CHENNAI MATHEMATICAL INSTITUTE

M.Sc. Data Science Entrance Examination 2020 - Solutions

Part A - Answers

A1. (d)

A2. (a), (b) and (d)

A3. (a)

A4. All.

A5. (c)

A6. (a) and (c)

A7. (d)

A8. (a), (b) and (d)

A9. (b) and (d)

A10. (b), (c) and (d)

A11. (b)

A12. (d)

A13. (c)

A14. (b)

A15. (b)

A16. (a), (b) and (c).

A17. (d)

A18. (a)

A19. (b) and (d)

A20. (a) and (c)

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 - Solutions

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

Solution: (d) is correct. This code computes x^y. f(6,3) will compute 6^3 which is 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) trace(A) = trace(B)

(c) $|A| = -|B|$

(d) trace(AB) = trace(BA)

Solution: (a), (b) and (d) are true. The given matrices are obtained from $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 & 9 \\ 0 & 0 & 10 & 11 & 12 \\ 0 & 0 & 0 & 13 & 14 \\ 0 & 0 & 0 & 0 & 15 \end{pmatrix}$ **by 10 row and 10 column exchanges respectively.**

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

Solution: (a) is correct. This code converts any given integer to its binary representation.25 is 11001 in binary.
Note: This question had a minor typographical error in the actual exam paper. This has been taken into account when evaluating the question.

4. Which of the following limits are correct?

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

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

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

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

   Solution: All are correct.
   (a) $\frac{x^2 + 2x}{2x} = \frac{x+2}{2}$ **so** $\lim_{x \to 0} \frac{x^2 + 2x}{2x} = \frac{2}{2} = 1$**;**
   (b) $\frac{2x^2 + x - 1}{2x - 1} = \frac{(2x-1)(x+1)}{2x-1} = x + 1$**, so** $\lim_{x \to 1/2} \frac{2x^2 + x - 1}{2x - 1} = \frac{1}{2} + 1 = \frac{3}{2}$**;**
   (c) $\lim_{x \to \infty} 18x^3 - 12x^2 + 1 = \lim_{x \to \infty} x^3(18 - \frac{12}{x} + \frac{1}{x^3}) = \infty$**;**
   (d) $\lim_{x \to -\infty} 18x^3 - 12x^2 + 1 = \lim_{x \to -\infty} x^3(18 - \frac{12}{x} + \frac{1}{x^3}) = -\infty$

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

Solution: (c) is correct. First, we compute the Annual Growth Rate (AGR). Then, we could compute the three year moving averages of AGR. Our calculations should result in the moving averages as shown in Table 1. Note that the moving averages average to 14.23. Starting with 141 in 1954, at a growth rate of 14.23%, we will have 141 * $(1 + 14.23/100)^6 6$ = 9,19,412.8. However, you may approximate to save some time. For instance, the rule of 72 suggests that at a rate of 14%, the number of Ph.D.s will double in approximately (72/14) 5 years. Therefore we have (1954) 150, (1959) 300, (1964) 600... and so on till (2019) 12,28,000. Hence, 9 Lakhs is the closest answer.

Year	Ph.D.s	Annual Growth Rate	Moving Avg
1948	57		
1949	88	54.39	
1950	125	42.05	
1951	135	8	34.81
1952	126	-6.67	14.46
1953	146	15.87	5.74
1954	141	-3.42	1.93
Avg			14.23

Table 1: Solution table for Ph.D. degrees earned in Earth Sciences

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.

