<u>Bio Statistics</u>

<u>Solution: Question. (01)</u>

a) Equal number in the four groups, so the overall mean is

(204 + 259 + 266 + 317) / 4 = 261.5 for men (by calculator)

(178 + 235 + 266 + 304) / 4 = 245.75 for women

The SDs are 0.9 × √1308 = 32.5 for men, and 0.8 × √1540 = 31.4 for women

The overall mean is (1308 × 261.5 + 1540 × 245.75) / 2848 = 253.0 (2) (by calculator)

(Award 1 for an attempt at weighted mean that result in the wrong answer)

b) Milk consumption is very low for both men and women in Q1 and Q2

But it rises sharply in Q3 and again in Q4.

So those who eat most fresh vegetables consume much more milk than those who eat less fresh vegetables. (1)

c) Rice consumption falls, for men and for women, as fresh vegetable consumption rises

It is the only food group to show this pattern.

d) Parallel bar chart, back-to-back bar charts or simple line graphs to compare men and women.

(Charts must be appropriate (1), accurate (1), labeled (1) and well executed (1))

e) Divide men's figures by 1.2 (or multiply women's figures by 1.2) (by calculator)

f)

Comparison Table

Fresh vegetable	170	178
fruit	26	28
rice	306	315
Wheat flour	66	56
meat	58	48
fish	19	19

No very large differences in patterns of consumption (1)

But men eat more meat and wheat flour (1)

While women eat more rice and fresh vegetables (1)

Solution: Question. (02)

a) **The purpose of census:** The purpose of a census is to enumerate, and collect data on, every member of a population

The census is a snapshot whereas various administrative records span continuous time and would have to be analyzed, with some difficulty, to get a spot figure for a particular date.

b) It differs from a sample survey in that, by definition, a sample survey does not attempt to reach the whole population

A census will differ from records held by government departments in that it aims to be complete whereas government department records will not be

Also it addresses particular questions (e.g. language spoken, religion) which government records are unlikely to contain

And participation in the census is a legal requirement

c) A participation rate of 94% is high

And to that extent might be regarded as giving very good information when compared with other data

However, the nature of the missing 6% is an issue. These people are likely to be untypical of the 94% who participate

E.g. estimating the homeless rate from the 94% reached would be very inaccurate

d) Since ‘Jedi Knight’ is not in any real sense a religion

This indicates that people do not always take the census seriously

This may therefore cast doubt on the accuracy of other responses they give

It may also indicate contempt for or a distrust of, government and the collection of data by government agencies

While this example indicates that not all responses can be taken seriously, there may still be value in asking the question

For example, the 2011 census quantified the decline in Christianity and the rise in Islam: these are likely to be real phenomena

e) Conducting the census online in 2021 will present problems for excluded groups, those without internet access or with limited internet capability such as those in poverty and the old

So enumerators will still need to be used to reach these groups

It may also be more difficult to persuade people to complete the census by going online than it is to persuade them to fill in a printed form

f) There may be additional concerns about security of information when it is supplied online.)

Additional information held by government agencies is unlikely to be complete

Record matching in combining databases is a notorious problem

Solution: Question. (03)

a) A.M,G.M,H.M,Median,Mode,Quartiles,Deciles,Percentiles,Range,M.D,Q.D,Variance,Standared Deviation, Coefficient of variation,Skewness for the following data.

Class	Frequency
20 - 24	1
25 - 29	3
30 - 34	5
35 - 39	8
40 - 44	5
45 - 49	2
50 - 54	0
55 - 59	1

Class (1)	Frequency (f) (2)	Mid value (x) (3)	$d = \frac{x - A}{h} = \frac{x - 42}{5}$ $A = 42, h = 5$ (4)	$f \cdot d$ (5) = (2) × (4)	$f \cdot d^2$ (6) = (5) × (4)	cf (7)
20 - 24	1	22	-4	-4	16	1
25 - 29	3	27	-3	-9	27	4
30 - 34	5	32	-2	-10	20	9
35 - 39	8	37	-1	-8	8	17
40 - 44	5	42=A	0	0	0	22
45 - 49	2	47	1	2	2	24
50 - 54	0	52	2	0	0	24
55 - 59	1	57	3	3	9	25
---	---	---	---	---	---	---
	$n = 25$	-----	-----	$\sum f \cdot d = -26$	$\sum f \cdot d^2 = 82$	-----

Mean $\bar{x} = A + \frac{\sum fd}{n} \cdot h$

$= 42 + - \frac{26}{25} \cdot 5$

$= 42 + - 1.04 \cdot 5$

$= 42 + - 5.2$

$= 36.8$

To find Median Class

= value of $\left(\frac{n}{2}\right)^{th}$ observation

= value of $\left(\frac{25}{2}\right)^{th}$ observation

= value of 12^{th} observation

From the column of cumulative frequency cf, we find that the 12^{th} observation lies in the class $35 - 39$.

