Tutorial 2 - Presenting Of Data
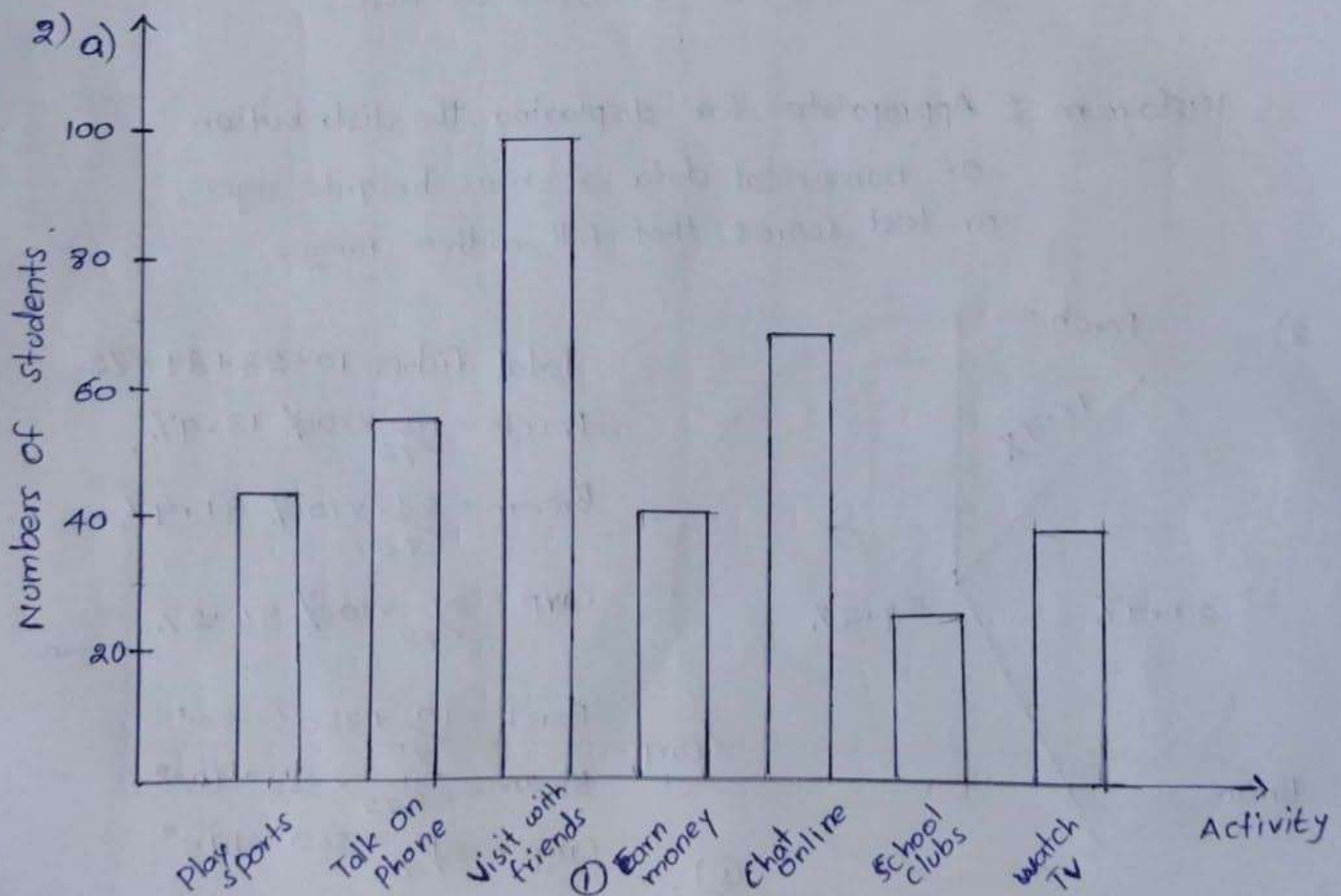
Assignment - 01

01) a) Population :- All 40 students enrolled in the course.

