(1) S9245

Q.1 A)

1) False, In case of secondary data, data obtained might be outdated.
2) False, A and B are independent if $(AB) = (A)(B)/N$.
3) False, Mode can be obtained graphically using histogram.
4) False, Width = (Max - Min)/k.
5) False, Standard deviation cannot be calculated for open-end intervals.

B)

1) Methods of primary data collection - Direct personal observation, Indirect oral interviews, Mailed questionnaire method, Schedule method, From local agents.
2) Consistency conditions in case of two attributes:
$(AB) \geq 0$, $(\alpha B) \geq 0$, $(A\beta) \geq 0$ and $(\alpha\beta) \geq 0$,
(or equivalently) $\max\{0, (A)+(B)-N\} \leq (AB) \leq \min\{(A), (B)\}$.
3) Merits of Harmonic mean (Any one)
i) Rigidly defined.
~~ii) Easy to understand and easy to calculate~~
iii) Based upon all the values of given data.
iii) Capable of further mathematical treatment.
iv) Not much affected by sampling fluctuations.

Demerits:
i) Not easy to understand and calculate.
ii) Cannot be calculated if any observations are missing.
iii) Cannot be calculated for open end intervals.
iv) Not a good representative of data.
v) Affected by extreme values.
vi) Cannot be located graphically.

4) Skewness is deviation from symmetry. Bowley's measure $= \frac{(Q_1 - Q_2) + (Q_3 - Q_2)}{(Q_3 - Q_1)}$.

5) Components of 5 point summary: Min, Max, Median, Q_1 and Q_3.

Q.2

1) Types of variables (Any 2): Nominal Scale, Ordinal scale, Interval, Ratio depending upon its measuring scale

2) Components of table: Table number, Table title, Captions, stubs, body footnote, source of the data.

Characteristics of a good table: 1) Should be precise & easy to understan
2) self explainatory
3) Table should contain: Title (appropriate) Row & Column headings arranged in proper order, body must contain numeric information, footnote and source if needed.

3) When two attributes are consistent,

$(AB) \geq 0 \quad -(1)$

$(\alpha B) \geq 0 \rightarrow (AB) \leq (B)$ $\left(\because (AB)+(\alpha B)=(B) \ \& \ (\alpha B)\geq 0 \Rightarrow (AB)-(B)\leq 0\right.$

$(A\beta) \geq 0 \rightarrow (AB) \leq (A)$ $\left(\because (AB)+(A\beta)=(A) \ \& \ (A\beta)\geq 0 \rightarrow (AB)\leq (A)\right)$

$(\alpha\beta) \geq 0 \rightarrow (AB) \geq (A)+(B)-N$

$\max\{0, (A)+(B)-N\} \leq (AB) \leq \min\{(A),(B)\}$

4) $N=2500$, $(A)=420$, $(B)=670$, $(AB)=85$

	B	β	
A	85	335	420
α	585	1495	2080
	670	1830	N = 2500

Consistent, Negatively associated

$Q = -0.2133$

Q.5 1) Representation of the data graphically where c.f. < type shows no. of values below certain boundary and c.f. ≥ type shows no. of values above certain boundary

- Procedure : i) c.f. < type, (upper class boundary, <cf) to be plotted
  ii) c.f. ≥ type, (lower class boundary, ≥cf) to be plotted
  iii) Points plotted in both the cases must be joined by smooth curves.

2) Calculating Median for grouped freq. data:

i) find Median class

ii) formula: Median $= l_1 + (l_2 - l_1)\left(\frac{N/2 - cf}{fm}\right)$

iii) Example.

3) GM (for ungrouped data) $= \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \text{antilog}\left(\frac{\sum \log x_i}{n}\right)$

GM (for grouped data) $= \sqrt[N]{x_1^{f_1} x_2^{f_2} \cdots x_n^{f_n}} = \text{antilog}\left(\frac{\sum f_i \log x_i}{N}\right)$

Merits: 1) Based on all values
2) Capable of further mathematical treatment
3) Not much affeted by sampling fluctuations

Demerits: 1) Not easy to understand & calculate
2) Any value is 0 then GM = 0
3) Cannot be calculated if any value is missing
4) cannot be calculated for open end classes.
5) Affected by extreme values
6) Cannot be located graphically
etc.