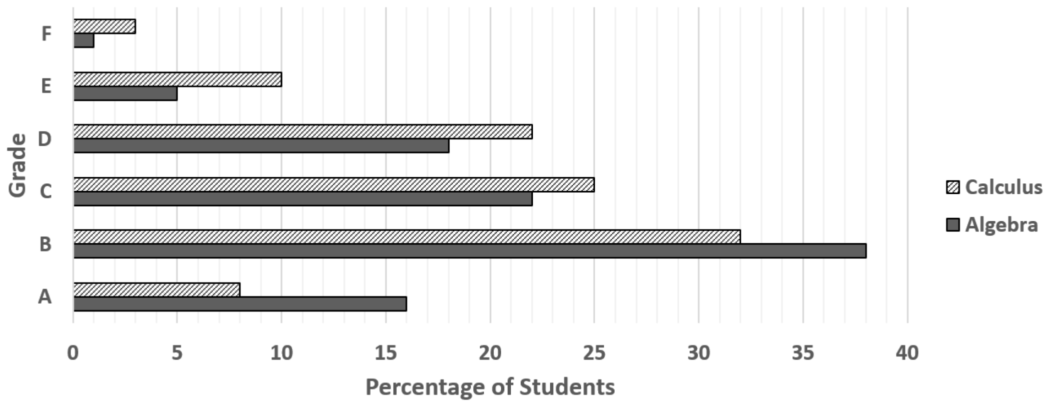
   Solution: All we can conclude is that (a) and (c) are true, for otherwise $Ax = b$ has solution for all b. That (b) is false can be seen by taking $A = (a_{ij})$ with $a_{11} = a_{21} = 1$, $a_{12} = a_{22} = 2$ and the rest of the entries as 0; $b = (b_i)$ with $b_1 = 3$ and $b_2 = 6$ and the rest of b_j's to be 0. Now $Ax = b$ does not admit a solution, but $x_1 = 1$ and $x_2 = 2$ and rest of the x_j's equal to 0 is a solution to $A^T x = b$. In the same example, taking $c_1 = 5$ and $c_2 = 6$ and rest of the entries as 0, we see that $Ax = c$ also does not admit a solution but b and c are linearly independent.

7. Consider the following bar chart:

   Which of the following are true?

   (a) Number of students who scored A grade in Algebra is higher than the number of students who scored A grade 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) Average percentage of students scoring A grade considering this data, is 12%.

Solution: (d) is correct. We do not have the total number of students enrolled in Algebra and Calculus. Hence, we cannot conclude that A graders count in Algebra is higher. 54% students scored A or B in Algebra. Only 40% scored A or B in Calculus. Hence the second option is also false. Nothing in the question tells that calculus is easier than algebra. In fact, the calculus distribution has a heavier tail. Moreover, the data distribution does not mention that the same students over the same semester took both classes. Hence, it would be inappropriate to compare the difficulty of these courses. The final option is indeed true. The average percent of students is $(16 + 8)/2 = 12$.

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

Solution: (a), (b) and (d) are true.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This is a permutation matrix, 6 interchanges of columns, 1st and 2nd, then 2nd and 3rd,... finally 6th and 7th will take this to identity matrix, so determinant is 1. Trace is sum of diagonal elements, which is 0. A^7 **is identity and** $A^{-1} = A^6$ **which is not equal to** A.

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

   Solution : (b) and (d) are correct. (a) is False, for if it were true, $P(A \cup B) = P(A) + P(B) = 0.9$**. (b) is true since** $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.2 = P(A) * P(B)$**. Here,** $P(A\Delta B) = P((A^c \cap B) \cup (A \cap B^c)) = P(A^c \cap B) + P(A \cap B^c))$ **which, by independence, equals** $0.6 * 0.5 + 0.4 * 0.5 = 0.5$**. (d)** $P(A^c \cup B^c) = P(A^c) + P(B^c) - P(A^c \cap B^c) = 0.6 + 0.5 - 0.6 \times 0.5 = 0.8$**.**

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

    Solution: (b), (c) and (d) are correct.

$$\mathbb{E}(X) = \frac{\alpha}{\beta} = 4$$

and

$$\mathbb{V}(X) = \frac{\alpha}{\beta^2} = 2 \implies \frac{\alpha}{\beta} . \frac{1}{\beta} = 2 \implies 4\frac{1}{\beta} = 2 \implies \beta = 2$$

$$\mathbb{E}(X) = \frac{\alpha}{\beta} = 4 \implies \frac{\alpha}{2} = 4 \implies \alpha = 8$$

Hence (b) is correct and (a) is wrong.

Now

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \implies \frac{\alpha}{\beta^2} = \mathbb{E}(X^2) - [\frac{\alpha}{\beta}]^2 \implies \mathbb{E}(X^2) = \frac{\alpha}{\beta}(\frac{1+\alpha}{\beta})$$

Hence, (c) is correct.

$$\mathbb{E}(X^2) = \frac{8}{2}(\frac{1+8}{2}) = 18$$

Hence (d) is correct.

