<u>Part A: Multiple Choice Questions:</u> Each correct answer is worth 2 marks. Please enter your answer in the table below.

Question	1	2	3	4	5	6	7	8	9	10
Answer	B	C	C	D	C	A	B	B	C	B

1. A fair octagonal (eight sided) die, with faces marked 1 to 8, is thrown as an experiment, the result being the number on the face on which the die rests. The following events are defined: $E_1 = \{1, 2, 3, 4, 5\}$; $E_2 = \{2, 4, 6, 8\}$; $E_3 = \{1, 3, 5, 7\}$. Find $P(E_3 | (E_1 or E_2))$

A. 0 **B.** 3/7 **C.** 2/5 **D.** 3/4

2. Let $P(A \mid B) = 0.5$, $P(B) = 0.25$, $P(A \cup B) = 0.75$.
$P(A)$ = ?

A. 0.633 **B.** 0.75 **C.** 0.625 **D.** 0.50

3. The mode, median, and mean of a distribution are 30, 40, and 60. What can you say about the shape of the distribution?

A. The distribution is symmetric. **B.** The distribution is left skewed.

C. The distribution is right skewed. **D.** Neither of the three distribution above.

4. The shape of the distribution of a midterm marks is not known, however the mean is 75 and the variance is 25. At least what percentage of the marks are between 60 and 90

A. 99.7%. **B.** 75% **C.** 95% **D.** 88.88%

5. Which one of the following summary measures is affected most by outliers (or extreme values)?

A. first quartile **B.** median **C.** mean **D.** interquartile range (IQR)

6. Suppose you draw a card randomly from a standard deck of 52 playing cards and observe the suit and rank of the card. Let *Q* be the event that we draw a "Queen" and *H* be the event that you draw a "Heart". What can you say about the relationship between the events *Q* and *H*?

A. *Q* and *H* are independent. **B.** *Q* and *H* are mutually exclusive.

C. *Q* and *H* are equiprobable. **D.** Both (A) and (B).

7. You are told that the equation of the least-squares regression line for predicting y from x, based on $n = 10$ pairs of observations is $y = 1.73 - 2x$. If $s_x^2 = 1$ and $s_y^2 = 16$ then the equation of the least-squares regression line for predicting x from y, based on the same $n = 10$ pairs of observations is $x = a + by$, where b is

A. –1/16 **B.** –1/8 **C.** –3/16 **D.** –1/2

8. Given two bivariate observations, (2, 3) and (3.5, b), where $b < 2$, the value of the correlation coefficient r

A. is always +1

B. is always –1

C. is always either +1 or –1, depending on the value of b

D. can be any number in the range 0 and +1, depending on the values of b.

9. Which of the following randomly selected measurements, X, might be considered a potential outlier if it was selected from the given population?

A. X=16 from a population with $\mu = 8$ and $\sigma = 4$

B. X=5 from a population with $\mu = 1$ and $\sigma = 4$

C. $X = -3$ from a population with $\mu = 4$ and $\sigma = 2$

D. $X = 4$ from a population with $\mu = 0$ and $\sigma = 2$

10. Which of the following is/are binomial experiment/experiments?

a. Drawing 7 balls without replacement from a box that contains 10 balls, 5 of which are red and 5 are blue, and observing the colours of the drawn balls.
b. Drawing 7 balls with replacement from a box that contains 10 balls, 5 of which are red and 5 are blue, and observing the colours of the drawn balls.
c. Selecting a few households from New York and observing whether or not they own stocks when it is know that 27% of all households in New York own stocks.

A. Only (a) is true. **B.** Only (b) and (c) are true.
C. Only (c) is true. **C.** Only (a) and (c) are true.

<u>Part B: Long Answer Questions:</u> FOR FULL MARKS, YOU MUST SHOW ALL OF YOUR WORK AND, WHEN APPLICABLE, DEFINE VARIABLES, THEIR DISTRIBUTIONS, AND PARAMETERS.

1. [7 marks] A certain city has one morning newspaper and one evening newspaper. It is estimated that 20% of the city's households subscribe to the morning paper and 60% subscribe to the evening paper. Of those who subscribe to the morning paper, 80% also subscribe to the evening paper.