4) Mean = 63.3

Median = 62.2727

Mode = 58.125

Q.4 1) Range = UL − LF1

Coeff of range = $\frac{ULI - LFI}{ULI + LFI}$

Merits: Rigidly defined, simple & easy to understand

Demerits: Based upon all the values, not capable of further mathematical treatment, affected by extreme values, affected by sampling variation, cannot be calculated for open end intervals etc.

2) Raw Moments = $\mu'_r := \frac{\sum x_i^r}{n}$ $(r = 0, 1, 2, \dots)$

r^{th} central moment = $\mu_r = \frac{\sum (x_i - \bar{x})^r}{n}$ $(r = 0, 1, 2, \dots)$

Derive: $\mu_r = \mu_r' - \binom{r}{1}\mu'_{r-1}\mu'_1 + \binom{r}{2}\mu'_{r-2}(\mu'_1)^2 + \dots + (-1)^{r-1}\left(\binom{r}{r-1} - 1\right)(\mu'_1)^r$.

(Till $r = 4$)

3) Skewness - Deviation from symmetry.

Types: negatively skewed, positively skewed and symmetric

Karl Pearson's skewness coeff := $S_{kp} := \frac{3(\text{Mean} - \text{Median})}{\sigma}$

4) Median:

{3, 7, 8, 5, 12, 14, 21, 15, 18, 14}

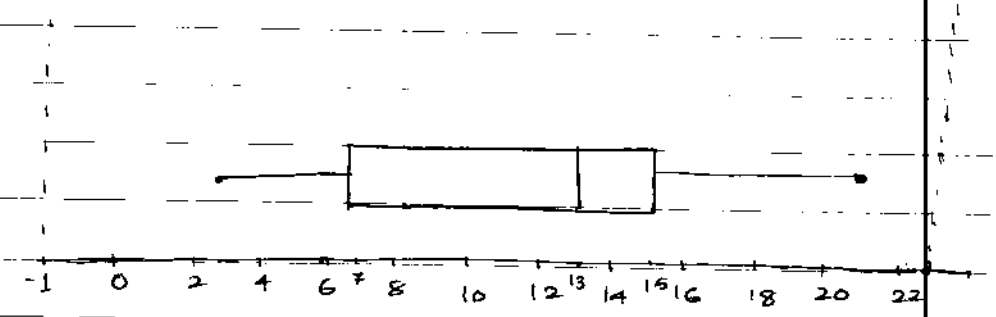
Data arranged: 3, 5, 7, 8, 12, 14, 14, 15, 18, 21

Median = 13

Q_1 = 7

Q_3 = 15

IQR = 8



Q 5

1) Sampling techniques: Simple random sampling (in detail)
Stratified random sampling (in detail)

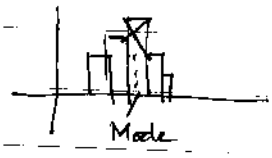
2) A, B are independent if $(AB) = \frac{(A)(B)}{N}$

Given A, B are NOT independent.

3) Arithmetic mean: formula for ungrouped: $\frac{\sum x_i}{n}$, For grouped: $\frac{\sum f_i x_i}{N}$.
Shifting origin to 'a' and changing scale by h: u_i be the new values then

$\bar{x} = a + h\bar{u}$.

4) Histogram: Procedure & Mode using Histogram



5) Stem & Leaf example with procedure.
Demerits: 1) Not very informative
2) Might be messy if the data is large.

6) Standard deviation (σ) := $\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$ (for ungrouped data)

$= \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i - \bar{x})}{N}}$ (for grouped data)

coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100$.

7) Kurtosis - Defn
types - Leptokurtic, Mesokurtic, Platykurtic. (With diagram)