**Section 1: Answers**

**1. A statistics test was conducted for 10 learners in a class. The mean of their score is 85, and the variance of the score is zero. What can you interpret about the score obtained by all learners?**

* **Answer**: Since the variance is zero, it means there is no deviation from the mean. All 10 learners scored exactly 85. In other words, the score of every learner is 85.

**2. In a residential locality, the mean size of the house is 2224 square feet, and the median value of the house is 1500 square feet. What can you interpret about the skewness in the distribution of house size? Are there bigger or smaller houses in the residential locality?**

* **Answer**: When the mean is greater than the median, it suggests that the distribution is **right-skewed** (positively skewed). This indicates that there are a few larger houses pulling the mean up. Therefore, there are **bigger houses** in the locality, but the majority of the houses are smaller, as indicated by the lower median.

**3. The following table shows the mean and variance of the expenditure for two groups of people. You want to compare the variability in expenditure for both groups with respect to their mean. Which statistical measure would you use to evaluate the variability in expenditure? Please provide an explanation for your answer.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr no** | **Expenditure** | **Group 1** | **Group 2** |
| 0 | Mean | $500,000.00 | $40,000.00 |
| 1 | Std Dev | $125,000.00 | $10,000.00 |

* **Answer**: The appropriate measure to compare the variability relative to the mean is the **Coefficient of Variation (CV)**. The CV is calculated as:

Cv = (std dev/mean) \* 100 %

This gives the percentage of variability relative to the mean, allowing you to compare variability between two groups with different means.

* Group 1 CV: (125,000/500,000)×100=25%
* Group 2 CV: (10,000/40,000)×100=25

Both groups have the same relative variability (25%).

**4. During the survey, the ages of 80 patients infected by COVID and admitted to one of the city hospitals were recorded. The data is represented in the less-than cumulative frequency distribution table.**

|  |  |
| --- | --- |
| **Age in years** | **Number of Patients** |
| 5 - 15 | 16 |
| 15 - 25 | 11 |
| 25 - 35 | 21 |
| 35 - 45 | 23 |
| 45 - 55 | 14 |
| 55 - 65 | 5 |

**a. Which class interval has the highest frequency?**

* **Answer**: The class interval **35-45** has the highest frequency with 23 patients.

**b. Which age group was affected the least?**

* **Answer**: The class interval **55-65** has the lowest frequency with 5 patients.

**c. How many patients aged 45 years and above were admitted?**

* **Answer**: To find the number of patients aged 45 and above, sum the frequencies for the class intervals 45-55 and 55-65: 14+5=19 patients14 + 5 = 19 \{ patients}14+5=19 patients

**d. Which is the modal class interval in the above dataset?**

* **Answer**: The modal class interval is the one with the highest frequency, which is **35-45**.

**e. What is the median class interval of age?**

* **Answer**: The median is the middle value in the ordered dataset. Since there are 80 patients, the median will be the value at the 40th patient (80/2).
  + Cumulative frequencies:
    - 5-15: 16 patients
    - 15-25: 16 + 11 = 27 patients
    - 25-35: 27 + 21 = 48 patients (Median lies here)
  + Therefore, the **median class interval is 25-35**.

**5. Assume you are the trader, and you have invested over the years, and you are worried about the average return on investment. What average method would you use to compute the average return for the data given below?**

* **Answer**: The best method to compute the average return on investment is the **Geometric Mean**, as it accounts for the compounding nature of investment returns over time.

**6. Suppose you have been told to measure the average height of all the males on the earth. What would be your strategy for the same? Would the average height be a parameter or a statistic? Justify your answer.**

* **Answer**: Since it is impractical to measure the height of all males on Earth, you would use a **sampling strategy**. You would randomly select a representative sample of males from different regions around the world to calculate the average height.
  + This average height is a **statistic** because it is derived from a sample. A **parameter** refers to a measure from the entire population, which is usually unknown.

**7. Calculate the z-score of the following numbers:**

**X = [4.5, 6.2, 7.3, 9.1, 10.4, 11]**

* **Step 1**: Find the mean (μ\muμ) of X.

μ=4.5+6.2+7.3+9.1+10.4+116=8.08\mu = {4.5 + 6.2 + 7.3 + 9.1 + 10.4 + 11}{6} = 8.08μ=64.5+6.2+7.3+9.1+10.4+11 =8.08

* **Step 2**: Find the standard deviation (σ\sigmaσ).

σ=(4.5−8.08)2+(6.2−8.08)2+(7.3−8.08)2+(9.1−8.08)2+(10.4−8.08)2+(11−8.08)26≈2.13\sigma = \sqrt{{(4.5-8.08)^2 + (6.2-8.08)^2 + (7.3-8.08)^2 + (9.1-8.08)^2 + (10.4-8.08)^2 + (11-8.08)^2}{6}} \approx 2.13σ=6(4.5−8.08)2+(6.2−8.08)2+(7.3−8.08)2+(9.1−8.08)2+(10.4−8.08)2+(11−8.08)2 ≈2.13

* **Step 3**: Compute the z-score for each value using the formula:

Z=X−μσZ = Z=σX−μ

For each value of X:

* + Z1=4.5−8.082.13≈−1.68Z\_1 = {4.5 - 8.08}{2.13} \approx -1.68Z1 =2.134.5−8.08 ≈−1.68
  + Z2=6.2−8.082.13≈−0.88Z\_2 = {6.2 - 8.08}{2.13} \approx -0.88Z2 =2.136.2−8.08 ≈−0.88
  + Z3=7.3−8.082.13≈−0.37Z\_3 = {7.3 - 8.08}{2.13} \approx -0.37Z3 =2.137.3−8.08 ≈−0.37
  + Z4=9.1−8.082.13≈0.48Z\_4 = {9.1 - 8.08}{2.13} \approx 0.48Z4 =2.139.1−8.08 ≈0.48
  + Z5=10.4−8.082.13≈1.09Z\_5 = {10.4 - 8.08}{2.13} \approx 1.09Z5 =2.1310.4−8.08 ≈1.09
  + Z6=11−8.082.13≈1.37Z\_6 = {11 - 8.08}{2.13} \approx 1.37Z6 =2.1311−8.08 ≈1.37