$\therefore$ The median class is $34.5 - 39.5$.

Now,

$\therefore L$ = lower boundary point of median class $= 34.5$

$\therefore n$ = Total frequency $= 25$

$\therefore cf$ = Cumulative frequency of the class preceding the median class $= 9$

$\therefore f$ = Frequency of the median class $= 8$

$\therefore c$ = class length of median class $= 5$

Median $M = L + \frac{\frac{n}{2} - cf}{f} \cdot c$

$= 34.5 + \frac{12.5 - 9}{8} \cdot 5$

$= 34.5 + \frac{3.5}{8} \cdot 5$

$= 34.5 + 2.1875$

$= 36.6875$

To find Mode Class
Here, maximum frequency is 8.

$\therefore$ The mode class is 34.5 - 39.5.

$\therefore L$ = lower boundary point of mode class = 34.5

$\therefore f_1$ = frequency of the mode class = 8

$\therefore f_0$ = frequency of the preceding class = 5

$\therefore f_2$ = frequency of the succedding class = 5

$\therefore c$ = class length of mode class = 5

$$Z = L + \left(\frac{f_1 - f_0}{2 \cdot f_1 - f_0 - f_2} \right) \cdot c$$

$$= 34.5 + \left(\frac{8 - 5}{2 \cdot 8 - 5 - 5} \right) \cdot 5$$

$$= 34.5 + \left(\frac{3}{6} \right) \cdot 5$$

$$= 34.5 + 2.5$$

Sample Variance $S^2 = \left(\frac{\sum f \cdot d^2 - \frac{(\sum f \cdot d)^2}{n}}{n - 1} \right) \cdot h^2$

$$= \left(\frac{82 - \frac{(-26)^2}{25}}{24} \right) \cdot 5^2$$

$$= \left(\frac{82 - 27.04}{24} \right) \cdot 25$$

$$= \frac{54.96}{24} \cdot 25$$

$$= 2.29 \cdot 25$$

$$= 57.25$$

Sample Standard deviation $S = \sqrt{\frac{\sum f \cdot d^2 - \frac{(\sum f \cdot d)^2}{n}}{n - 1}} \cdot h$

$$= \sqrt{\frac{82 - \frac{(-26)^2}{25}}{24}} \cdot 5$$

$$= \sqrt{\frac{82 - 27.04}{24}} \cdot 5$$

$$= \sqrt{\frac{54.96}{24}} \cdot 5$$

$$= \sqrt{2.29} \cdot 5$$

$$= 1.5133 \cdot 5$$

$$= 7.5664$$

Co-efficient of Variation (Sample) $= \frac{S}{\bar{x}} \cdot 100\ \%$

$= \frac{7.5664}{36.8} \cdot 100\ \%$

$= 20.56\ \%$

Here, $n = 25$

Q_3 class :

Class with $\left(\frac{3n}{4}\right)^{th}$ value of the observation in cf column

$= \left(\frac{3 \cdot 25}{4}\right)^{th}$ value of the observation in cf column

$= (18.75)^{th}$ value of the observation in cf column

and it lies in the class 40 - 44.

$\therefore Q_3$ class : 39.5 - 44.5

The lower boundary point of 39.5 - 44.5 is 39.5.

$\therefore L = 39.5$

$$Q_3 = L + \frac{\frac{3n}{4} - cf}{f} \cdot c$$

$= 39.5 + \frac{18.75 - 17}{5} \cdot 5$

$= 39.5 + \frac{1.75}{5} \cdot 5$

$= 39.5 + 1.75$

$= 41.25$

D_7 class :

Class with $\left(\frac{7n}{10}\right)^{th}$ value of the observation in cf column

$= \left(\frac{7 \cdot 25}{10}\right)^{th}$ value of the observation in cf column

$= (17.5)^{th}$ value of the observation in cf column

and it lies in the class 40 - 44.

$\therefore D_7$ class : 39.5 - 44.5

The lower boundary point of 39.5 - 44.5 is 39.5.

