PACK 1 – SOLUTION SET (STANDALONE, WITH QUESTIONS RESTATED)

SECTION A — BIOSTATISTICS

Q1. HDL Cholesterol Dataset (n = 14)

The HDL cholesterol values (mg/dL) for 14 adults are:

$$42,\ 48,\ 52,\ 50,\ 61,\ 58,\ 54,\ 49,\ 63,\ 68,\ 72,\ 70,\ 66,\ 55$$

Compute:

a) Mean
b) Median
c) Mode
d) Range
e) Q1, Q3, and IQR
f) Sample variance
g) Sample standard deviation
h) Coefficient of variation (CV%)

Solution 1

Step 1 – Sort the data:

$$42,\ 48,\ 49,\ 50,\ 52,\ 54,\ 55,\ 58,\ 61,\ 63,\ 66,\ 68,\ 70,\ 72$$

Number of observations: (n = 14) (even).

---

a) Mean

Formula:

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

$$\sum x_i = 808$$

Substitute:

$$\bar{x} = \frac{808}{14} = 57.71\ \mathrm{mg/dL}\ \mathrm{(approx.)}$$

Interpretation:

The average HDL cholesterol in this sample is about **57.7 mg/dL**. The mean uses all data points and is therefore sensitive to extreme values.

---

b) Median

For **even n**, the median is the average of the $\frac{n}{2}$-th and $\frac{n}{2}+1$-th values.

Formula:

$$Median = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

Here, $n/2 = 7$:

- 7th value = 55
- 8th value = 58

$$Median = \frac{55 + 58}{2} = 56.5 \text{ mg/dL}$$

Interpretation:
Half of the individuals have HDL < 56.5 mg/dL, and half have HDL > 56.5 mg/dL. Median is robust to outliers.

---

c) Mode
Definition: Mode is the value that occurs most frequently.
Each value appears only once → **no mode**.
Interpretation:
For continuous biochemical data, having "no mode" in raw data is common unless we group values into classes.

---

d) Range
Formula:

$$Range = \max(x) - \min(x)$$

$$Range = 72 - 42 = 30 \text{ mg/dL}$$

Interpretation:
There is a **30 mg/dL** spread between the lowest and highest HDL values. Range is very sensitive to extreme values and should be interpreted alongside other measures.

---

e) Q1, Q3, and IQR
For **even n**, we split into two equal halves (no actual median value to remove):

- Lower half (first 7 values):
  42, 48, 49, 50, 52, 54, 55
- Upper half (last 7 values):
  58, 61, 63, 66, 68, 70, 72

For an **odd number within each half** (7), the median is the middle (4th) value.

- **Q1** = 4th value of lower half = 50
- **Q3** = 4th value of upper half = 66

Formula:

$$IQR = Q3 - Q1$$

$$IQR = 66 - 50 = 16 \text{ mg/dL}$$

Interpretation:
The middle 50% of HDL values lie in a 16 mg/dL band. The IQR is not influenced by outliers and is a good measure of **central spread**.

---

f) Sample Variance (shortcut formula)
Formula (sample variance, shortcut):

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

We have:

$$\sum x = 808, \quad \sum x^2 = 47732$$

Substitute:

$$s^2 = \frac{47732 - \frac{808^2}{14}}{13}$$

$$\frac{808^2}{14} = \frac{652864}{14} = 46633.14$$

$$s^2 = \frac{47732 - 46633.14}{13} = \frac{1098.86}{13} = 84.53 \ (\text{mg/dL})^2 \ (\text{approx.})$$

Interpretation:
The variance is about **84.5 (mg/dL)²**. This quantifies the average squared deviation from the mean. It's mathematically useful but less intuitive than SD because it's in squared units.

---

g) Sample Standard Deviation
Formula:

$$s = \sqrt{s^2}$$

$$s = \sqrt{84.53} \approx 9.19 \text{ mg/dL}$$

Interpretation:
On average, HDL values differ from the mean by about **9.2 mg/dL**. This indicates **moderate variability** within the group.

---

h) Coefficient of Variation (CV%)

Formula:

$$CV = \left(\frac{s}{\bar{x}}\right) \times 100\%$$

$$CV = \frac{9.19}{57.71} \times 100 \approx 15.9\%$$

Interpretation:
A CV of about **16%** suggests **moderate relative variability** in HDL values. CV is useful when comparing variability between datasets with different means or units.

Q2. Hemoglobin and Z-Score (with Z-table)
In a population of adult men, Hb levels follow a normal distribution with:

- Mean (μ) = 14.8 g/dL
- SD (σ) = 1.2 g/dL

A man has Hb = 12.4 g/dL.
Use the following Z-table extract (showing **P(Z < z)**):

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262

Tasks:
a) Z-score
b) Proportion of men with Hb below 12.4 g/dL
c) Clinical interpretation

Solution 2
a) Z-score
Formula:

$$Z = \frac{X - \mu}{\sigma}$$

Where
X = observed value,
μ = population mean,
σ = population SD.

