Probability and Statistics

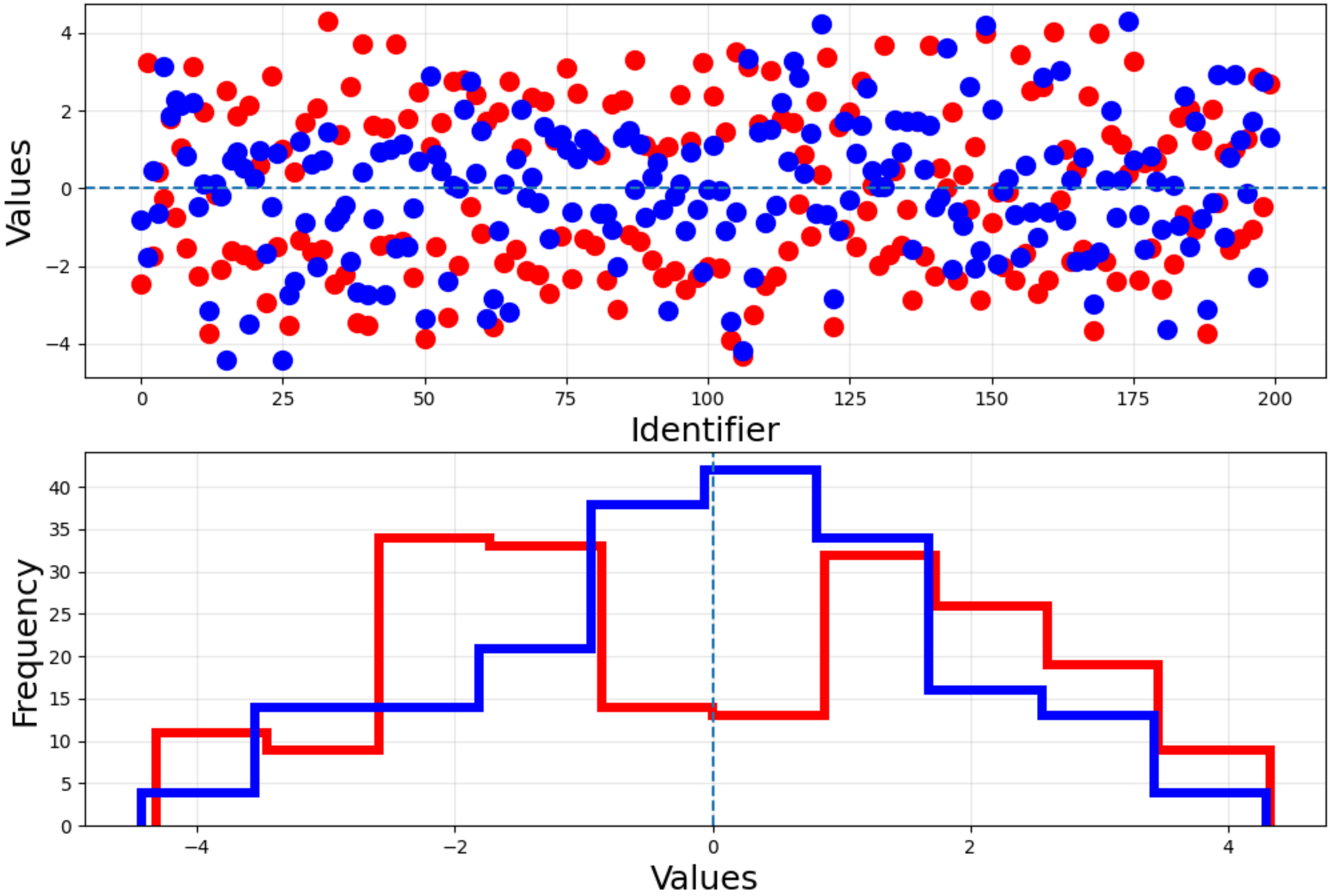
MA2103 - 2023

Rajesh Kumble Nayak

Two set of data with same mean and median. The spread is different

Variability or dispersion is a very important characteristic of data.