$\therefore L = 39.5$

$$D_7 = L + \frac{\frac{7n}{10} - cf}{f} \cdot c$$

$$= 39.5 + \frac{17.5 - 17}{5} \cdot 5$$

$$= 39.5 + \frac{0.5}{5} \cdot 5$$

$$= 39.5 + 0.5$$

$$= 40$$

P_{20} class :

Class with $\left(\frac{20n}{100}\right)^{th}$ value of the observation in cf column

$= \left(\frac{20 \cdot 25}{100}\right)^{th}$ value of the observation in cf column

$= (5)^{th}$ value of the observation in cf column

and it lies in the class 30 - 34.

$\therefore P_{20}$ class : 29.5 - 34.5

The lower boundary point of 29.5 - 34.5 is 29.5.

$\therefore L = 29.5$

$$P_{20} = L + \frac{\frac{20n}{100} - cf}{f} \cdot c$$

$$= 29.5 + \frac{5 - 4}{5} \cdot 5$$

$$= 29.5 + \frac{1}{5} \cdot 5$$

$$= 29.5 + 1$$

$$= 30.5$$

Skewness :

Mean $\bar{x} = \frac{\sum fx}{\sum f}$

$$= \frac{920}{25}$$

$$= 36.8$$

Class (1)	Mid value (x) (2)	f (3)	$f \cdot x$ (4) = (2) × (3)	$(x - \bar{x})$ (5)	$f \cdot (x - \bar{x})^2$ (6) = (3) × (5)	$f \cdot (x - \bar{x})^3$ (7) = (5) × (6)
20 - 24	22	1	22	-14.8	219.04	-3241.792
25 - 29	27	3	81	-9.8	288.12	-2823.576
30 - 34	32	5	160	-4.8	115.2	-552.96
35 - 39	37	8	296	0.2	0.32	0.064
40 - 44	42	5	210	5.2	135.2	703.04
45 - 49	47	2	94	10.2	208.08	2122.416
50 - 54	52	0	0	15.2	0	0
55 - 59	57	1	57	20.2	408.04	8242.408
---	---	---	---	---	---	---
--	--	$n = 25$	$\sum f \cdot x = 920$	`--	= 1374	= 4449.6

Sample Standard deviation $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

$= \sqrt{\frac{1374}{24}}$

$= \sqrt{57.25}$

$= 7.5664$

Sample Skewness $= \frac{\sum (x - \bar{x})^3}{(n - 1) \cdot S^3}$

$= \frac{4449.6}{24 \cdot (7.5664)^3}$

$= \frac{4449.6}{24 \cdot 433.1749}$

$= 0.428$

Geometric mean,Harmonic mean :

Class	Mid value (x) (2)	f	$f\log(x)$	$\frac{f}{x}$
20 - 24	22	1	3.091	0.0455
25 - 29	27	3	9.8875	0.1111
30 - 34	32	5	17.3287	0.1562
35 - 39	37	8	28.8873	0.2162
40 - 44	42	5	18.6883	0.119
45 - 49	47	2	7.7003	0.0426
50 - 54	52	0	0	0
55 - 59	57	1	4.0431	0.0175
---	---	---	---	---
--	--	$n = 25$	$\sum f\log(x) = 89.6263$	$\sum \left(\frac{f}{x}\right) = 0.7082$

GM of X $= Antilog\left(\frac{\sum f\log(x)}{n}\right)$

GM of X $= Antilog\left(\frac{\sum f\log(x)}{n}\right)$

$= Antilog\left(\frac{89.6263}{25}\right)$

$= Antilog(3.5851)$

$= 36.0552$

HM of X $= \frac{n}{\sum\left(\frac{f}{x}\right)}$

$= \frac{25}{0.7082}$

$= 35.3019$

Mean deviation :

Mean $\bar{x} = \frac{\sum fx}{\sum f}$

$= \frac{920}{25}$

$= 36.8$

Class (1)	f (2)	Mid value (x) (3)	$f \cdot x$ (4) = (2) × (3)	$\lvert x - \bar{x}\rvert = \lvert x - 36.8\rvert$ (5)	$f \cdot \lvert x - \bar{x}\rvert$ (6) = (2) × (5)
20 - 24	1	22	22	14.8	14.8
25 - 29	3	27	81	9.8	29.4
30 - 34	5	32	160	4.8	24
35 - 39	8	37	296	0.2	1.6
40 - 44	5	42	210	5.2	26
45 - 49	2	47	94	10.2	20.4
50 - 54	0	52	0	15.2	0
55 - 59	1	57	57	20.2	20.2
---	---	---	---	---	---
--	$n = 25$	--	$\sum f \cdot x = 920$	--	$\sum f \cdot \lvert x - \bar{x}\rvert = 136.4$

