Assignment #01

Name: M. Imran Butt

Reg no: L1F24BSCSO546

Course: Probability & Statistic

Section: C8

Question #1

(a) Name the graph

The graph is "Histogram".

The graph is used to represent the frequency distribution of continuous numerical data (weights of Athletes). The weights are grouped into continuous intervals of equal size on the x-axis, and the height of each bar (the y-axis) represents the frequency and the bars are adjacent to each other to show that data is continuous. So graph is "Histogram."

(b) Average weight of athletes?

we know that to find the average (mean) of a grouped data we use this formula:

$$= \frac{\sum(f \cdot f_m)}{\sum f}$$

Class Intervals	f	f_m	$f \cdot g_m$
30—37	5	33.5	167.5
37—44	5	40.5	202.5
44—51	70	47.5	3325
51—58	220	54.5	11990
58—65	300	61.5	18450
65—72	290	68.5	19865
72—79	250	75.5	18875
79—86	200	82.5	16500
86—93	105	89.5	9397.5
93—100	80	96.5	7720
100—107	30	103.5	3105
107—114	10	110.5	1105
114—121	15	117.5	1762.5
121—128	5	124.5	622.5
128—135	5	131.5	657.5
135—142	3	138.5	415.5
142—149	2	145.5	291
149—156	0	152.5	0
156—163	2	159.5	319
163—170	0	166.5	0
170—177	0	173.5	0
177—184	0	180.5	0
184—191	0	187.5	0
191—198	0	194.5	0
198—205	0	201.5	0
205—212	0	208.5	0
212—219	2	215.5	431

$\Sigma f = 1599$

$\Sigma f \cdot f_m = 115201.5$

So $A \cdot W = \dfrac{115201.5}{1599} \approx 72.04$

So Average weights of athletes is 72.04 Kg.

(c) Most of weight exist within what interval?

By looking at the graph I can see that most of weight exist within the interval "(52, 65]".

(d) estimated sample size?

The estimated sample size is the total number of athletes. So In part (b) we did find the Σf which represents the total number of athletes. so the estimated sample size is **1599.**

(e) Shape of distribution?

The distribution is positively skewed (right-skewed).

$\longleftrightarrow$

Question # 2

(a) Name the graph?

The graph is a Frequency Polygon.

As it represents the frequency on y-axis and midpoints of the class intervals on x-axis.

(b) total Students of both sections:

orange line: $0 + 3 + 0 + 0 + 1 + 2 + 4 + 13 + 7 +$
$7 + 10 + 0 = 47$

Blue line: $1 + 0 + 10 + 10^{+12} + 7 + 4 + 0$
$= 44$

Total Students: 91

(C) which section show stable result

Section 1 blue shows relatively more stable results.

The blue line frequencies change more smoothly, no sharp sudden jumps as compared to orange so that's why blue is more stable.

(d) minimum and maximum marks of both sections?

→ Series 1 (Blue):
Estimated Minimum: The lowest non-zero midpoint is 10. The interval is 5 to 15 marks

Estimated Maximum: The highest non-zero midpoint is 90. The interval is 85 to 95 marks.

→ Series 2 (Orange):
Estimated Minimum: The lowest non-zero midpoint is 10. The interval is 5 to 15 marks.

Estimated Maximum: The highest non-zero midpoints is 100. The interval is 95 to 105 marks.

Question # 3

data: 62, 72, 66, 79, 83, 61, 62, 85, 72, 64, 74, 71, 42, 38, 91, 66, 77, 90, 74, 63, 64, 68, 42

Sorted data: 38, 42, 42, 61, 62, 62, 63, 64, 64, 66, 66, 68, 71, 72, 72, 74, 74, 77, 79, 83, 85, 90, 91

(a) Find mean, Median, Mode

$$\text{mean} = \bar{x} = \frac{\Sigma x_i}{n}$$

$$\bar{x} = \frac{38+42+42+61+62+62+63+64+64+66+66+68+71+72+72+74+74+77+79+83+85+90+91}{23}$$

$$\bar{x} = \frac{1566}{23}$$

$$\boxed{\bar{x} = 68.08} \checkmark$$

$$\boxed{\text{Median} = 68} \checkmark$$

$$\boxed{\text{Mode} = 42, 62, 64, 66, 72 \text{ and } 74} \checkmark$$

(b) Which measure of central tendency d escribes best?

The Median (68) because it represents the central value more accurately when there are outliers or multiple modes.

(c) Check outliers:

$Q_1 = 62$

$Q_3 = 77$

$IQR = Q_3 - Q_1 = 77 - 62$

$$\boxed{IQR = 15}$$

Outlier boundaries:

for lower $= Q_1 - 1.5(IQR)$
$= 62 - 1.5(15)$

lower $= 39.5$

for upper $= Q_3 + 1.5(IQR)$
$= 77 + 1.5(15)$

upper $= 99.5$

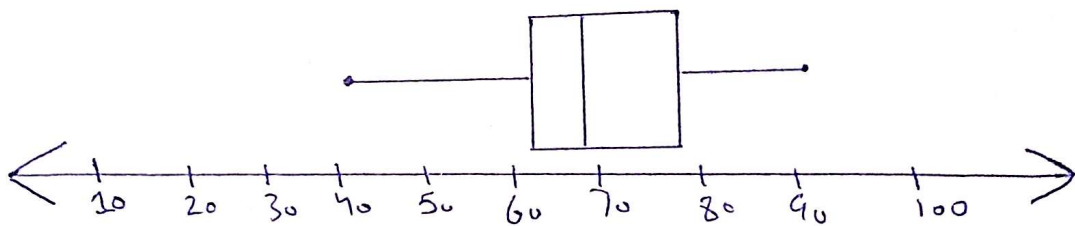So the value 38 is lower than 39.5 and no values above 99.5, So

$$\boxed{Outlier: 38}$$

(d) Make Whisker plot and discuss data distribution:

$Q_1 = 62$, $Q_2 = 68$ and $Q_3 = 77$

as 38 is the outlier lower than 39.5 we'll consider small whisker the next smaller value but not an outlier. So that value is 42 so

smallest = 42, largest = 91

→ Whisker plot:

→ Discussion of data distribution:

→ The meadian (68) lies closer to the lower quartile (62) than to the upper quartile (77), which means more than half of the data are slightly concentrated toward the lower scores.

→ The upper whisker (77 – 91) is ~~longer~~ smaller than the lower whisker (42 – 62), so the data is left - skewed (negative skewed)

→ There's one outlier (38) below the lower whisker.

→ The IQR = 15 means the middle 50% of score lie b/w 62 and 77.