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

Solution: (b) is correct.

Let X is the life-span of the battery. The percentage of batteries that will last for more than 8 years is

$$\begin{array}{rcl} \mathbb{P}(X > 8) & \times & 100\% \\ \displaystyle\int_8^\infty f(x)dx & \times & 100\% \\ 1 - \displaystyle\int_0^8 f(x)dx & \times & 100\% \end{array}$$

(c) and (d) are not correct, because it is working on

$$\int xf(x)dx,$$

which is not correct to calculate the probability.

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

(a) 49

(b) 204

(c) 203

(d) 140

Solution: (d) is correct. There exist 49 1x1 squares, 36 2x2 squares and so on till 1 7x7 square. Therefore the total is $\sum_{k=1}^{7} k^2 = 1 + 2^2 + ... + 7^2 = 140$. ($\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$).

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 10:00am to 1:00pm 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

    Solution: (c) is correct. Suppose the first seat is occupied by a CS student. Then there are three choices for who this student could be. The second seat must be occupied by a DS student, and there are three choices for this. The third seat must be occupied by a CS student, and there are two choices for this. The fourth seat must be occupied by a DS student, and there are two choices for this. The last two occupants are fixed, so there is one choice for these. In total there are 3 * 3 * 2 * 2 = 36 ways of seating the students if the first seat is occupied by a CS student. Symmetrically, there are 36 (other) ways of seating the students if the first seat is occupied by a DS student. In total, there are 72 ways of seating these students.

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

    Solution: (b) is correct.

    (a) There are 36 outcomes (the first die can turn up anything from 1 to 6, and so can the second die) of which exactly one sums up to 12. So the first probability is 1/36.

    (b) (This requires writing out the outcomes which add up to an even number.) There are 18 outcomes which sum up to an even number (For any number that turns up on the first die, three numbers on the second die give an even sum.), of which exactly one sums up to 12. So the probability is 1/18.

    (c) (The list of outcomes from part 2 comes in handy here.) There are 14 outcomes which sum up to an even number larger than 4 (One outcome if the first die turns up 1, two outcomes each if the first die turns up 2 or 3, and three outcomes each if the first die turns up 4, 5, or 6.), of which exactly one sums up to 12. So the probability is 1/14.

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 $f(x) = \frac{x^5}{20} - \frac{x^4}{2} + 3x + 1$, 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.

   Solution: (b) is correct. The point of this question is to check if the concept of inflection point is understood; the given function will have one inflection point, but also points where second derivative is zero but is not an inflection point. Here $f'(x) = \frac{x^4}{4} - 2x^3 + 3$ and $f''(x) = x^2(x-6)$, so $f''(x)$ vanishes at 0 and 6; but only 6 is an inflection point since $f'''(0) = 0$ while $f'''(6) \neq 0$.

16. Which of the following are true?

   (a) $\frac{2019}{2020} < \frac{2020}{2021}$

   (b) $x + \frac{1}{x} \geq 2$ for all $x > 0$

   (c) $2^{60} > 5^{24}$

   (d) $2^{314} < 31^{42}$

   Solution: (a), (b), (c) are correct.
   (a) $2019 \times 2021 = 2020^2 - 1 < 2020^2$
   (b) Since $(x-1)^2 \geq 0$**.**
   (c) $2^{60} = (2^5)^{12} = 32^{12} > 25^{12} = (5^2)^{12} = 5^{24}$**.**
   (d) $2^{314} = 2^{210+104} = (2^5)^{42} \times 2^{104} = (32)^{42} \times 2^{104} > 31^{42}$.
   Note: This question had a minor typographical error in the actual exam paper. The correction has been taken into account for evaluating the question.

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 $-\frac{1}{2} < r < \frac{1}{2}$

   Solution: (d) is correct. If $|r| \geq \frac{1}{2}$ then the series on the right does not converge as the n^{th} term does not go to 0, while if $|r| < \frac{1}{2}$, this is a geometric series and the sum is given by the LHS.