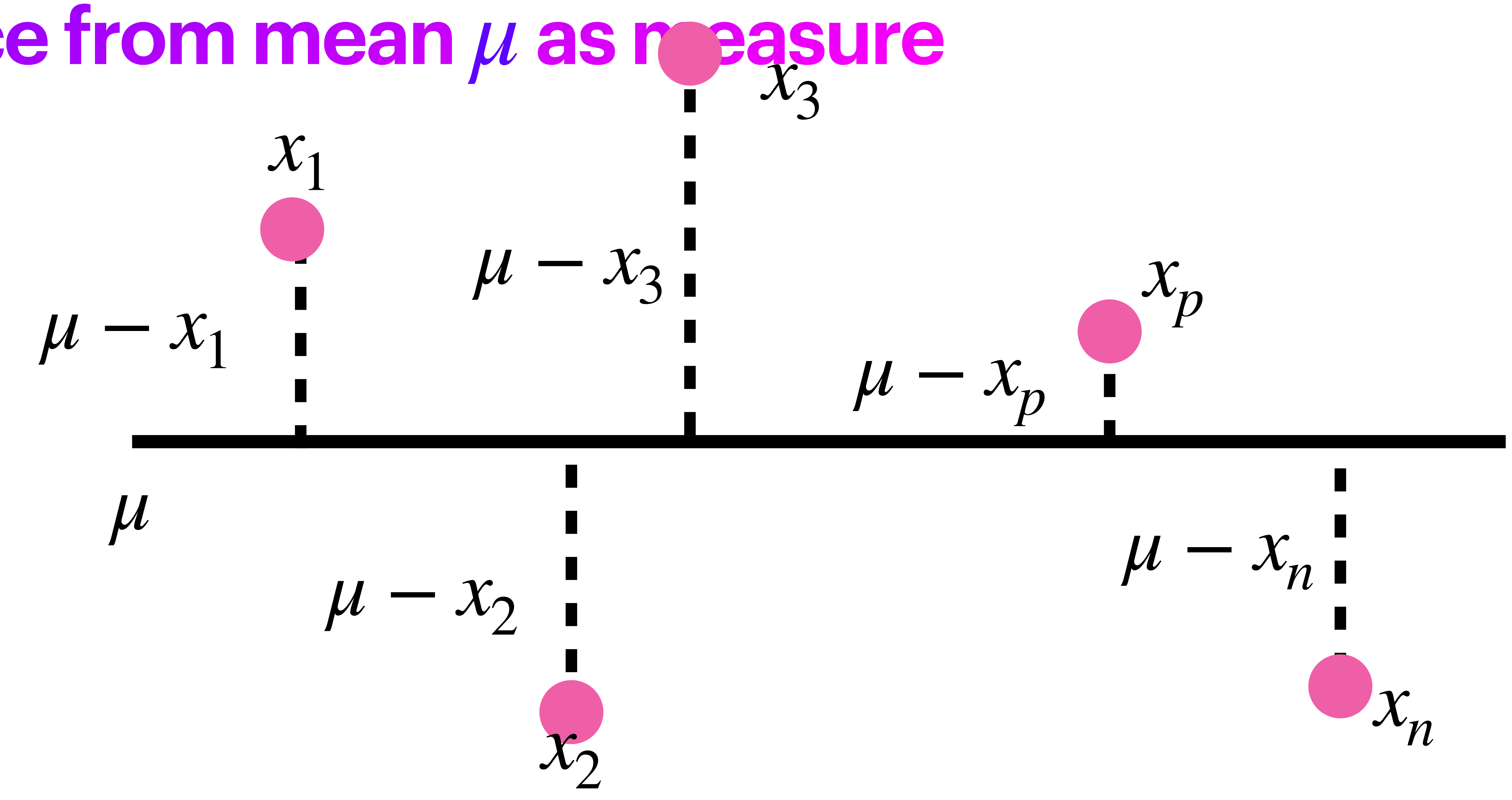
One may propose the range or maximum - minimum and simple measure of dispersion

