Theoretical Part

1. Why the Median Minimizes the Sum of Absolute Deviations

Let’s say we have a dataset: $x_1, x_2, \ldots, x_n.$

We’re trying to find a “center” value c. One way to measure how good a choice of c is is to add up the **absolute deviations** of all points from c:

$$S(c) = \sum_{i=1}^{n} |x_i - c|.$$

- **Case 1: c < Median**
  If c is smaller than the median, most of the data lies above it. If we move c upward, we reduce the distance to many points, while only slightly increasing the distance to the few below. So overall, S(c) goes down as c moves toward the median.
- **Case 2: c = Median**
  At the median, the data is balanced: about half the points are below and half are above. If we move c in either direction, this balance is lost, and the total distance starts to increase. That means the minimum is reached exactly at the median.
- **Case 3: c > Median**
  Now most of the data lies below c. If we shift c downward, we again reduce many distances at once, so S(c) decreases until we get back to the median.

Conclusion:
The median is the one and only “balance point” where the sum of absolute deviations is smallest:

$$\text{Median} = \arg\min_{c} \sum_{i=1}^{n} |x_i - c|.$$

This is why the median is often called a *robust* measure of center—it naturally minimizes overall distance in a way that resists being pulled around by extreme values.

2. Different Ways to Define a "Center"

The idea of a “center” or *location statistic* is broader than just the mean or median. Depending on what you care about (minimizing errors, balancing probabilities, ignoring outliers, etc.), there are different definitions.

A. Optimization-Based Definitions

- **Arithmetic Mean:** The average we all know. It's the value that minimizes the *sum of squared deviations*:

$$\bar{x} = \arg\min_c \sum_{i=1}^{n} (x_i - c)^2.$$

- **Median:** Minimizes the *sum of absolute deviations*:

$$\text{Median} = \arg\min_c \sum_{i=1}^{n} |x_i - c|.$$

- **General Loss Minimization:** In general, if you pick any loss function L(x,c), you can define the "center" as:

$$c^* = \arg\min_c \sum_{i=1}^{n} L(x_i, c).$$

B. Quantile-Based Definitions

- **Median as the 50th Percentile:** By definition, the median is the point where half the data lies below it:

$$P(x \leq c) = 0.5.$$

- **General Quantiles:** More generally, the p-quantile is the value where a fraction p of the data is below it:

$$Q(p) = \inf\{c : F(c) \geq p\}.$$

So by choosing different p, we get a whole family of "centers."

C. Frequency-Based Definitions

- **Mode:** The most common value in the dataset, or for continuous distributions, the point where the probability density is highest:

$$\text{Mode} = \arg\max_c f(c).$$

D. Robust and Composite Measures

Sometimes we want definitions that reduce the effect of outliers. Examples include:

- **Trimmed Mean:** Throw away a percentage of the largest and smallest values, then average the rest.
- **Winsorized Mean:** Similar, but instead of discarding extremes, replace them with the nearest kept value.
- **Weighted Mean:** Give each point a weight w_i:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

- **Trimean:** A mix of quartiles and the median:

$$\text{Trimean} = \frac{Q_1 + 2 \cdot \text{Median} + Q_3}{4}.$$

3.Generalization and Infinite Possibilities

Each of these definitions is based on a specific criterion for centrality—whether it is minimizing a certain loss (squared, absolute, etc.), balancing probabilities (quantiles), or maximizing frequency (mode). Since one can choose any loss function L(x,c), any quantile level Q(p), or even combine multiple criteria, the framework for defining a location statistic is infinitely flexible.

Thus, beyond standard measures such as the mean, median, and mode, there exists an infinite variety of ways to define the "center" of a distribution, tailored to the characteristics of the data or the needs of the analysis.