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

Solution: (a) is correct.
The sum and product of roots of a quadratic polynomial $ax^2 + bx + c = 0$ are given by $-b/a$ and c/a respectively. In the given example, $a = 9, b = 171, c = -81$, so sum of roots equals $-171/9 = -19$ and product of roots equals $-81/9 = -9$.

Note: This question had a minor typographical error in the actual exam paper. This has been taken into account when evaluating the question.

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.

Solution: (b) and (d) are correct.
(This is a rewording of one of Lewis Carroll's problems.
Lets assign symbols
P: "is an owl"
Q: "is hopelessly ignorant of music"
R: "subscribes to Rahman's YouTube channel'
S: "really appreciates Rahman".
Then the given statements mean: S implies R, P implies Q and Q implies negation of R. So P implies Q, Q implies negation of R and negation of R implies negation of S. Thus we deduce: P implies negation of S, i.e. Owls don't really appreciate A.R. Rahman, this is (b).
Q implies negation of R and negation of R implies negation of S. So Q implies negation of S, which means S implies negation of Q, this is part (d).

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

Solution: (a) and (c) are true.
There are many ways to answer this. For example, using the second derivative test. Let $f(x) = e^x - 1 - x$ using the second derivative test we see that $x = 0$ is the global minimum. Hence, $f(x) \geq 0$ (in fact) for all real values of x. If one sets $g(x) = 1 + x - \ln(x)$ then $x = 1$ is the global minima and $g(x) \geq 2$ for $x > 0$. One could also appeal to the Taylor expansion; for $x \geq 0$ we have $e^x = 1 + x + \cdots$.

