CHENNAI MATHEMATICAL INSTITUTE
M.Sc. Data Science Entrance Examination
4^{th} October 2020

Enter your *Admit Card Number*: D – ☐☐☐ – ☐☐☐☐

IMPORTANT INSTRUCTIONS

- This booklet has 11 printed sheets, including this cover page. Sheet 2 (page 1) contains space to answer part A, sheets 3 to 6 contain *twenty* questions in part A and sheets 7 to 11 contain *twenty* questions in part B.
  - For questions in part (A), you only have to write your answer on the designated page on the appropriate line. For example, if the answer is parts (a) and (c), write only (a) and (c) on the line and if the answer is 2781, write this number on the line.
  - For questions in part (B), you have to write your answer with short explanation in the space provided below the question.
  - For numerical answers, the following forms are acceptable: fractions, decimals, symbolic e.g.: $\binom{n}{r}$, ${}^nP_r$, $n!$ etc.

  For rough work use the blank pages at the end.
- **Time allowed is 3 hours. Total points: 100** = 40 for part A + 60 for part B.
- **Part A will be used for screening.** Part B will be graded only if you score a certain minimum in part A. However your scores in both parts will be used while making the final decision.

For office use only

	Points	Remarks
Part A		
Part B		
Total		

B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	Total
B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	Total

This page intentionally left blank.

Part A - Answers

This is the only place that will be seen for grading part A. So carefully and clearly write the answers to each question on the designated line below. Write only the final answers, do not show any intermediate work. Illegible/unclear answers will not be considered.

A1.____________________ A11. ____________________

A2.____________________ A12. ____________________

A3.____________________ A13. ____________________

A4.____________________ A14. ____________________

A5.____________________ A15. ____________________

A6.____________________ A16. ____________________

A7.____________________ A17. ____________________

A8.____________________ A18. ____________________

A9.____________________ A19. ____________________

A10.____________________ A20. ____________________

This page intentionally left blank.

Notation

- A function f from a set A to a set B is said to be **injective (or one-to-one)** if $f(x) = f(y)$ implies $x = y$ for all $x, y \in A$;
- f is said to be **surjective (or onto)** if for every $y \in B$ there exists $x \in A$ such that $f(x) = y$;
- f is said to be **bijective** if it is both injective and surjective;
- f is said to be **invertible** if there exists a function g from B to A such that $f(g(y)) = y$ for all $y \in B$ and $g(f(x)) = x$ for all $x \in A$ and then g is said to be inverse of f and is denoted by f^{-1}.
- For a matrix A, $|A|$ denotes the determinant of A and A^T denotes the transpose of A.
- For a set X, X^c denotes the complement of X.

Part A - Questions

This section consists of some questions requiring a single answer and some multiple choice questions. For questions that ask you to calculate a value, you will be assessed based on the answer you provide. No explanation is required. In multiple choice questions, there may be multiple correct choices. You have to select all the correct options and no incorrect option to get full marks. There is no partial credit. Write the correct options / answer in the space provided on page 1. Only page 1 will be seen for grading part A.

1. Consider the following program. Assume that x and y are integers.

```
f(x, y)
{
    if (y != 0)
        return (x * f(x, y-1));
    else
        return 1;
}
```

What is `f(6,3)`?

(a) 243
(b) 729
(c) 125
(d) 216

2. Consider the matrices

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 15 \\ 0 & 0 & 0 & 13 & 14 \\ 0 & 0 & 10 & 11 & 12 \\ 0 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 4 & 3 & 2 & 1 \\ 9 & 8 & 7 & 6 & 0 \\ 12 & 11 & 10 & 0 & 0 \\ 14 & 13 & 0 & 0 & 0 \\ 15 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Which of the following hold true?

(a) $|A| = |B|$
(b) $\text{trace}(A) = \text{trace}(B)$
(c) $|A| = -|B|$
(d) $\text{trace}(AB) = \text{trace}(BA)$

3. Consider the following program. Assume that all variables are integers. Note that x%y computes the remainder after dividing x by y. The division is an integer division. For example, 1/3 will return zero while 10/3 will return 3.

```
g(n)
{
    result = 0;
    i = 1;

    repeat until (n == 0)
    {
        remainder = n%2;
        n = n / 2;
        result = result + (remainder * i);
        i = i * 10;
    }

    return result;
}
```

What is `g(25)`?