Sample : The 5 students who always sit in the back
of the classroom and were questioned
by Roger.

b) No, this is not likely to be a representative Sample.

Reason :- The sample only includes students who sit
in the back of the classroom, which may not
represent the options of all 40 students in
the class.

2) a)

b) Difference.

Bar chart : Used for categotical or discrete data.
The bars are separated, indicating distinct
Categories.

Histogram : Used for continuous data that has been
grouped into intervals (bins). The bars are
adjacent (no gaps) to show the continuity
of the data.

Appropriate Data Type.

Bar chatt : Appropriate for displaying the frequency of
different categories, like favourite after-
school activities or types of fish.

Histogram : Appropriate for displaying the distribution
of numerical data, such as heights, ages,
or test scores that fall within ranges.

3)

Pie chart labelled: Perch 13.9%, Bream 31.9%, Carp 54.2%
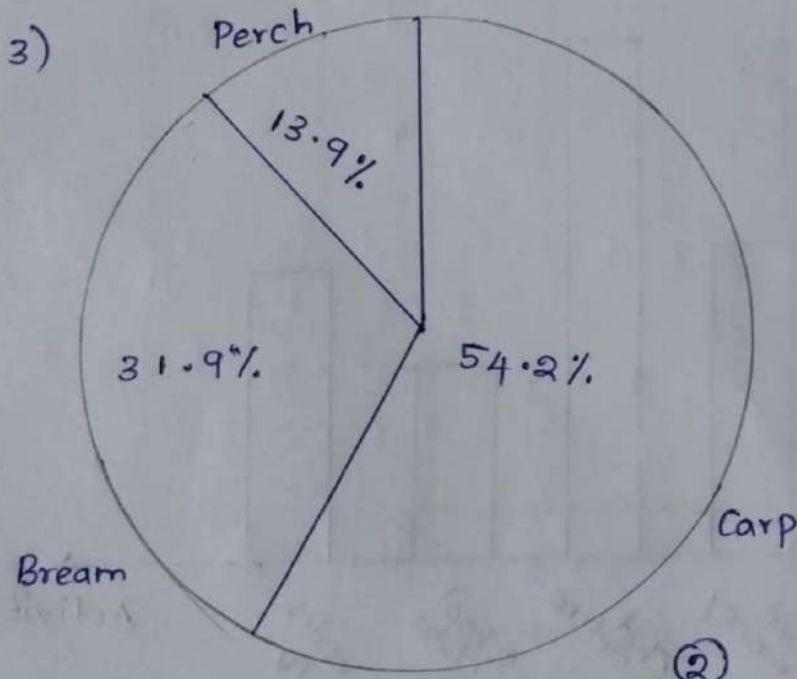
Total fishes = 10 + 23 + 39 = 72