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;
}
```

Solution: This pseudo-code reverses a sentence word by word. For example, if the input string is "I love India", the output is "India love I".

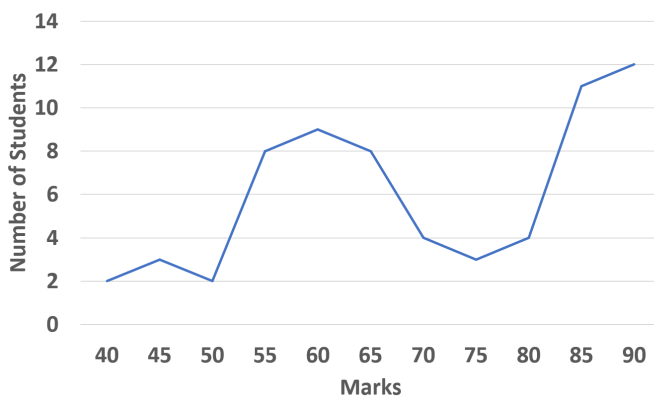
2. Consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Find A^n, in terms of n, for $n \geq 2$.

Solution: $A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $A^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $A^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. So the next powers of A will cycle through the above matrices. This means for $n \geq 1$,

$$A^{4n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A^{4n+1} = A$$

$$A^{4n+2} = A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } A^{4n+3} = A^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

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 $t = \frac{j*n}{100}$ and let k be an integer such that $k \leq t < (k+1)$ and let $s = t - k$. Then $u^* = x_{(k)} + s * (x_{(k+1)} - x_{(k)})$. Here, $x_{(n+1)}$ is defined to be $x_{(n)}$.

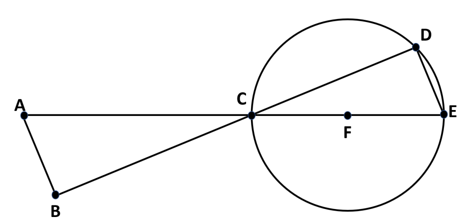


Based on the information presented in the graph, answer the following questions.

(a) Compute the 10th percentile of marks.
Solution: If the student fell into the 10th percentile, there must be 10% students scoring lesser marks. There are 66 students. So, to beat 6 students, the student should score 50 marks.

(b) Is the median score higher than the mean score?
Solution: Median falls at 70 and mean is 70.53. So, median is not higher than the mean.

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?



Solution: The area of the triangle ABC is 24. CAB and CED are similar triangles. Note that the triangle CED is a right triangle since the angle D spans the diameter. The diameter CE of the circle forms the hypotenuse of the triangle CED. Therefore, CD = 4. Since AB = 6 which is twice DE, BC scales up twice as CD as well. Therefore, AB = 6, BC = 8 and area of triangle ABC = $\frac{1}{2}(6)(8) = 24$.

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

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix},$$

e.g. if $n = 3$ then $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ 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:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 1 & 6 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 3 & 2 & 6 & 5 \end{pmatrix}.$$

Solution:

- $sgn(\sigma) = (-1)^{inv(\sigma)} = (-1)^4 = 1$**,**
- $sgn(\tau) = (-1)^{inv(\tau)} - (-1)^5 = -1$**,**
- $A_\sigma = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$
- $A_\tau = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$

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

Solution: $|A_\sigma| = 1$ **and** $|A_\tau| = -1$**. These can be found by observing that the matrices are obtained from the** 6×6 **identity matrix by row/column exchanges.** A_σ **requires even number of row exchanges, while** A_τ **requires odd number of row exchanges.**
In general, for any permutation σ**,**

$$sgn(\sigma) = |A_\sigma|.$$

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? **Solution: Apply Bayes' formula. Here**

   - **Let A be the event that an elderly Pandemic-20 patient in Gondwanaland survives the disease;**
   - **Let A' denote the complement of A; namely, the event that an elderly Pandemic-20 patient in Gondwanaland dies from the disease;**
   - **Let B be the event that an elderly Pandemic-20 patient in Gondwanaland was put on a ventilator as part of the treatment.**

   We have been told that

   - $P(A) = 0.80$
   - $P(A') = 0.20$
   - $P(B \mid A) = 0.6$
   - $P(B \mid A') = 0.1$

   And we want to find $P(A \mid B)$. Observe that $P(B) = P(B \mid A) \cdot P(A) + P(B \mid A') \cdot P(A') = 0.6 \cdot 0.8 + 0.1 \cdot 0.2 = 0.50$. Applying Bayes' formula we get that $P(A \mid B) = \frac{0.6 \cdot 0.8}{0.5} = 0.96$.

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.)
   Solution: 0.999^{1000}. The probability that any one mask is not effective is 0.001. The probability that any one mask is effective is thus $1 - 0.001 = 0.999$. The probability that the first 1000 (in fact, *any* 1000) masks are *all* effective is 0.999^{1000}.