Substitute:

$$Z = \frac{12.4 - 14.8}{1.2} = \frac{-2.4}{1.2} = -2.0$$

---

b) Proportion below Hb = 12.4 g/dL

We need $P(Z < -2.0)$. From the Z-table row for Z = −2.0, column 0.00:

$$P(Z < -2.0) = 0.0228$$

So:

$$\text{Proportion} = 0.0228 = 2.28\%$$

---

c) Interpretation

- A Z-score of **−2.0** means this man's Hb is **2 standard deviations below the mean**.
- Being at **2.28th percentile** implies that **only about 2–3%** of men in this population have Hb at or below 12.4 g/dL.
- Clinically, this strongly suggests **anemia**, assuming standard Hb cutoffs are used for adult males.
- This example illustrates how the **normal distribution** and Z-scores allow us to place an individual value in the context of the population distribution and interpret how "unusual" it is.

SECTION B — DEMOGRAPHY

Q3. Fertility Indicators

District X:

- Live births = 8,400
- Mid-year population = 510,000
- Women 15–49 = 120,000
- Women 20–24 = 32,000
- Births to women 20–24 = 2,100

Compute CBR, GFR, ASFR (20–24).

Solution 3

a) Crude Birth Rate (CBR)

Formula:

$$CBR = \frac{\text{Number of live births in a year}}{\text{Mid-year population}} \times 1000$$

Substitute:

$$CBR = \frac{8400}{510000} \times 1000 = 16.47 \approx 16.5 \text{ per 1000 population}$$

Interpretation:

About **16–17 births per 1,000 population**. As a crude measure, it does not account for how many people are women of reproductive age.

---

b) General Fertility Rate (GFR)
Formula:

$$GFR = \frac{\text{Live births in a year}}{\text{Number of women aged 15–49}} \times 1000$$

Substitute:

$$GFR = \frac{8400}{120000} \times 1000 = 70 \text{ per } 1000 \text{ women } (15\text{–}49)$$

Interpretation:
There are **70 births per 1,000 women of reproductive age**. This is a more refined measure of fertility than CBR.

---

c) Age-Specific Fertility Rate (ASFR), 20–24 years
Formula:

$$ASFR_{20-24} = \frac{\text{Births to women aged 20–24}}{\text{Number of women aged 20–24}} \times 1000$$

Substitute:

$$ASFR_{20-24} = \frac{2100}{32000} \times 1000 = 65.625 \approx 65.6 \text{ per } 1000 \text{ women } (20\text{–}24)$$

Interpretation:
Women aged 20–24 have a fertility rate of about **66 births per 1,000 women**. This is often one of the highest fertility age groups, informing family planning priorities.

Q4. Total Fertility Rate (TFR)
The Age-Specific Fertility Rates (ASFR) for a district are given below (per 1,000 women in each age group):

Age Group (years)	ASFR
15–19	30
20–24	110
25–29	145
30–34	80
35–39	20

Using 5-year age intervals, calculate the Total Fertility Rate (TFR).

Solution 4

Formula (5-year groups):

$$TFR = \frac{\sum(ASFR \times 5)}{1000}$$

Compute contribution of each age group:

- 15–19: $30 \times 5 = 150$
- 20–24: $110 \times 5 = 550$
- 25–29: $145 \times 5 = 725$
- 30–34: $80 \times 5 = 400$
- 35–39: $20 \times 5 = 100$

$$\sum(ASFR \times 5) = 150 + 550 + 725 + 400 + 100 = 1925$$

$$TFR = \frac{1925}{1000} = 1.925 \approx 1.93$$

Interpretation:
This population has a TFR of **about 1.9 children per woman**, which is **below the replacement level** (~2.1). If sustained over time (and ignoring migration), the population will eventually start to shrink and age.

Q5. Dependency Ratios
In a given population:

- Number of people aged 0–14 years = 24,000
- Number of people aged 15–59 years = 52,000
- Number of people aged 60+ years = 12,000

Compute:
a) Young dependency ratio
b) Old dependency ratio
c) Total dependency ratio
(Express each as number of dependents per 100 persons in the 15–59 age group.)

Solution 5
Let "working age" = 15–59 years = 52,000.
a) Young dependency ratio
Formula:

$$Young\ DR = \frac{\text{Population 0–14}}{\text{Population 15–59}} \times 100$$

$$Young\ DR = \frac{24000}{52000} \times 100 \approx 46.15 \approx 46.2$$

So **46 young dependents per 100 workers**.

---

b) Old dependency ratio
Formula:

$$Old\ DR = \frac{\text{Population } 60+}{\text{Population } 15\text{–}59} \times 100$$

$$Old\ DR = \frac{12000}{52000} \times 100 \approx 23.08 \approx 23.1$$

So **23 elderly dependents per 100 workers**.