Perch = $\frac{10}{72} \times 100\% = 13.9\%$

Bream = $\frac{23}{72} \times 100\% = 31.9\%$

Carp = $\frac{39}{72} \times 100\% = 54.2\%$

Perch = $\frac{10}{72} \times 360° = 50°$

Bream = $\frac{23}{72} \times 360° = 115°$

Carp = $\frac{39}{72} \times 360° = 195°$

②

4) a) Food
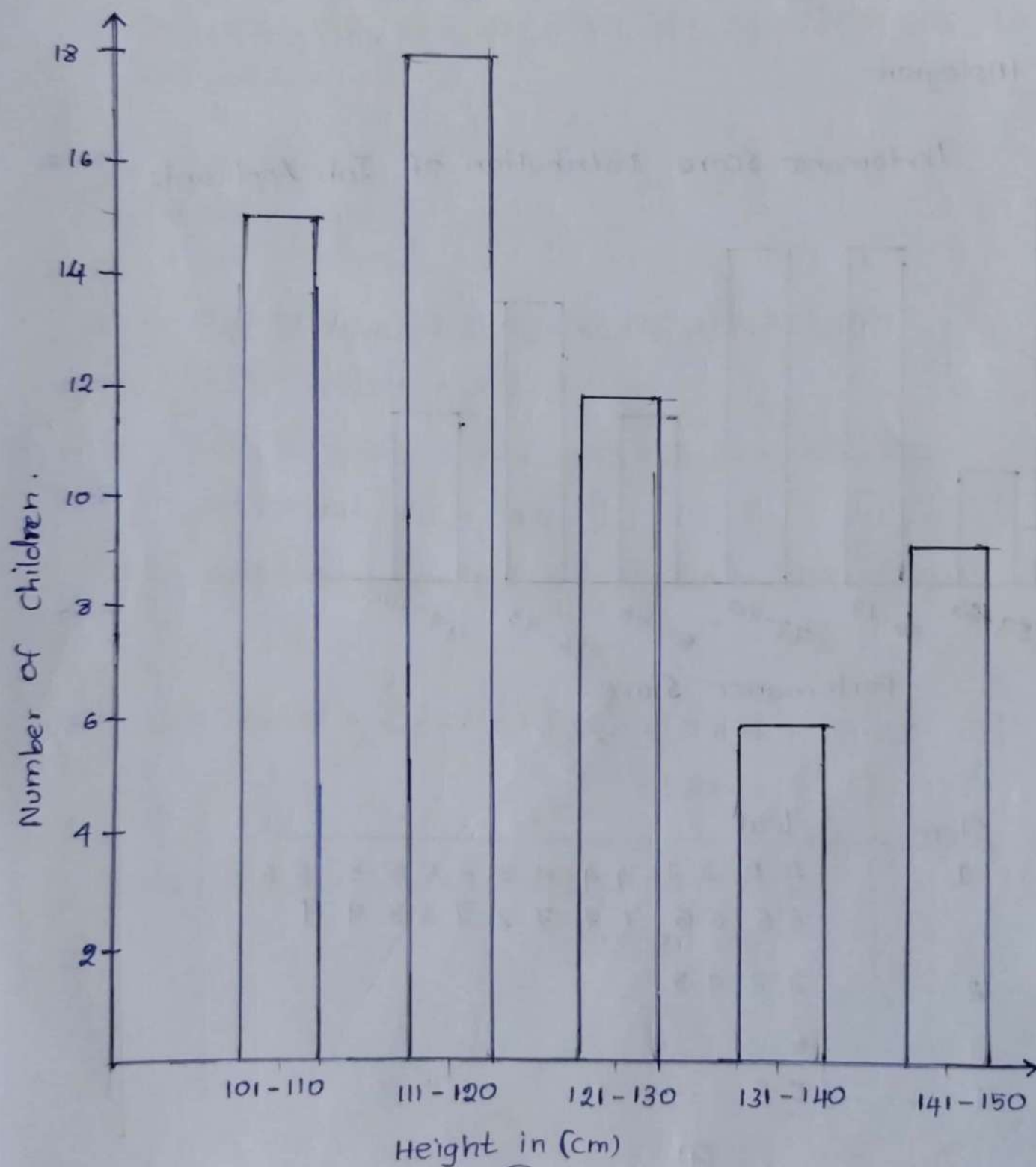
b) Petrol = 90°

Fraction on Petrol = $\frac{90}{360} = \frac{1}{4}$

c) Total money spent = \$ 25 × 4 = \$ 100.

5)

Height Frequency Distribution of Children

5) Frequency Distribution Table.

Score Range	Frequency
59 – 66	2
66 – 73	6
73 – 80	6
80 – 86	3
86 – 93	5
93 – 100	3

Histogram

Performance Score Distribution of Job Applicants.

Performance Score.

7) a)

Stem	Leaf
1	2 2 3 3 4 4 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 8 8 9 9
2	3 3 4 8
3	6
4	5 8

④

b) 34 People

c) yongest person - 12 years
   oldest person - 43 years.

d) 10 - 19

8) a) ascending order.
   308, 314, 316, 317, 321, 321, 321, 324, 325, 326, 326,
   327, 332.