Mean deviation of Mean

$\delta\bar{x} = \frac{\sum f \cdot \lvert x - \bar{x}\rvert}{n}$

$\delta\bar{x} = \frac{136.4}{25}$

$\delta\bar{x} = 5.456$

Co-efficient of Mean deviation $= \frac{\delta\bar{x}}{\bar{x}}$

$= \frac{5.456}{36.8}$

$= 0.1483$

Quartile deviation :

Class	Frequency f	cf
20 - 24	1	1
25 - 29	3	4
30 - 34	5	9
35 - 39	8	17
40 - 44	5	22
45 - 49	2	24
50 - 54	0	24
55 - 59	1	25
---	---	---
	n = 25	--

Here, $n = 25$

Q_1 class :

Class with $\left(\frac{n}{4}\right)^{th}$ value of the observation in cf column

$= \left(\frac{25}{4}\right)^{th}$ value of the observation in cf column

$= (6.25)^{th}$ value of the observation in cf column

and it lies in the class 30 - 34.

$\therefore Q_1$ class : 29.5 - 34.5

The lower boundary point of 29.5 - 34.5 is 29.5.

$\therefore L = 29.5$

$$Q_1 = L + \frac{\frac{n}{4} - cf}{f} \cdot c$$

$$= 29.5 + \frac{6.25 - 4}{5} \cdot 5$$

$$= 29.5 + \frac{2.25}{5} \cdot 5$$

$$= 29.5 + 2.25$$

$$= 31.75$$

Q_3 class :

Class with $\left(\frac{3n}{4}\right)^{th}$ value of the observation in cf column

$= \left(\frac{3 \cdot 25}{4}\right)^{th}$ value of the observation in cf column

$= (18.75)^{th}$ value of the observation in cf column

and it lies in the class 40 - 44.

$\therefore Q_3$ class : 39.5 - 44.5

The lower boundary point of 39.5 - 44.5 is 39.5.

$\therefore L = 39.5$

$$Q_3 = L + \frac{\frac{3n}{4} - cf}{f} \cdot c$$

$$= 39.5 + \frac{18.75 - 17}{5} \cdot 5$$

$$= 39.5 + \frac{1.75}{5} \cdot 5$$

$$= 39.5 + 1.75$$

$$= 41.25$$

Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{41.25 - 31.75}{2} = \frac{9.5}{2} = 4.75$

Coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{41.25 - 31.75}{41.25 + 31.75} = \frac{9.5}{73} = 0.1301$

b) Convert the above given data in the form of ungrouped and then find A.M,G.M,H.M,Median,Mode,Quartiles,Deciles,Percentiles,Range,M.D,Q.D,Variance,Standar ed Deviation, Coefficient of variation, Skewness for the converted data.

Solution:

`x`	`dx = x - A = x - 39`	`dx^2`
59	20	400
49	10	100
39	0	0
29	-10	100
20	-19	361
---	---	---
`sum x=196`	`sum (dx)=1`	`sum (dx)^2=961`

Mean `bar x = (sum x)/n`

`=(59 + 49 + 39 + 29 + 20)/5`

`=196/5`

`=39.2`

`bar x = 39.2` is not an integer, use assumed mean method

`A = 39`

Median :
Observations in the ascending order are :
`20, 29, 39, 49, 59 `

Here, `n=5` is odd.

`M=` value of `((n+1)/2)^(th)` observation

`=` value of `((5+1)/2)^(th)` observation

`=` value of `3^(rd)` observation

`=39`

Mode :
In the given data, no observation occurs more than once.
Hence the mode of the observations does not exist, means mode=0.

Sample Variance `S^2 = (sum dx^2 - (sum dx)^2/n)/(n-1)`

`=(961 - (1)^2/5)/4`

`=(961 - 0.2)/4`

`=960.8/4`

`=240.2`

Sample Standard deviation $S = \sqrt{\dfrac{\sum dx^2 - \dfrac{(\sum dx)^2}{n}}{n - 1}}$

$= \sqrt{\dfrac{961 - \dfrac{(1)^2}{5}}{4}}$

$= \sqrt{\dfrac{961 - 0.2}{4}}$

$= \sqrt{\dfrac{960.8}{4}}$

$= \sqrt{240.2}$

$= 15.4984$

Co-efficient of Variation (Sample) $= \dfrac{S}{\bar{x}} \cdot 100\ \%$

$= \dfrac{15.4984}{39.2} \cdot 100\ \%$

$= 39.54\ \%$

Skewness :

