THE CHINESE UNIVERSITY OF HONG KONG

Department of Statistics

Subject Code:	STAT1011	Course Title:	Introduction to Statistics
Session:	Semester 1, 2018/2019, Midterm Examination		
Date:	24 October 2018	Time:	15:00pm -17:00pm
Time Allowed:	2 Hours		

This question paper has 2 pages.

Instructions to Candidates:

1. Attempt **ALL** questions
2. This paper has **7** questions.
3. Give **full details** of your working to the questions in the answer book.
4. Round the answers to the 4^{th} decimal place, whenever necessary.
5. A standard normal table is attached.

Subject Examiner: Professor Yuanyuan LIN

DO NOT TURN OVER THE PAGE UNTIL YOU ARE TOLD TO DO SO

1. (**15 marks**) The data values below represent the prices of 10 actively traded stocks from the Hong Kong Stock Exchange (in dollars)

$$6 \quad 2 \quad 28 \quad 80 \quad 26 \quad 19 \quad 23 \quad 14 \quad 37 \quad 47$$

   (a) (**5 marks**) Find the mean, median, range, and standard deviation (using sample standard deviation formula);

   (b) (**5 marks**) Find the first quartile and the third quartile, and the interquartile range;

   (c) (**5 marks**) Identify any outliers (show your steps clearly).

2. (**20 marks**) Find the probability that

   (a) (**6 marks**) at least one ace turns up in four rolls of a fair die;

   (b) (**6 marks**) at least one double-ace turns up in 25 rolls of two fair dice.

   (c) (**7 marks**) getting exactly 4 sixes in a roll of 7 fair dice.

3. (**10 marks**) Among 65-year-old college professors, 95% are nonsmoker and 5% are smokers. The probability of a nonsmoker dying in a year is 0.005, and the probability for smoker dying in a year is 0.05. Given that one of the college professors died last year, what is the conditional probability that the professor was a smoker?

4. (**15 marks**)

   (a) (**8 marks**) The average income in city A is \$14000 and the standard deviation is \$1500. Use Chebyshev's Theorem to evaluate the probability that a randomly selected individual's income is above \$17100.

   (b) (**7 marks**) At a particular university in city B, illness is stated as the reason for 75% of class absents. Find the probability that among the next 5 class absents at most 3 resulted from illness.

5. (**15 marks**) A PARKnSHOP manager knows that, on average, 100 people enter his store per hour. Find the probability that

   (a) (**7 marks**) in a given 3-minute period, nobody enters the store;

   (b) (**8 marks**) in a given 3-minute period, more than 5 people enter the store.

6. (**15 marks**) Suppose that the heights of all people in city H are normally distributed with a mean of 1.65m and a standard deviation of 0.05m.

   (a) (**7 marks**) What is the probability that a randomly selected person in city H will be between 1.6m and 1.7m?

   (b) (**8 marks**) If 16 people are selected randomly in city H, what is the chance that their average heights are below 1.6m?

7. (**10 marks**) Suppose that a fair coin is tossed independently for 20 times. Given that there were 12 heads in the 20 independent tosses, calculate

   (a) (**5 marks**) the chance that first toss landed head;

   (b) (**5 marks**) the chance that the last two tosses landed heads.

*** End ***

1. (15 marks)

(a) (5 marks)

mean = 28.2, median = $\dfrac{23+26}{2}$ = 24.5, range = 80 - 2 = 78,

standard deviation = $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} = 22.63625.$

(b) (5 marks)

The first quartile Q_1: 14;
the third quartile Q_3: 37;
the interquartile range IQR: 37-14=23.

(c) (5 marks)

A data value less than $Q_1 - 1.5(IQR)$ or greater than $Q_3 + 1.5(IQR)$ can be considered an outlier.
$Q_1 - 1.5(IQR) = 14 - 1.5 \cdot 23 = -20.5$ $\qquad Q_3 + 1.5(IQR) = 37 + 1.5 \cdot 23 = 71.5$
A number out of the range $[-20.5, 71.5]$ can be considered an outlier. Therefore, 80 is the outlier.

2. (20 marks)

(a) (6 marks)

$$\begin{aligned} P(\text{at least 1 ace in 4 rolls}) =& 1 - P(\text{no ace in 4 rolls}) = 1 - (P(\text{no ace in 1 roll}))^4 \\ =& 1 - (1 - P(\text{1 ace in 1 roll}))^4 = 1 - (1 - \frac{1}{6})^4 = 0.518 \end{aligned}$$

(b) (6 marks)

$$\begin{aligned} P(\text{at least 1 double-ace in 25 rolls}) =& 1 - P(\text{no double-ace in 25 rolls}) = 1 - (P(\text{no double-ace in 1 roll}))^{25} \\ =& 1 - (1 - P(\text{1 double-ace in 1 roll}))^{25} = 1 - (1 - \frac{1}{6} \cdot \frac{1}{6})^{25} = 0.5055 \end{aligned}$$

(c) (8 marks)

This is a binomial distribution, and the probability of success, i.e. getting a 6, for each trial is $\frac{1}{6}$. Therefore,

$$P(\text{4 sixes in 7 rolls}) = \binom{7}{4}(\frac{1}{6})^4(1 - \frac{1}{6})^{7-4} = 0.0156$$

3. (10 marks)

Let A denotes the event that the professor die, and B be the event that he was a smoker.
Given that $P(B) = 0.05$, $P(B^C) = 1 - 0.05 = 0.95$, $P(A|B) = 0.05$, $P(A|B^C) = 0.005$, we can get

$$\begin{aligned} P(B|A) =& \frac{P(A \cap B)}{P(A)} = \frac{P(B) \cdot P(A|B)}{P(B) \cdot P(A|B) + P(B^C) \cdot P(A|B^C)} \\ =& \frac{0.05 \times 0.05}{0.05 \times 0.05 + 0.95 \times 0.005} = 34.4828\% \end{aligned}$$

4. (15 marks)

(a) (8 marks)

$\mu = 14000$, $\sigma = 1500$, $k = \dfrac{17100 - \mu}{\sigma} = \dfrac{31}{15}$

Let X be a randomly selected individual's income. Use Chebyshev's Theorem,

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) \geq 1 - \frac{1}{k^2} = 0.7659.$$

Thus, the probability that a randomly selected individual's income is above 17100 is at most $1 - 0.7659 = 0.2341$.