Q_2 = 13 total
(Q_2) middle = 7th.
Q_2 = 321

Q_1 = First 6 values : 308, 314, 316, 317, 321, 321
Q_1 = $(316 + 317)/2$ = 316.5

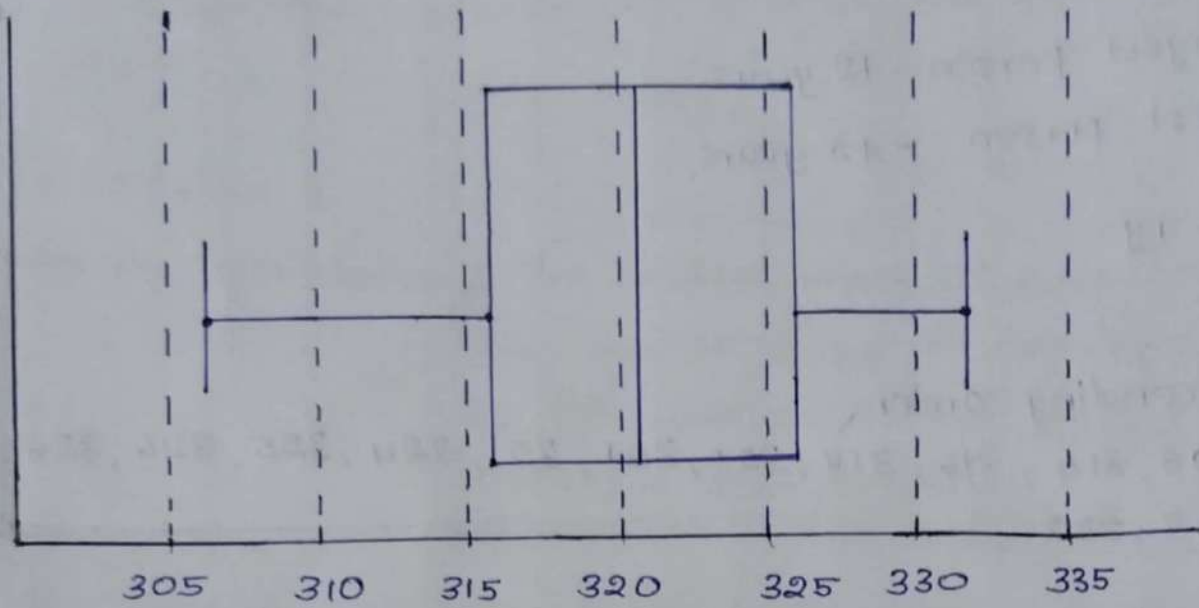Q_3 = last 6 values :- 324, 325, 326, 326, 327, 332
Q_3 = $(326 + 326)/2$ = 326.

Interquartile Range (IQR) = $Q_3 - Q_1$ = 326 - 316.5

= 9.5

Lower Bound = $Q_1 - 1.5 \times IQR$ = 316.5 - 14.25

= 302.5.

Upper Bound = $Q_3 + 1.5 \times IQR$ = 326 + 14.25

= 340.25.

⑤

a)

Boxplot of Crop Harvested Areas.

305   310   315   320   325   330   335

Crop Harvested Area.

(millions of Acres)

b)

Tourists Arrivals to Ceylon.
(2001 - 2005)