A. [2] What proportion of households subscribes to both papers?

We know that $P(M) = 0.20$, where M=event of subscribing to the morning paper
$P(E) = 0.60$, where E=event of subscribing to the evening paper
$P(E \mid M) = 0.80$

[1] $P(M \cap E) = P(M).P(E \mid M) = (0.20).(0.80) =$ **[1]** $\underline{\underline{0.16}}$

B. [2] If a household subscribes to the evening paper, what is the probability that they subscribe to the morning paper?

[1] $P(M \mid E) = \dfrac{P(M \cap E)}{P(E)} = \dfrac{0.16}{0.60} =$ **[1]** $\underline{\underline{0.2666}}$

C. [2] What proportion of households subscribes to at least one paper?

[1] $P(at\ least\ 1\ \ paper) = P(M \cup E) = P(M) + P(E) - P(M \cap E) =$
$= 0.20 + 0.60 - 0.16 =$ **[1]** $\underline{\underline{0.64}}$

D. [1] What proportion of households does not subscribe to either paper?

$P(no\ \ paper) = 1 - P(M \cup E) = 1 - 0.64 =$ **[1]** $\underline{\underline{0.36}}$

2. [11 marks] An encyclopedia salesperson estimates the probabilities for the number of encyclopedia sold in a week as follows:

Number sold (x)	0	1	2	3	4	5
$p(x)$	0.05	0.30	0.40	0.10	0.10	0.05

Assume that the number of sales in one week is independent of the number in any other.

A. [1] What is the probability, for a given week, that fewer than 3 encyclopedias will be sold?

Let X = number of encyclopedias sold in a week, then
$P(X < 3) = P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.05 + 0.30 + 0.40 =$ **[1]** $\underline{\underline{0.75}}$

B. [2] Find the expected number of encyclopedias sold in a week.

[1] $E(X) = \sum_i x_i p(x_i) = 0(0.05) + 1(0.30) + 2(0.40) + 3(0.10) + 4(0.10) + 5(0.05) =$ **[1]** 2.05

i.e. Salesperson can expect to sell on average 2.05 encyclopedias each week.

C. [3] Find the standard deviation of the number of encyclopedias sold in a week.

[2; 1 for formula, 1 for final answer]

$$V(X) = E(X^2) - \left[E(X)\right]^2 = \sum_i x_i^2 p(x_i) - \left[E(X)\right]^2 = 0^2(0.05) + 1^2(0.30) + 2^2(0.40) + 3^2(0.10) + 4^2(0.10) + 5^2(0.05) -$$
$$-(2.05)^2 = 0 + 0.30 + 1.60 + 0.90 + 1.60 + 1.25 - 4.2025 = 5.65 - 4.2025 = 1.4475$$

[1] $\therefore s.d.(X) = \sqrt{V(X)} = \sqrt{1.4475} = \underline{\underline{1.2032}}$

D. [2] The salesperson receives a weekly salary of $250, plus a commission of $60 for each encyclopedia sold. Find the expected value of the total amount the salesperson will receive each week.

[1] Let P= profit, hence $P = 60X + 250$
[1] $\therefore E(P) = E(60X + 250) = 60E(X) + 250 = 60(2.05) + 250 = \underline{\underline{373}}$

i.e. The salesperson is expected to receive on average $373 each week.

E. [3] Find the standard deviation of the total weekly amount.

[2; 1 for formula, 1 for final answer]

$$V(P) = V(60X + 250) = 60^2 V(X) = 60^2(1.4475) = 5211$$

[1] $\therefore s.d.(P) = \sqrt{V(P)} = \sqrt{5211} = 72.1873 \cong \$\underline{\underline{72.19}}$

3. [11 marks] The probability of a telesales representative making a sale on a customer call is 0.15.

[1] X = number of sales in 10 calls

[2] X~Binomial (10, 0.15)

A. Find the probability that

(i) [3] no sales are made in 10 calls,

$P(X=0)$ = **[2]** $(0.15)^0 (0.85)^{10}$ = **[1]** 0.197

(ii) [6] more than 3 sales are made in 10 calls.

[1] $P(X>3) = 1 - P(X \leq 3)$ = **[1]** 1–0.95 = **[1]** 0.05

Note: Students may also use the binomial density function to compute this probability.