Remark: Since we did not assume the distribution is symmetric, the answer $\frac{0.2341}{2}$ is incorrect.

(b) (7 marks)

Let X be the number of class absents resulted from illness, then $X \sim B(n = 5, p = \frac{3}{4})$, thus,

$$\begin{aligned} P(X \leq 3) =& \sum_{x=0}^{3} b(x; n = 5, p = \frac{3}{4}) = 1 - P(X = 4) - P(X = 5) \\ =& 1 - \binom{5}{4}(\frac{3}{4})^4(1 - \frac{3}{4})^{5-4} - \binom{5}{5}(\frac{3}{4})^5(1 - \frac{3}{4})^{5-5} = 0.3672 \end{aligned}$$

5. (15 marks)

Let X be the number of people arriving in a 3-minute period, then $X \sim Poisson(\lambda = 5)$ since $\frac{\text{100 visits}}{\text{60 minutes}}$ is the same as $\frac{\text{5 visits}}{\text{3 minutes}}$.

(a) (7 marks)

$P(X = 0) = \dfrac{e^{-5}5^0}{0!} = 0.0067$

(b) (8 marks)

$P(X > 5) = 1 - P(X \le 5) = 1 - \sum_{x=0}^{5} \dfrac{e^{-5}5^x}{x!} = 0.384$

6. (15 marks)

Let X be the height of an individual in city H, thus, $X \sim N(\mu = 1.65, \sigma^2 = 0.05^2)$

(a) (7 marks)

$$\begin{aligned} P(1.6 < X < 1.7) =& P(\frac{1.6-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1.7-\mu}{\sigma}) \\ =& P(\frac{1.6-1.65}{0.05} < Z < \frac{1.7-1.65}{0.05}) \\ =& P(-1 < Z < 1) = 1 - 2P(Z \le -1) = 1 - 2 \times 0.1587 = 0.6826 \end{aligned}$$

(b) (8 marks)

By Central Limit Theorem, $\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \sim N(\hat{\mu} = \mu, \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n})$, thus,

$$\begin{aligned} P(\overline{X} < 1.6) =& P(\frac{\overline{X}-\hat{\mu}}{\sigma_{\bar{X}}} < \frac{1.6-\hat{\mu}}{\sigma_{\bar{X}}}) = P(\frac{\overline{X}-\hat{\mu}}{\sigma_{\bar{X}}} < \frac{1.6-1.65}{0.05/\sqrt{16}}) \\ =& P(Z < -4) = 3.167 \times 10^{-5} \approx 0 \end{aligned}$$

7. (10 marks)

Let $A = \{12 \text{ heads in } 20 \text{ tosses}\}$. Then $P(A)$ is the binomial probability:
$P(A) = \binom{20}{12}\left(\frac{1}{2}\right)^{12}\left(1-\frac{1}{2}\right)^{20-12} = \dfrac{20!}{12!8!}\left(\frac{1}{2}\right)^{20}$

(a) (5 marks)

Let $B = \{1\text{st toss landed heads}\}$. We want to find $P(B|A)$. Note that

$$\begin{aligned} B \cap A =& \{1\text{st toss landed head and } 12 \text{ heads in } 20 \text{ tosses}\} \\ =& \{1\text{st toss landed head and } 11 \text{ heads in the remaining } 19 \text{ tosses}\} \end{aligned}$$

Let $C = \{11 \text{ heads in tosses } 2 \text{ through } 20\}$. By the independence of the trials, the events B and C are independent. By the definition of C, we can get $B \cap A = B \cap C$.
$P(B) = \dfrac{1}{2}, \qquad P(C) = \binom{19}{11}\left(\frac{1}{2}\right)^{19}.$

Thus,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(C \cap B)}{P(A)} = \frac{P(B)P(C)}{P(A)}$$
$$= \frac{\frac{1}{2} \cdot \binom{19}{11} \left(\frac{1}{2}\right)^{19}}{\binom{20}{12} \left(\frac{1}{2}\right)^{20}} = \frac{\binom{19}{11}}{\binom{20}{12}} = \frac{12}{20} = \frac{3}{5}$$

Alternatively, given that there are 12 heads in 20 tosses, and since all outcomes are equally likely, any ordering of 12 heads and 8 tails is equally likely. The number of ordering that have a head on the first toss is $\binom{19}{11}$. It is because after fixing the first toss as a head, the other 11 heads must be assigned to the remaining 19 tosses. The total number of orderings is $\binom{20}{12}$. Thus,

$$P(B|A) = \frac{\binom{19}{11}}{\binom{20}{12}} = \frac{12}{20} = \frac{3}{5}$$

(b) (5 marks)

Again, all ordering are equally likely, so after fixing the last two tosses as heads there are $\binom{18}{10}$ ways to order the remaining 10 heads and 8 tails, so the probability that the last two tosses landed heads given there were 12 heads is $\frac{\binom{18}{10}}{\binom{20}{12}} = \frac{12 \cdot 11}{20 \cdot 19} = \frac{33}{95} \approx 0.3474$.