Legend:
- Total
- other
- Business
- Pleasure

year
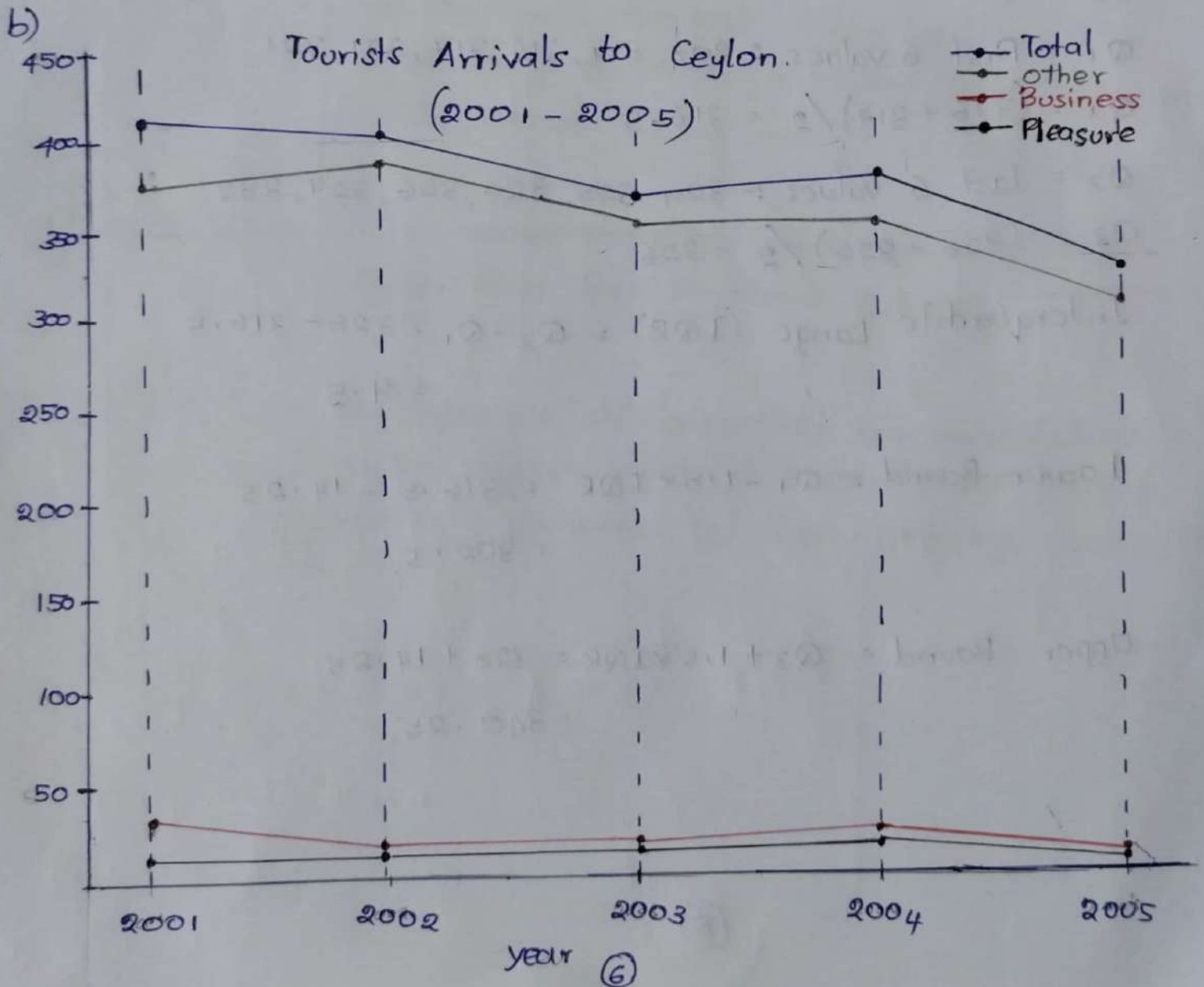
9) a. Yes, we can conclude that there is a difference between the salaries of women and men in this Plant

Justification :-

• Median Difference :- the median salary of men is higher the median salary of women as shown by the position of the central line (median) in each boxplot.

• Box Overlap :- The boxes themselves represent the interquartile range (IQR), Containing the middle 50% of the data. While there is some Overlap in the ranges of the boxes, the male salary box is shifted Significantly to the right (Higher salaries) Compared to the female salary box.

• Overall Range :- The overall range of male salaries extends higher than that of female salaries.

• Outliers :- The presence of an outlier for male salaries a much higher value ($73000-$79000) Further Suggests a potential for higher earnings among men.

b. Outlier threshold = $Q_3 + 1.5 \times IQR$.

For Male Salary

$Q_1 = 35000$

$Q_3 = 47000$

$IQR = Q_3 - Q_1 = 47000 - 35000 = 12000$

Outlier threshold $(Q_3 + 1.5 \, IQR) = 47000 + 1.5 \times 12000$

$= 65000$.

