675$. 1 pt for isolating for $x$.

**Question 5:** 1 pt for defining a binomial random variable, i.e. then indicate what $X$ represents. 1pt for (a), 1 pt for (b).

**Question 7:** For each part give 1/2 for rewriting the prob in terms of the cdf. For each part give 1/2 for providing the `R` output with the correct final answer.