**Statistics Project Guide – Descriptive and Inferential Analysis**

**Objective:** Build foundational and applied skills in statistics, including descriptive measures, confidence intervals, hypothesis testing, and data-driven decision-making.

**Table of Contents**

1. Key Statistical Definitions (Q1)
2. Measures of Central Tendency – Definitions (Q2)
3. Manual Calculation of Mean, Median, Mode (Q3)
4. Levels of Measurement (Q4)
5. Variance and Standard Deviation – Theory (Q5)
6. Manual Calculation – Variance and Standard Deviation (Q6)
7. Range and Interquartile Range (IQR) (Q7)
8. Five-number Summary and Boxplot Concept (Q8)
9. Confidence Interval for the Mean (Q9)
10. Hypothesis Testing – One Sample Z-Test (Q10)
11. Dataset Guidance
12. Expected Outcomes

**Q1. Key Statistical Definitions**

**Objective:** Understand foundational statistical terms.

* **Population and Sample:**
  + Population = entire set of interest (e.g., all students).
  + Sample = subset of the population for study.
* **Descriptive vs Inferential Statistics:**
  + Descriptive = summarize data (mean, median, charts).
  + Inferential = make predictions/generalizations (hypothesis testing, CI).
* **Parameter vs Statistic:**
  + Parameter = numerical description of population (μ).
  + Statistic = numerical description of sample (x̄).
* **Qualitative vs Quantitative Data:**
  + Qualitative = categories (gender, colors).
  + Quantitative = numeric (age, height, income).

**Q2. Measures of Central Tendency – Definitions**

* **Mean:** Sum of all values ÷ number of values.
  + Example: {4,6,8} → Mean = 6
* **Median:** Middle value in ordered data.
  + Example: {2,5,7} → Median = 5
* **Mode:** Most frequent value.
  + Example: {3,4,4,5} → Mode = 4

**Q3. Manual Calculation of Mean, Median, Mode**

**Dataset:** 12, 18, 14, 16, 18, 20, 18, 15, 12, 18, 14, 16, 18, 20, 18, 15

* Mean = 16.38
* Median = 17
* Mode = 18

**Q4. Levels of Measurement**

* **Nominal:** Categories, no order (e.g., blood type).
* **Ordinal:** Ordered categories (e.g., satisfaction rating).
* **Interval:** Ordered, equal spacing, no true zero (e.g., Celsius).
* **Ratio:** Ordered, equal spacing, true zero (e.g., weight).

**Q5. Variance and Standard Deviation – Theory**

* **Variance:** Average squared deviation from mean.
* **Standard Deviation:** √Variance; interpretable in original units.
* **Why SD > Variance:** SD is in same units as data; easier to understand.

**Q6. Manual Calculation – Variance & SD**

**Dataset:** 8, 10, 12, 14, 16, 10, 12, 14, 16

* Sample Variance = 7.77
* Sample SD = 2.79

**Q7. Range and IQR**

**Dataset:** 22, 29, 25, 31, 35, 40, 45, 48, 50

* Ascending order: 22, 25, 29, 31, 35, 40, 45, 48, 50
* Range = 28
* Q1 = 27, Q3 = 46.5
* IQR = 19.5

**Q8. Five-number Summary & Boxplot Concept**

* **Five-number summary:** Minimum, Q1, Median, Q3, Maximum
* **Boxplot:** Visual summary, detects outliers
* **Outliers:** Values < Q1–1.5×IQR or > Q3+1.5×IQR

**Q9. Confidence Interval for the Mean**

**Dataset:** Sample n=36, mean=162 cm, SD=6 cm

* SE = 6/√36 = 1
* CI = 162 ± 1.96×1 → (160.04, 163.96)

**Q10. Hypothesis Testing – One Sample Z-Test**

**Dataset:** Population μ=30,000, Sample n=49, x̄=31,000, SD=4,900

* **H0:** μ=30,000, **Ha:** μ>30,000
* SE = 4,900/√49 = 700
* Z = (31,000–30,000)/700 = 1.43
* Critical Z = 1.645 → Fail to reject H0
* Conclusion: Not enough evidence for increase