(a) 11001
(b) 10011
(c) 11011
(d) 10101

4. Which of the following limits are correct?

   (a) $\lim_{x \to 0} \dfrac{x^2 + 2x}{2x} = 1$

   (b) $\lim_{x \to 1/2} \dfrac{2x^2 + x - 1}{2x - 1} = \dfrac{3}{2}$

   (c) $\lim_{x \to \infty} 18x^3 - 12x^2 + 1 = \infty$

   (d) $\lim_{x \to -\infty} 18x^3 - 12x^2 + 1 = -\infty$

5. As per the data released by the US Department of Health, Education and Welfare, the number of Ph.D. degrees conferred in Earth Sciences from the year 1948 to 1954 is as given in Table 5.

Year	Degrees
1948	57
1949	88
1950	125
1951	135
1952	126
1953	146
1954	141

   Based on the average of all available three year moving averages of annual growth rate, and the number of degrees in 1954, what will be the (approximate) predicted number of degrees in earth sciences in 2020? Note that a moving average or a rolling average is an average of a subset of data points. Choose the best answer.

   (a) 900

   (b) 9,000

   (c) 9,00,000

   (d) 90,00,000

6. Suppose that A is an $n \times n$ matrix with $n = 10$ and b is an $n \times 1$ vector. Suppose that the equation $Ax = b$ for an $n \times 1$ vector does not admit any solution. Which of the following conclusions can be drawn from the given information?

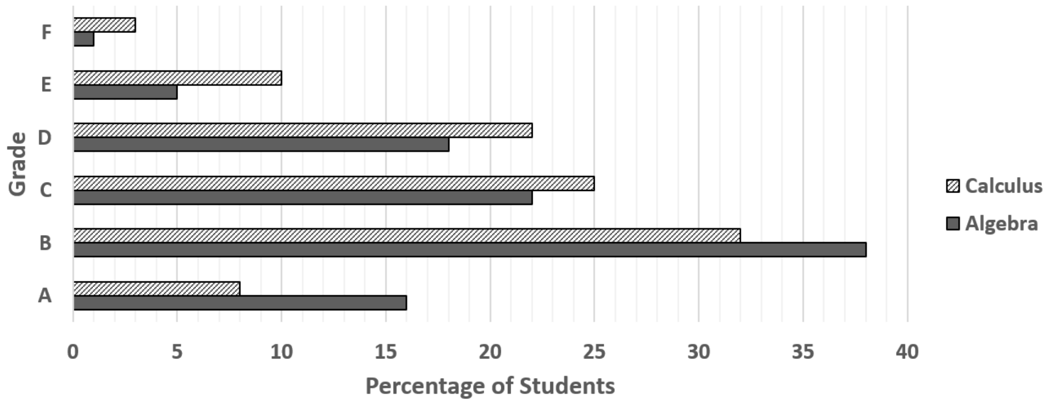
   (a) A^{-1} does not exist.

   (b) The equation $A^T x = b$ also does not admit any solution.

   (c) $|A| = 0$.

   (d) Suppose c is another $n \times 1$ vector such that $Ax = c$ also does not admit a solution. Then the vector c is a constant multiple of the vector b.

7. Consider the following bar chart:



Which of the following are true?

(a) Number of students who scored A in Algebra is higher than the number of students who scored A in Calculus.

(b) Percentage of students who scored A or B in algebra is lower than the percentage of students who scored A or B in calculus.

(c) Calculus is easier than algebra.

(d) Considering this data, the average percentage of students scoring A is 12%.

8. Let $A = ((a_{ij}))$ be a 7×7 matrix with $a_{i,i+1} = 1$ for $1 \leq i \leq 6$, $a_{7,1} = 1$ and all the other elements of the matrix are zero. Which of the following statements are true?

(a) $|A| = 1$

(b) $\text{trace}(A) = 0$

(c) $A^{-1} = A$

(d) $A^7 = I$, where I is the identity matrix

9. Let A and B be events such that $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cup B) = 0.7$. Which of the following are true? (For sets A, B, $A \Delta B = (A^c \cap B) \cup (A \cap B^c)$).

(a) A and B are mutually exclusive

(b) A and B are independent

(c) $P(A \Delta B) = 0.1$

(d) $P(A^c \cup B^c) = 0.8$

Description for the following 2 questions:

The lifespan of a battery in a car follows Gamma distribution with probability density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}, \quad 0 < x < \infty,$$

where $\alpha > 0$ and $\beta > 0$. The mean and variance of a Gamma distribution are $\mathbb{E}(X) = \frac{\alpha}{\beta}$ and $\mathbb{V}(X) = \frac{\alpha}{\beta^2}$ respectively. From historical data the mean and variance of the lifespan of a battery are estimated as 4 years and 2 years respectively.