---

c) Total dependency ratio
Formula:

$$Total\ DR = \frac{\text{Population } 0\text{–}14 + 60+}{\text{Population } 15\text{–}59} \times 100$$

$$Total\ DR = \frac{24000 + 12000}{52000} \times 100 = \frac{36000}{52000} \times 100 \approx 69.2$$

Interpretation:
There are about **69 dependents per 100 working-age people**, with **young dependents (46)** forming a larger share of the burden than **elderly (23)**. This has implications for planning education, childcare, and future employment.

SECTION C — VITAL STATISTICS

Q6. Infant, Neonatal, and Perinatal Mortality
District Y reports for one year:

- Live births = 15,000
- Infant deaths (0–11 months) = 450
- Neonatal deaths (0–27 days) = 270
- Early neonatal deaths (0–6 days) = 150
- Stillbirths (≥ 28 weeks gestation) = 190

Compute:
a) Infant Mortality Rate (IMR)
b) Neonatal Mortality Rate (NMR)
c) Perinatal Mortality Rate (PMR)

Solution 6
a) Infant Mortality Rate (IMR)
Formula:

$$IMR = \frac{\text{Infant deaths } (<1 \text{ year})}{\text{Live births}} \times 1000$$

$$IMR = \frac{450}{15000} \times 1000 = 30 \text{ per 1000 live births}$$

Interpretation:
30 infants die before age 1 for every 1,000 babies born alive. IMR is a key indicator of **overall health system performance** and socioeconomic development.

---

b) Neonatal Mortality Rate (NMR)
Formula:

$$NMR = \frac{\text{Neonatal deaths } (<28 \text{ days})}{\text{Live births}} \times 1000$$

$$NMR = \frac{270}{15000} \times 1000 = 18 \text{ per } 1000 \text{ live births}$$

Interpretation:
Most infant deaths here occur in the **neonatal period**. High NMR suggests problems with **antenatal care, intrapartum care, prematurity, infections**, etc.

---

c) Perinatal Mortality Rate (PMR)
Definition: Stillbirths (≥28 weeks) + early neonatal deaths (0–7 days), per 1,000 **total births**.
Formula:

$$PMR = \frac{\text{Stillbirths} + \text{Early neonatal deaths}}{\text{Live births} + \text{Stillbirths}} \times 1000$$

Total births = 15,000 + 190 = 15,190

Perinatal deaths = 190 + 150 = 340

$$PMR = \frac{340}{15190} \times 1000 \approx 22.4 \text{ per } 1000 \text{ total births}$$

Interpretation:
Perinatal mortality reflects late fetal and very early neonatal losses, and is strongly influenced by **maternal health, obstetric care, birth practices, and immediate newborn care**.

Q7. Maternal Mortality Ratio (MMR)
In a district:

- Live births = 22,500
- Maternal deaths (as per WHO definition) = 18

Compute the Maternal Mortality Ratio (MMR).

Solution 7
Formula:

$$MMR = \frac{\text{Maternal deaths}}{\text{Live births}} \times 100000$$

$$MMR = \frac{18}{22500} \times 100000 = 80 \text{ per } 100{,}000 \text{ live births}$$

Interpretation:
An MMR of **80 per 100,000** suggests a moderate level of maternal mortality. It reflects performance of the **antenatal, intranatal, and emergency obstetric care systems**, and can be compared with district, state, or national benchmarks.

Q8. Attack Rate (AR)
At a school function:

- 180 children ate sandwiches.
- 54 of these children developed symptoms of food poisoning.

Compute the **attack rate (AR)** among those who ate sandwiches, and write a one-line interpretation.

Solution 8
Formula:

$$AR = \frac{\text{Number of people who became ill}}{\text{Number of people exposed}} \times 100$$

$$AR = \frac{54}{180} \times 100 = 30\%$$

Interpretation:
An attack rate of **30%** among those who ate the sandwiches suggests a strong likelihood that the sandwiches were contaminated. In outbreak investigations, this would be compared with the attack rate in non-exposed children to confirm the hypothesis.

Q9. Standardized Mortality Ratio (SMR)
In a factory population, mortality was studied for one year:

- Observed deaths in the factory population = 37
- Expected deaths (based on applying standard population death rates to the factory's age structure) = 30

Compute the Standardized Mortality Ratio (SMR) and interpret the result.

Solution 9
Formula:

$$SMR = \frac{\text{Observed deaths}}{\text{Expected deaths}} \times 100$$

$$SMR = \frac{37}{30} \times 100 \approx 123.3$$

Interpretation:
An SMR of about **123** means mortality in this factory population is **23% higher than expected** based on standard rates. This may suggest:

- Occupational exposures
- Poor access to healthcare
- Underlying high-risk profile

It warrants further investigation into working conditions and health services for workers.