   (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? **Solution: $1 - (0.999^{10^7})$. By a similar argument as for part (a), the probability that the first (or any set of) 10^7 masks are *all* effective is 0.999^{10^7}. So the probability that at least one of these masks is *not* effective is $1 - (0.999^{10^7})$.**

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.

Solution: At any point in time, the possible values for the number of games completed by any one player are $\{0, 1, 2, \ldots, 19\}$. Pick an arbitrary point t in time. If a player P_1 has completed 19 games at time t, then the number of games completed by any other player (say P_2) at time t cannot be 0. This is because P_1 must have completed one game with P_2 by time t, in order to have completed 19 games in total. So both 0 and 19 cannot appear *together* in the list of *actual* values for the number of games completed by all the players at time t.

Thus the count of players is 20, and the count of possible values for the number of games completed by all the players at time t is *at most* 19. It follows from the Pigeonhole Principle that there are *at least two players* who would have completed the *same* number of games in the tournament till time t.

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.

Solution:

(a) $R \Rightarrow P$ **(**R **implies** P**), or equivalently,** $\neg P \Rightarrow \neg R$ **(not(**P**) implies not(**R**))**

(b) $P \wedge (\neg Q) \wedge R$ **(**P **and not(**Q**) and** R**)**

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.

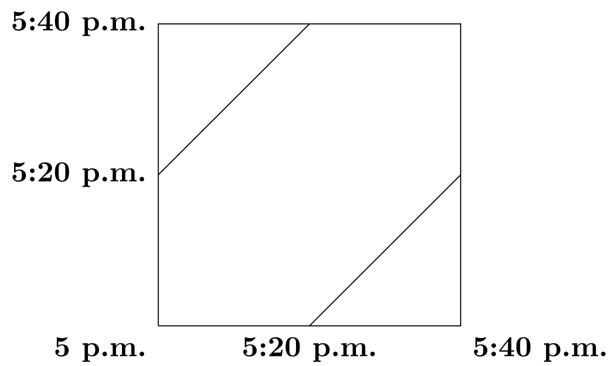
(b) What is the probability that Amar and Prem will meet? (Hint: Plot the arrival times on the x- and y-axes.)

Solution:

(a) i. The latest time by which any one of them can enter is $\underline{\ 5:40\ p.m.\ }$**;**

ii. If Amar and Prem are to meet, then their entry times can be separated by at most an interval of $\underline{\ 20\ }$ **minutes.**

(b) The sample space is the area of the square, each side of which represents 40 minutes. The event space is represented by the polygon in the square:



So the probability that the two will meet is the ratio of the two areas, which equals: $\frac{3}{4}$ **or** 0.75**. (Assuming that each side is** 2 **units, the area of the square is** 4 **units and the area of the polygon inside it is** $[4 - (\text{area of the two triangles})] = [4 - 1] = 3$ **units.)**

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 $b = 6$.

Solution: $f(1) = 4$, $f(6) = 364$ **and** $f'(c) = 6c^2 - 6c + 7$. **Then**

$$\begin{aligned} 6c^2 - 6c + 7 &= \frac{364 - 4}{6 - 1} = \frac{360}{5} = 72 \\ 6c^2 - 6c - 65 &= 0 \end{aligned}$$

$$c = \frac{6 \pm \sqrt{36 - 4.6.(-65)}}{2.6} = \frac{6 \pm \sqrt{1596}}{12} = \frac{6 \pm 39.95}{12} = 3.82 \quad and \quad -2.83$$

So the final solution is

$$c = \frac{6 + \sqrt{1596}}{12} = 3.82$$

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?
Solution: Number of symmetric binary relations on a set of size n is $2^{\frac{n(n+1)}{2}}$. Total number of binary relations are the relations that can be formed from the nine elements of the set, {(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)} which is 2^9. Therefore, the answer is $\frac{2^6}{2^9} = 1/8$.

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?
Solution: Peppa cannot have a knight mask, else her statement would be a lie which is a contradiction.
Peppa cannot be wearing an elf mask, for if that were true, then Suzy would be lying - hence Suzy would have a knight mask which means Emily would have the pirate mask, a contradiction since then her statement would have to be a lie so Emily would have to be a wearing an elf mask.
Thus, Peppa has to be wearing a pirate mask, which means Suzy is wearing an elf mask and Emily is wearing a knight mask.

15. For the function $f(x) = \ln(x)/x$, show that the maximum value of y occurs when $x = e$. Use this to show that $x^e < e^x$ for all positive values of x.