10. Which of the following statements are correct?

    (a) $\alpha = 16$ and $\beta = 4$

    (b) $\alpha = 8$ and $\beta = 2$

    (c) $\mathbb{E}(X^2) = \frac{\alpha}{\beta}(\frac{1+\alpha}{\beta})$

    (d) $\mathbb{E}(X^2) = 18$

11. Out of a large number of cars produced by the automaker, the percentage of batteries that will last for more than 8 years is

    (a)
    $$\left[\int_0^8 \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} dx\right] \times 100\%.$$

    (b)
    $$\left[1 - \int_0^8 \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} dx\right] \times 100\%.$$

    (c)
    $$\left[\int_8^\infty \frac{x\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} dx\right] \times 100\%.$$

    (d)
    $$\left[\int_0^8 \frac{x\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} dx\right] \times 100\%.$$

12. How many squares are there on a 7 x 7 chessboard?

    (a) 49

    (b) 204

    (c) 203

    (d) 140

13. It is mid-semester exam week at CMI and first-year students from both M.Sc. Data Science (DS) and M.Sc. Computer Science (CS) have their exams scheduled for Monday from 10 a.m. to 1 p.m. in Lecture Hall 1. The first row in Lecture Hall 1 has six seats. In how many different ways can three M.Sc. DS students - Anish, Binish and Finish - and three M.Sc. CS students - Ramesh, Suresh, and Ragesh - be seated in this row, in such a way that two students from the same course do not sit next to each other?

    (a) 36
    (b) 48
    (c) 72
    (d) 96

14. Suppose you roll two six-sided fair dice with faces numbered from 1 to 6 and take the sum of the two numbers that turn up. What is the probability that:

    - the sum is 12;
    - the sum is 12, *given that* the sum is even;
    - the sum is 12, *given that* the sum is an even number greater than 4?

    (a) $\frac{1}{36}$, $\frac{1}{18}$, and $\frac{1}{12}$, respectively
    (b) $\frac{1}{36}$, $\frac{1}{18}$, and $\frac{1}{14}$, respectively
    (c) $\frac{1}{36}$, $\frac{1}{16}$, and $\frac{1}{14}$, respectively
    (d) $\frac{1}{36}$, $\frac{1}{16}$, and $\frac{1}{12}$, respectively

15. Let $f(x)$ be a real-valued function all of whose derivatives exist. Recall that a point x_0 in the domain is called an *inflection point* of $f(x)$ if the second derivative $f''(x)$ changes sign at x_0. Given the function $f(x) = \dfrac{x^5}{20} - \dfrac{x^4}{2} + 3x + 1$, which of the following statements are true?

    (a) $x_0 = 0$ is not an inflection point.
    (b) $x_0 = 6$ is the only inflection point.
    (c) $x_0 = 0$ and $x_0 = 6$, both are inflection points.
    (d) The function does not have an inflection point.

16. Which of the following are true?

    (a) $\dfrac{2019}{2020} < \dfrac{2020}{2021}$
    (b) $x + \dfrac{1}{x} \geq 2$ for all $x > 0$
    (c) $2^{60} > 5^{24}$
    (d) $2^{314} < 31^{42}$

17. The identity

$$\frac{1}{(1-2r)} = \sum_{k=0}^{\infty}(2r)^k$$

is true

(a) if and only if $r \neq \frac{1}{2}$

(b) if and only if $0 \leq r < \frac{1}{2}$

(c) if and only if $-\frac{1}{2} \leq r < \frac{1}{2}$

(d) if and only if $-\frac{1}{2} < r < \frac{1}{2}$

18. The sum and product of the roots of the polynomial $9x^2 + 171x - 81$ are, respectively:

(a) -19 and -9

(b) 19 and 9

(c) -9 and 19

(d) 9 and -19

19. Choose the conclusions that follow logically from the statements given below.

i Nobody who really appreciates A.R.Rahman fails to subscribe to his YouTube channel.

ii Owls are hopelessly ignorant of music.

iii No one who is hopelessly ignorant of music ever subscribes to A.R.Rahman's YouTube channel.

(a) Anyone who subscribes to A.R.Rahman's YouTube channel is hopelessly ignorant of music.

(b) Owls don't really appreciate A.R.Rahman.