For Female salary

$Q_1 = 30,000$

$Q_3 = 42000$

$IQR = Q_3 - Q_1 = 42000 - 30000 = 12000$

Outlier threshold $(Q_3 + 1.5 \times IQR) = 42000 + 1.5 \times 12000$

$= 60000$.

- For Male-Salary, a salary must be greater than approximately 60000 to be an outlier on the high side.

How many Outliers:

For Male-Salary, there is clearly one outlier marked as a purple square. Which is significantly higher than the upper whisker, appearing to be arround 78000-79000.

For Female-Salary, there are no visible outliers on the high side. The upper whisker extends to roughly 55000-57000, and there are no individual points beyond it. The

Therefore, there is a total of one outlier in these data.

③

C. For Male_salary :-

    * The median is slightly closer to the lower quartile (Q_1) than to the upper quartile (Q_3).

    * The upper whisker is longer than the lower whisker (excluding the outlier)

    * The presence of a high outlier suggests a long tail on the right.

    * Conclusion - The male salary distribution appears to be positively skewed (right-skewed). This means there are few individuals earning significantly higher salaries, pulling the mean to the right of the median.

For Female_Salary :-

    * The median appears to be roughly in the middle of the box or slightly closer to the upper quartile (Q_3)

    * The lower whisker appears to be slightly longer than the upper whisker.

    * There are no outliers on the high side.

    * Conclusion :- The female salary distribution appears to be approximately symmetric or slightly negatively skewed (left-skewed). While it's difficult to be definitive without more detail, the box and whisker lengths don't strongly suggest positive skewness like the male distribution. If anything, the slightly longer lower whisker might him at a very mild left skew, but for practical purpose it could be considered reasonably symmetric.

10) a.

class Boundaries	Frequency I	Frequency II
E : (0.05 - 29.5)	8	8
D : (29.5 - 39.5)	4	4
C : (39.5 - 54.5)	3	3
B : (54.5 - 69.5)	1	0
A : (69.5 - 99.5)	4	5

b. Cumulative frequency I

E : 8
D : 8 + 4 = 12
C : 12 + 3 = 15
B : 15 + 1 = 16
A : 16 + 4 = 20

Cumulative Frequency II

E : 8
D : 8 + 4 = 12
C : 12 + 3 = 15
B : 15 + 0 = 15
A : 15 + 5 = 20

Cumulative Frequency table.

class Boundaries	Cumulative frequency I	Cumulative frequency II
E : (0.05 - 29.5)	8	8
D : (29.5 - 39.5)	12	12
C : (39.5 - 54.5)	15	15
B : (54.5 - 69.5)	16	15
A : (69.5 - 99.5)	20	20

⑩

Graph showing Cumulative frequency I (solid line) and Cumulative frequency II (dashed line) plotted against Percentage Marks (Class Boundaries). Y-axis: Cumulative frequency (0 to 20). X-axis values: 0.05, 29.5, 39.5, 54.5, 69.5, 99.5.

Percentage Marks.
(Class Boundaries)