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{59 + 49 + 39 + 29 + 20}{5}$

$= \frac{196}{5}$

$= 39.2$

x	$(x - \bar{x}) = (x - 39.2)$	$(x - \bar{x})^2 = (x - 39.2)^2$	$(x - \bar{x})^3 = (x - 39.2)^3$
59	19.8	392.04	7762.392
49	9.8	96.04	941.192
39	-0.2	0.04	-0.008
29	-10.2	104.04	-1061.208
20	-19.2	368.64	-7077.888
---	---	---	---
196	0	960.8	564.48

Sample Standard deviation $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

$= \sqrt{\frac{960.8}{4}}$

$= \sqrt{240.2}$

$= 15.4984$

Co-efficient of Variation (Sample) $= \frac{S}{\bar{x}} \cdot 100\,\%$

$= \frac{15.4984}{39.2} \cdot 100\,\%$

$= 39.54\,\%$

Skewness :

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{59 + 49 + 39 + 29 + 20}{5}$

$= \frac{196}{5}$

$= 39.2$

x	$(x - \bar{x}) = (x - 39.2)$	$(x - \bar{x})^2 = (x - 39.2)^2$	$(x - \bar{x})^3 = (x - 39.2)^3$
59	19.8	392.04	7762.392
49	9.8	96.04	941.192
39	-0.2	0.04	-0.008
29	-10.2	104.04	-1061.208
20	-19.2	368.64	-7077.888
---	---	---	---
196	0	960.8	564.48

Sample Standard deviation $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

$= \sqrt{\frac{960.8}{4}}$

$= \sqrt{240.2}$

$= 15.4984$

Sample Skewness $= \frac{\sum (x - \bar{x})^3}{(n - 1) \cdot S^3}$

$= \frac{564.48}{4 \cdot (15.4984)^3}$

$= \frac{564.48}{4 \cdot 3722.7126}$

$= 0.0379$

Geometric mean,Harmonic mean :

x	$\log(x)$	$\frac{1}{x}$
59	4.0775	0.0169
49	3.8918	0.0204
39	3.6636	0.0256
29	3.3673	0.0345
20	2.9957	0.05
---	---	---
	$\sum \log(x) = 17.9959$	$\sum \frac{1}{x} = 0.1475$

GM of X $= Anti\log\left(\frac{\sum f\log(x)}{n}\right)$

$= Anti\log\left(\frac{17.9959}{5}\right)$

$= Anti\log(3.5992)$

$= 36.5686$

HM of X $= \dfrac{n}{\sum\left(\frac{f}{x}\right)}$

$= \dfrac{5}{0.1475}$

$= 33.9026$

Mean deviation :

Mean $\bar{x} = \dfrac{\sum x}{n}$

$= \dfrac{59 + 49 + 39 + 29 + 20}{5}$

$= \dfrac{196}{5}$

$= 39.2$

x	$\lvert x - \bar{x} \rvert = \lvert x - 39.2 \rvert$
59	19.8
49	9.8
39	0.2
29	10.2
20	19.2
---	---
196	59.2

Mean deviation of Mean

$$\delta\bar{x} = \frac{\sum |x - \bar{x}|}{n}$$

$$\delta\bar{x} = \frac{59.2}{5}$$

$$\delta\bar{x} = 11.84$$

Co-efficient of Mean deviation $= \dfrac{\delta\bar{x}}{\bar{x}}$

$$= \frac{11.84}{39.2}$$

$$= 0.302$$

Quartile deviation :

Arranging Observations in the ascending order, We get :

20, 29, 39, 49, 59

Here, $n = 5$

$Q_1 = \left(\dfrac{n+1}{4}\right)^{th}$ value of the observation

$= \left(\dfrac{6}{4}\right)^{th}$ value of the observation

$= (1.5)^{th}$ value of the observation

$= 20 + 0.5(9)$

$= 20 + 4.5$

$= 24.5$

$Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ value of the observation

$= \left(\frac{3 \cdot 6}{4}\right)^{th}$ value of the observation

$= (4.5)^{th}$ value of the observation

$= 4^{th}$ observation $+0.5\left[5^{th} - 4^{th}\right]$

$= 49 + 0.5[59 - 49]$

$= 49 + 0.5(10)$

$= 49 + 5$

$= 54$

Inter Quartile range $= Q_3 - Q_1 = 54 - 24.5 = 29.5$

Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{54 - 24.5}{2} = \frac{29.5}{2} = 14.75$

Coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{54 - 24.5}{54 + 24.5} = \frac{29.5}{78.5} = 0.3758$