(c) Owls are not really appreciated by A.R.Rahman.

(d) Anyone who really appreciates A.R.Rahman is not hopelessly ignorant of music.

20. Which of the following inequalities are true?

(a) $e^x \geq (1+x)$ for $x \geq 0$

(b) $e^x \leq (1+x)$ for $x < 0$

(c) $\ln(x) < (1+x)$ for $x > 0$

(d) $e^x < x^2$ for all real numbers x

This page intentionally left blank.

Part B

For questions in part (B), you have to write your answer with a short explanation in the space provided below the question. For numerical answers, the following forms are acceptable: fractions, decimals, symbolic e.g.: $\binom{n}{r}$*,* ${}^nP_r$*,* $n!$ *etc.*

1. For any string `str`, `length(str)` returns the length of the string, `append(str1,str2)` concatenates `str1` with another string `str2`, and `trim(str)` removes any spaces that exist at the end of the string `str`. The function `reverse(str, i, j)` reverses the part of the string from position `i` to position `j`. Assume that position 0 refers to the first character in the string. What does the following pseudo-code do?

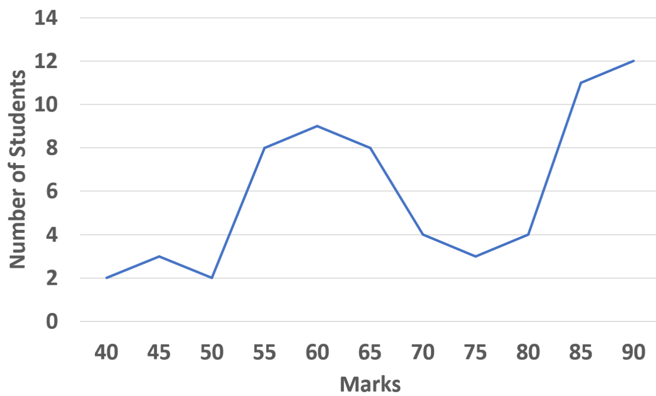
```
def manipulate(string str)
{
    reverse(str, 0, length(str)-1);
    append(str, ‘ ’);
    n = length(str);
    j = 0;

    for (i = 0; i < n; i=i+1) {
        if (str[i] is ‘ ’) {
            reverse(str, j, i-1);
            j = i + 1;
        }
    }

    trim(str);
    return str;
}
```

2. Consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Find A^n, in terms of n, for $n \geq 2$.

3. The following graph shows the performance of students in an exam. The marks scored by every student are a multiple of five. The j^{th}-percentile u^* for a discrete data $x_1, x_2, \ldots, x_n$ is defined as follows. Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the ordering of the data in ascending order. Let $t = \frac{j \cdot n}{100}$ and let k be an integer such that $k \le t < (k+1)$ and let $s = t - k$. Then $u^* = x_{(k)} + s * (x_{(k+1)} - x_{(k)})$. Here, $x_{(n+1)}$ is defined to be $x_{(n)}$.

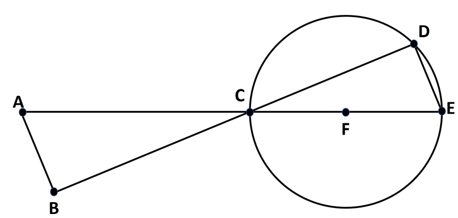


Based on the information presented in the graph, answer the following questions.

(a) Compute the 10th percentile of marks.

(b) Is the median score higher than the mean score?

4. In the figure shown below, the circle has diameter 5. Moreover, AB is parallel to DE. If DE = 3 and AB = 6, what is the area of triangle ABC?



The following description holds for the two problems below.
A *permutation* σ is a bijection from the set $[n] = \{1, 2, \ldots, n\}$ to itself. We denote it using the notation

$$\left(\begin{array}{cccc} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{array} \right),$$

e.g. if $n = 3$ then $\left(\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 3 & 1 \end{array} \right)$ denotes the permutation defined by $\sigma(1) = 2$, $\sigma(2) = 3$ and $\sigma(3) = 1$. An *inversion* in σ is a pair (i, j) such that $i < j$ but $\sigma(i) > \sigma(j)$. The *sign* of a permutation σ (denoted $sgn(\sigma)$) is defined to be $(-1)^{inv(\sigma)}$, where $inv(\sigma)$ denotes the total number of inversions in σ. In the above example, there are 2 inversions corresponding to the pairs $(1, 3)$ and $(2, 3)$ so that $sgn(\sigma) = (-1)^2 = 1$. For each permutation σ, define a matrix A_σ as follows:

$$A_\sigma(i, j) = \begin{cases} 1 & \text{if } \sigma(i) = j \\ 0 & \text{otherwise} \end{cases}$$