⑪

Differences
─────
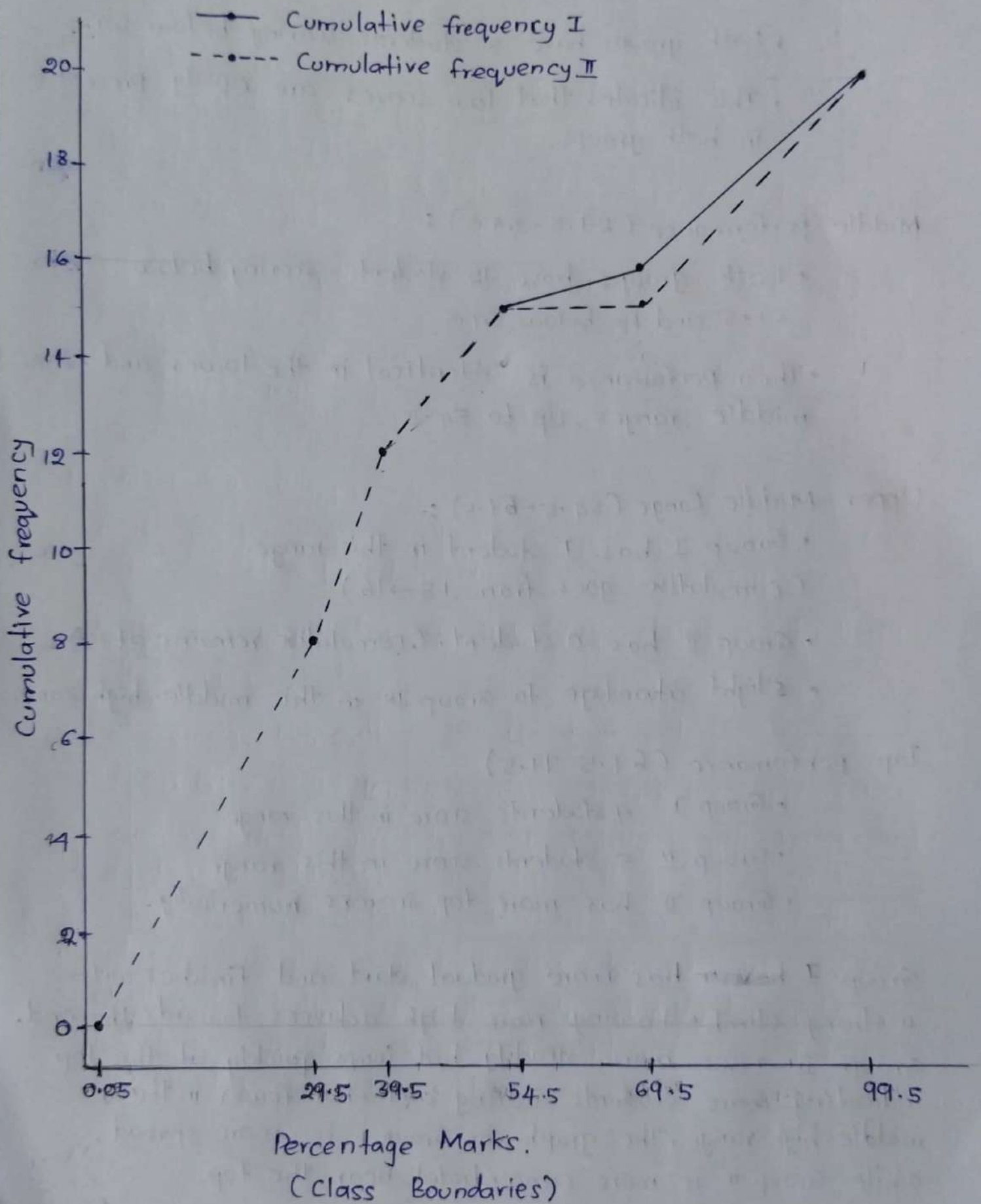
Initial Performance (0.05 – 29.5) :-

    • Both groups have 8 students scoring below 29.5.

    • This idicates that low scores are equally present in both groups.

Middle Performance (29.5 – 54.5) :-

    • Both groups have 12 students scoring below 39.5 and 15 below 54.5.

    • Their perfomance is indentical in the lower and middle ranges, up to 54.5.

Upper - Middle Range (54.5 – 69.5) :-
    • Group I has 1 student in this range.
    (cumulative goes from 15→16)

    • Group II has 0 student. (cumulative remains at 15)
    • slight advantage to Group II in this middle-high zone.

Top perfomance (69.5 – 99.5)
    • Group I 4 students score in this range
    • Group II 5 students score in this range.
    • Group II has more top scorers numerically.

Group I has more gradual start and finishes with a strong climb, showing more high achivers towards the end. Group II rises more steadily but jumps quickly at the top, indicating more students scoring well, but fewer in the middle-high range. The graph for Group I is more spread, while Group II is more concentrated near the top.