B. [2] How many representatives are required to achieve a mean of at least 5 sales each day.

$E(X)=np$= **[1]** $n(0.15) \geq 5$ then $n \geq 5/0.15$=**[0.5]**33.3 at least **[0.5]** 34 sales needed

4. [10 marks]Consider the following stem and leaf plot for the ages of 20 subjects on a medical study

3	2
3	79
4	01112234
4	5556779
5	01

Leaf unit = 1.0

A. [2] Give the 5th and 9th percentiles.

5th percentile:
[0.5] Rank = 21(0.05) = 1.05
[0.5] 5th = 32 + (37–32)0.05 = 32.25

9th percentile:
[0.5] Rank = 21(0.09) = 1.89
[0.5] 5th = 32 + (37–32)0.89 = 36.45

B. [3] Give the first quartile, median, and third quartile.

Q1:
[0.5] Rank = 21(0.25) = 5.25
[0.5] Q1 = 41 + (41 – 41)0.25 = 41. Simply 41, is also an acceptable answer.

Median:
[0.5] Rank = 21(0.5) = 10.5
[0.5] Median = 43 + (44 – 43)0.5 = 43.5.

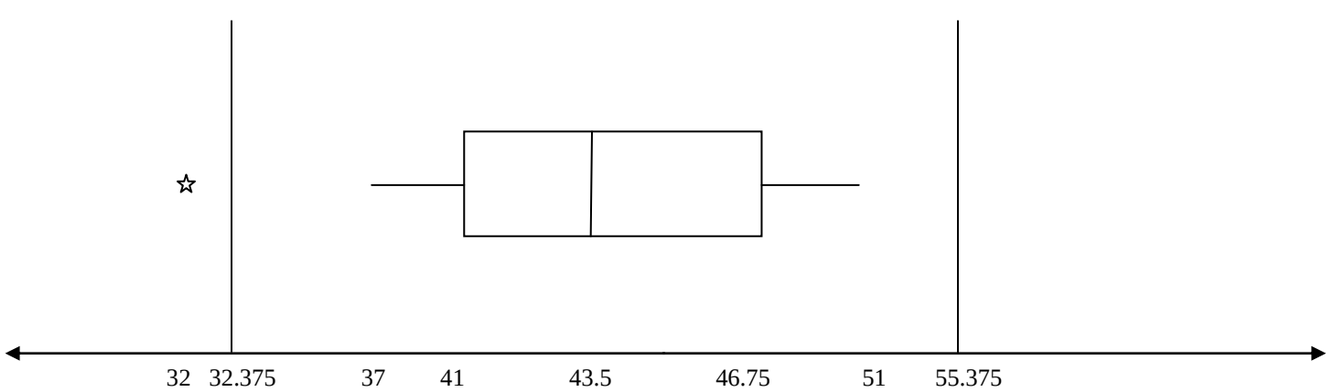
Q3:
[0.5] Rank = 21(0.75) = 15.75
[0.5] Q3 = 46 + (47 – 46)0.75 = 46.75.

C. [5] Sketch the box plot.

[1]*IQR* = 46.75 – 41 = 5.75

Fence:
[0.5] Lower bound = 41 – 1.5(5.75) = 32.375
[0.5] Upper bound = 46.75 + 1.5(5.75) = 55.375

[2.5; 1.5 of graph, 1 for labelling whiskers correctly]



[0.5] The observation 32 is an outlier.

5. [6 marks] A diagnostic blood test for performance enhancement drugs used among athletes was thought to be reasonably accurate. Among drug users, 98% had a positive result. Among nondrug users, 95% had a negative test. Assume that 1% of all athletes are drug users.

Let D = drug user, D_T = test says a tested person is a drug user
$P(D)$ =0.01, $P(D_T|D)$ =0.98, $P(D_T^c|D^c)$ = 0.95

A. [3] What proportion of athletes test positive using this test?

[0.5]$P(D_T) = P(D)P(D_T|D) + P(D^c)P(D_T|D^c)$ = **[2]**0.01 (0.98) + 0.99(0.05) = **[0.5]**0.0593

B. [3] If an athlete was tested positive, what the probability he/she is indeed a drug user?

[0.5]$P(D|D_T) = P(D \cap D_T)/P(D_T)$ = **[2]**(0.01)(0.98)/0.0593 ≈ **[0.5]**0.165

Note: Students can use a probability tree to do this.