5. Find $sgn(\sigma)$, $sgn(\tau)$, A_σ and A_τ for the following permutations:

$$\sigma = \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 1 & 6 \end{array} \right), \quad \tau = \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 3 & 2 & 6 & 5 \end{array} \right).$$

6. What are the determinants $|A_\sigma|$ and $|A_\tau|$? Can you relate these with the signs of permutations σ and τ?

7. The *case fatality rate* (CFR) of a disease is the ratio of the number of deaths from the disease to the total number of people diagnosed with the disease ("patients"), and is usually expressed as a percentage. It has been reported that the CFR of Pandemic-20 among elderly people in Gondwanaland is 20%. One way of treating a Pandemic-20 patient involves putting the patient on a ventilator. It has been observed that in Gondwanaland, 60% of the elderly Pandemic-20 patients who survived the disease had been put on a ventilator, and 10% of the elderly Pandemic-20 patients who died from the disease had been put on a ventilator. What is the probability that an elderly Pandemic-20 patient in Gondwanaland survives the disease if they were put on a ventilator as part of the treatment?

8. Owing to a defect in a certain machine which makes N95 masks, there is a 0.1% probability that a mask it makes is **not** effective in preventing airborne viruses from being inhaled.

   (a) What is the probability that the first 1000 masks that the machine produces are effective? (You may leave your solutions as arithmetic expressions; there is no need to compute their decimal representations.)

   (b) What is the probability that among the first one crore (10^7) masks that the machine produces, there is at least one mask which is not effective?

9. The International Chess Federation is organizing an online chess tournament in which 20 of the world's top players will take part. Each player will play exactly one game against each other player. The tournament is spread over three weeks; it starts at 9 a.m. on the Monday of Week 1 and ends at 6 p.m. on the Friday of Week 3. Note that *before* 9 a.m. on the Monday of Week 1 *every* player would have completed the *same* number of games in the tournament; namely, zero. Also, *after* 6 p.m. on Friday in Week 3, *every* player would have completed the *same* number of games in the tournament, namely, nineteen.

   Prove that at *any point in time* between 9 a.m. on the Monday of Week 1 and 6 p.m. on the Friday of Week 3, there are *at least two players* who would have completed the *same* number of games in the tournament till that point.

10. Your class has a textbook and a final exam. Let P, Q and R be the following propositions:

    - P: You get an A on the final exam.
    - Q: You do every exercise in the book.
    - R: You get an A in the class.

    Translate the following assertions into propositional formulas using P, Q, R and the propositional connectives $\wedge$ (*and*), $\vee$ (*or*), $\neg$ (*not*) and $\Rightarrow$ (*implies*).

    (a) To get an A in the class, it is necessary for you to get an A on the final.

    (b) You get an A on the final, but you don't do every exercise in this book; nevertheless, you get an A in this class.

11. Two friends Amar and Prem wish to meet at a theme party between 5 p.m. and 6 p.m. (They are said to meet if they are in the room at the same time or if one of them leaves as the other enters.) Once they enter, they stay for exactly 20 minutes.

   (a) Fill in the blanks:

      i. The latest time by which any one of them can enter is ___________;

      ii. If Amar and Prem are to meet, then their entry times can be separated by at most an interval of ___________ minutes.

   (b) What is the probability that Amar and Prem will meet? (Hint: Plot the arrival times on the x- and y-axes.)

12. Let f be a continuous function on the closed interval $[a, b]$ and differentiable on the open interval (a, b), where $a < b$. It is known that there exists c in (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Determine c for $f(x) = 2x^3 - 3x^2 + 7x - 2$ on $[a, b]$ with $a = 1$ and $b = 6$.

13. Let R be the set of all binary relations on the set $\{1, 2, 3\}$. Suppose a relation is chosen from R at random. What is the probability that the chosen relation is symmetric?

14. Peppa and her friends Suzy and Emily are making 3 masks: a knight mask, a pirate mask and an elf mask. The one wearing the knight mask must always tell the truth, the one wearing the pirate mask must always lie and the one wearing the elf mask could sometimes lie and sometimes tell the truth. After making three masks, one of each kind, they all wear one each and make the following statements:

Peppa	:	I am a wearing an elf mask.
Suzy	:	That is true.
Emily	:	I am not wearing an elf mask.