Solution: $f'(x) = \frac{1-\ln x}{x^2}$. **Equating** $f'(x)$ **to zero gives** $\ln x = 1$ **which is true iff** $x = e$. **Further this implies that for all** $x > 0$, $\frac{\ln x}{x} < f(e)$, **i.e.** $\frac{\ln x}{x} < \frac{\ln e}{e}$ **i.e.** $e \ln x < x$ **i.e** $\ln(x^e) < \ln(e^x)$. **Since** $\ln$ **is an injective function, this implies** $x^e < e^x$ **for** $x > 0$.

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`

Solution: *Previous permutation:* `bcjameflkihdg` *Next permutation:* `bcjamegdfhikl` *Explanation: The lexicographically smallest permutation of any subset of letters is in ascending order and the lexicographically largest is in descending order. To get the next permutation, we should disturb as small a suffix as possible. Look for a maximal suffix in descending order that cannot be incremented, in this case* ***`lkihgd`****. There is no further permutation whose suffix starts with* `f`*, so increment* `f` *to* `g` *and reorder the remaining letters in ascending order to get the next suffix* ***`gdfhikl`****. Reverse this reasoning to get the previous permutation.*

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`?
Solution: The contents of the array are sorted in descending order. This is called "pancake sort". Each iteration of the loop finds the position `p` of the largest element from `A[j]` to the end of the array and flips the segment from `A[j]` to `A[p]` to bring this largest element to position `j`.

(b) Suppose `size(A)` is `1000`. How many times is the test How many times is the test `A[i] > A[p]` executed?
Solution: 500,500 comparisons. For $j \in \{0, 1, \ldots, 999\}$, in iteration j, there are $1000 - j$ comparisons made. So in all there are $1000 + \cdots + 2 + 1 = (1000 \times 1001)/2 = 500500$ comparisons.

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?
    Solution: The answer is $2 * 6! = 1440$. There are $6!$ ways to seat the remaining students with B sitting to the left of A; similarly, there are $6!$ ways to seat the remaining students with B sitting to the right of A.

19. Let f be a continuous bijection from closed unit interval $[0,1]$ onto itself. (Recall: *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 $\alpha \in (f(c), f(d))$ is an intermediate value. Then there exists $y \in [a,b]$ such that $f(y) = \alpha$.)

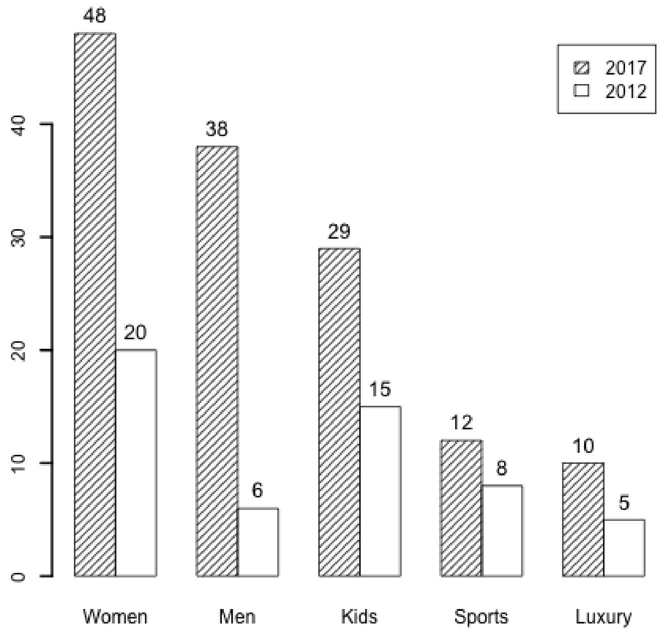
    (i) Show that $f(0)$ equals 0 or 1.
    Solution: (i) If $f(0) = a \in (0,1)$, pick $b, c \in [0,1]$ such that $b < a < c$. Let $x, y \in [0,1]$ be such that $f(x) = b$ and $f(y) = c$. Since $b \neq c$, we conclude $x \neq y$. Since f is continuous, intermediate value theorem will yield existence of z in the interval (x,y) (if $x < y$) or (y,x) such that $f(z) = a$. Since $z \neq 0$, this contradicts the assumption that f is a bijection.

    (ii) Show that $f(1)$ equals 0 or 1.
    Solution: (ii) follows similarly.

    (iii) Show that f admits a fixed point.
    Solution:(iii) if $f(0) = 0$ then nothing to prove. If $f(0) = 1$, we can conclude that $f(1) = 0$. Now take $g(x) = f(x) - x$. This is a continuous function such that $g(0) = 1$ and $g(1) = -1$ and hence intermediate value theorem will yield z such that $g(z) = 0$- a fixed point of f.

    (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.
    Solution: An example: $f(x) = 1 - x$ has unique fixed point $x = 0.5$. $f(x) = x$ has uncountably many fixed points!.

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?
Solution: Required percentage $= \frac{48-38}{38} \times 100 \approx 26.32\%$

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

-
- **Rate of increase in women's segment** $= \frac{48-20}{20} \times 100 \approx 140\%$ **;**
- **Rate of increase in men's segment** $= \frac{38-6}{6} \times 100 \approx 533.33\%$**,**
- **Rate of increase in kid's segment** $= \frac{29-15}{15} \times 100 \approx 93.33\%$**,**
- **Rate of increase in sports segment** $= \frac{12-8}{8} \times 100 \approx 50\%$**,**
- **Rate of increase in luxury segment** $= \frac{10-5}{5} \times 100 \approx 100\%$**.**
  So correct answer is sports.