What kind of mask is each one wearing?

15. For the function $f(x) = \ln(x)/x$, show that the maximum value of $f(x)$ occurs when $x = e$. Use this to show that $x^e < e^x$ for all positive values of x.

16. Given the set of letters {`a`,`b`,`c`,`d`,`e`,`f`,`g`,`h`,`i`,`j`,`k`,`l`,`m`}, we can list out all permutations of these letters in lexicographic (dictionary) order. The first three permutations in this list are `abcdefghijklm`, `abcdefghijkml` and `abcdefghijlkm` and the last one is `mlkjihgfedcba`. What permutations would appear immediately before and after the following one in this lexicographically ordered list of permutations?

`bcjameflkihgd`

17. Consider the following code, where `A` is an array of integers of size `size(A)` with values `A[0]` to `A[size(A)-1]`, and `reverse(A,i,j)` reverses the segment `A[i]` to `A[j]` if `i <= j` and has no effect otherwise. For instance, if `A = [0,1,2,3,4,5]`, then `reverse(A,2,4)` would modify `A` to `[0,1,4,3,2,5]`.

```
def mystery(A){
  for j in [0,1,..,size(A)-1] {
    p = j;
    for i in [j,j+1,..,size(A)-1] {
      if A[i] > A[p] {
         p = i;
      }
    reverse(A,j,p);
    }
  }
}
```

(a) What is the effect of this code on an input array `A`?

(b) Suppose `size(A)` is `1000`. How many times is the test `A[i] > A[p]` executed?

18. Eight students are to be seated around a circular table in a circular room. Two seatings are regarded as defining the same arrangement if each student has the same student on his or her right in both seatings: it does not matter which way they face. How many arrangements of these 8 students are there with 2 chosen students, say student A and student B, always sitting together?

19. Let f be a continuous bijection from closed unit interval $[0, 1]$ onto itself. (Recall the *Intermediate Value Theorem*: let f be a real valued continuous function on an interval $[a, b]$. Let $c, d \in [a, b]$ be such that $f(c) < f(d)$ and let $\alpha \in (f(c), f(d))$ be an intermediate value. Then there exists $y \in [a, b]$ such that $f(y) = \alpha$.)

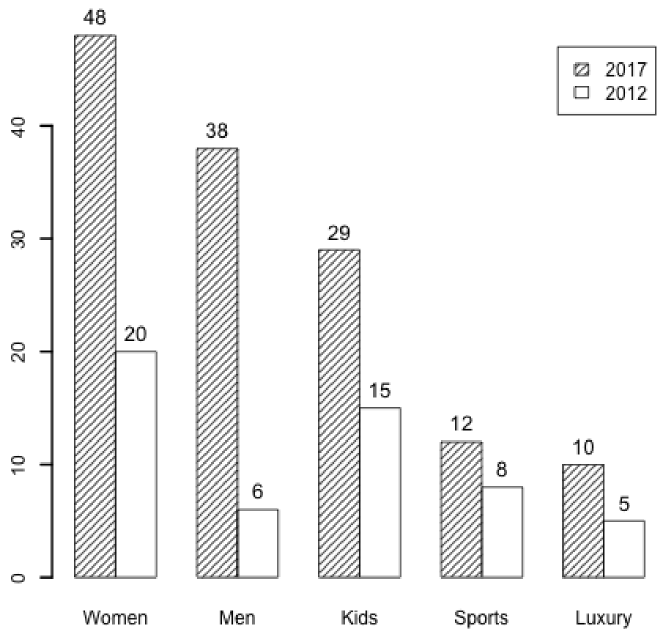
   (i) Show that $f(0)$ equals 0 or 1.

   (ii) Show that $f(1)$ equals 0 or 1.

   (iii) Show that f admits a fixed point.

   (iv) Give an example of such a function wherein the fixed point is unique and an example of a function with more than one fixed point.

20. A multi-national conglomerate sells soap products for five different market segments, namely (i) Women, (ii) Men, (iii) Kids, (iv) Sports, and (v) Luxury. The sales of these five segments (in lakh number of packs) during 2012 and 2017 are shown in the following figure.



(a) By what percentage were the sales of the Women segment in 2017 more than the sales of the Men segment in 2017?

(b) During the period 2012-2017, which segment experienced the minimum rate of increase in sales?