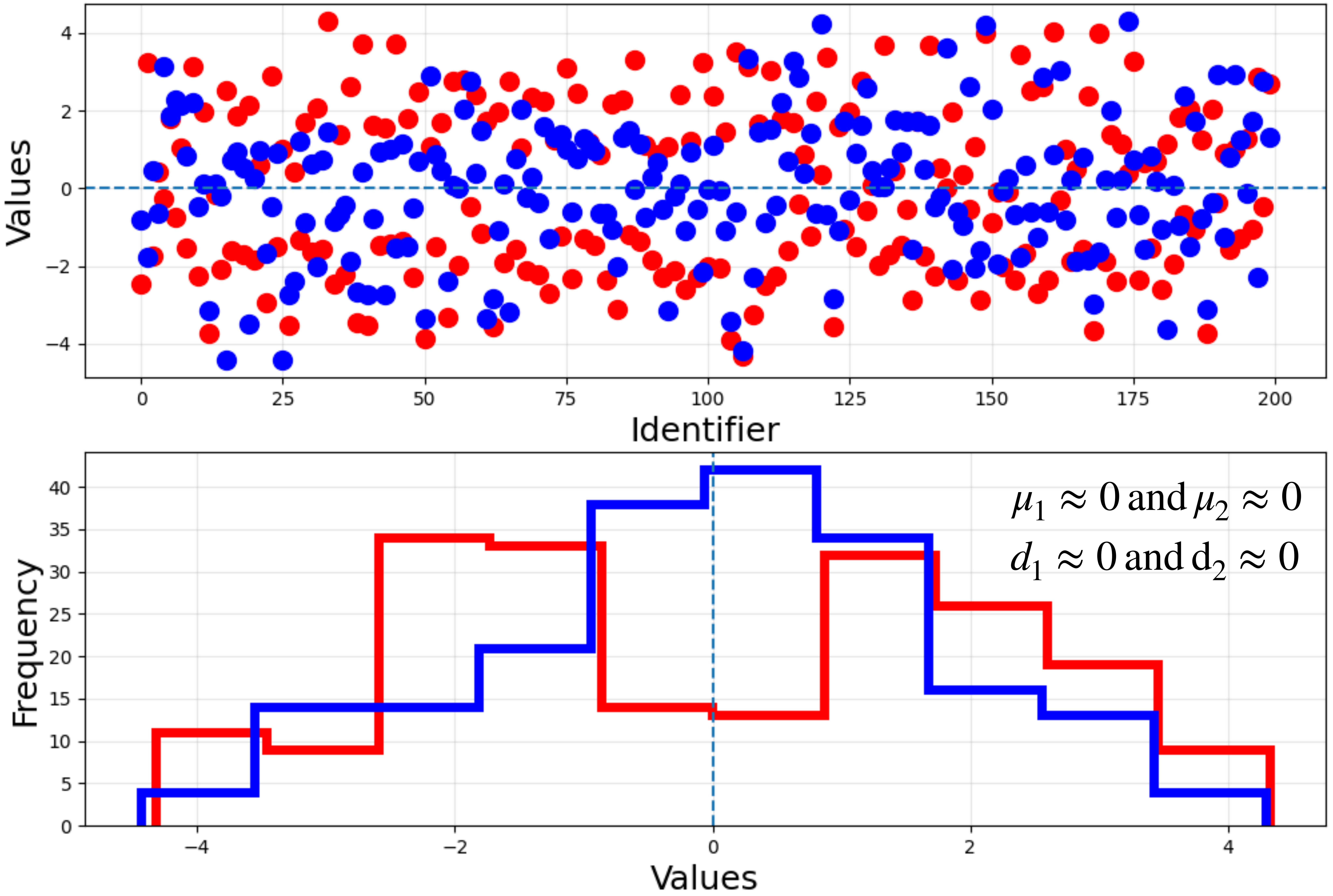

Values
Identifier
Frequency
Values

Let's use distance from mean μ as measure

mean $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$

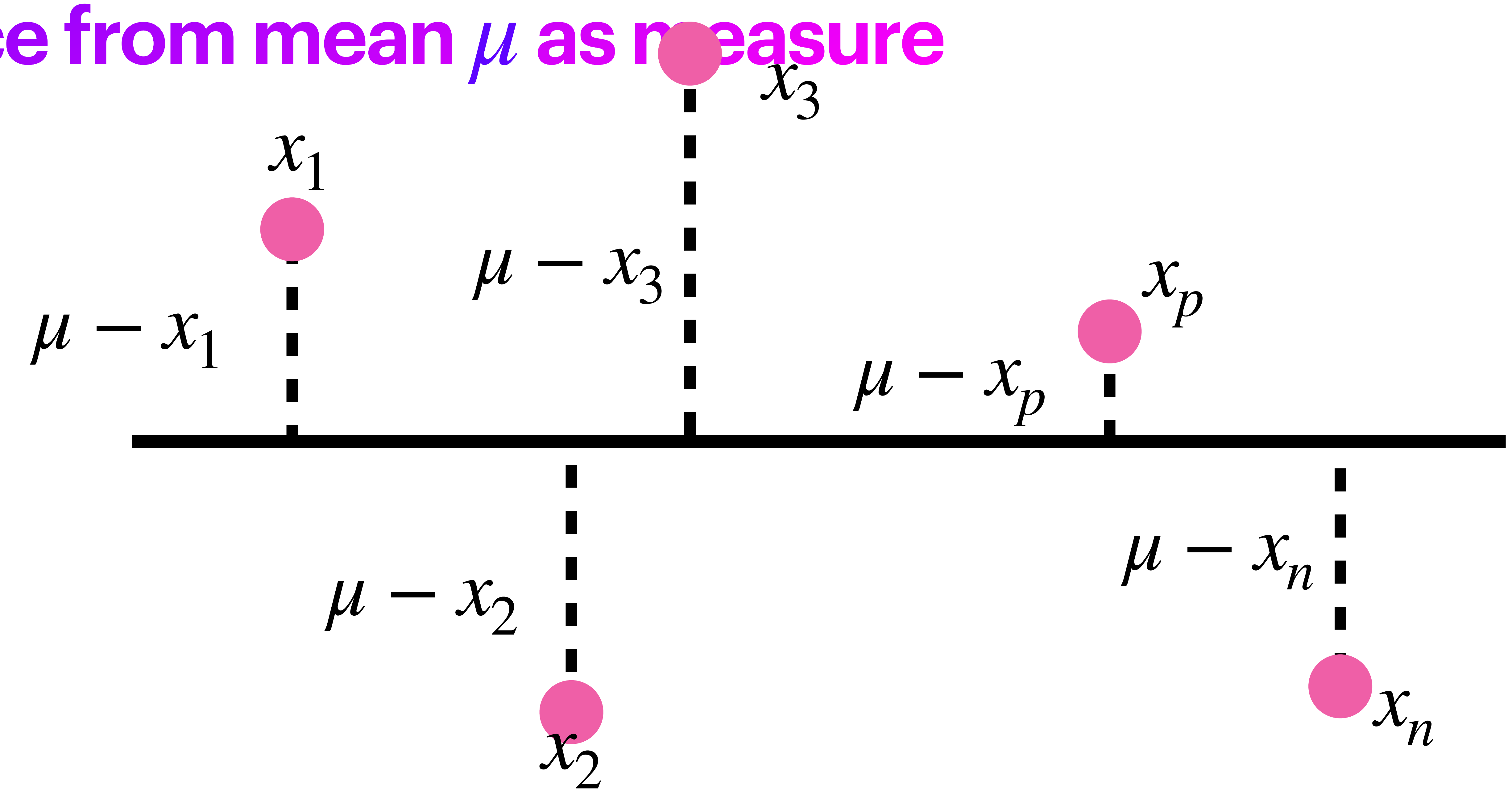
Let's define $d = \sum_{n=1}^{M} (x_n - \mu)$




Values
Identifier
Frequency
Values
$\mu_1 \approx 0$ and $\mu_2 \approx 0$
$d_1 \approx 0$ and $\mathrm{d}_2 \approx 0$

Let's use distance from mean μ as measure

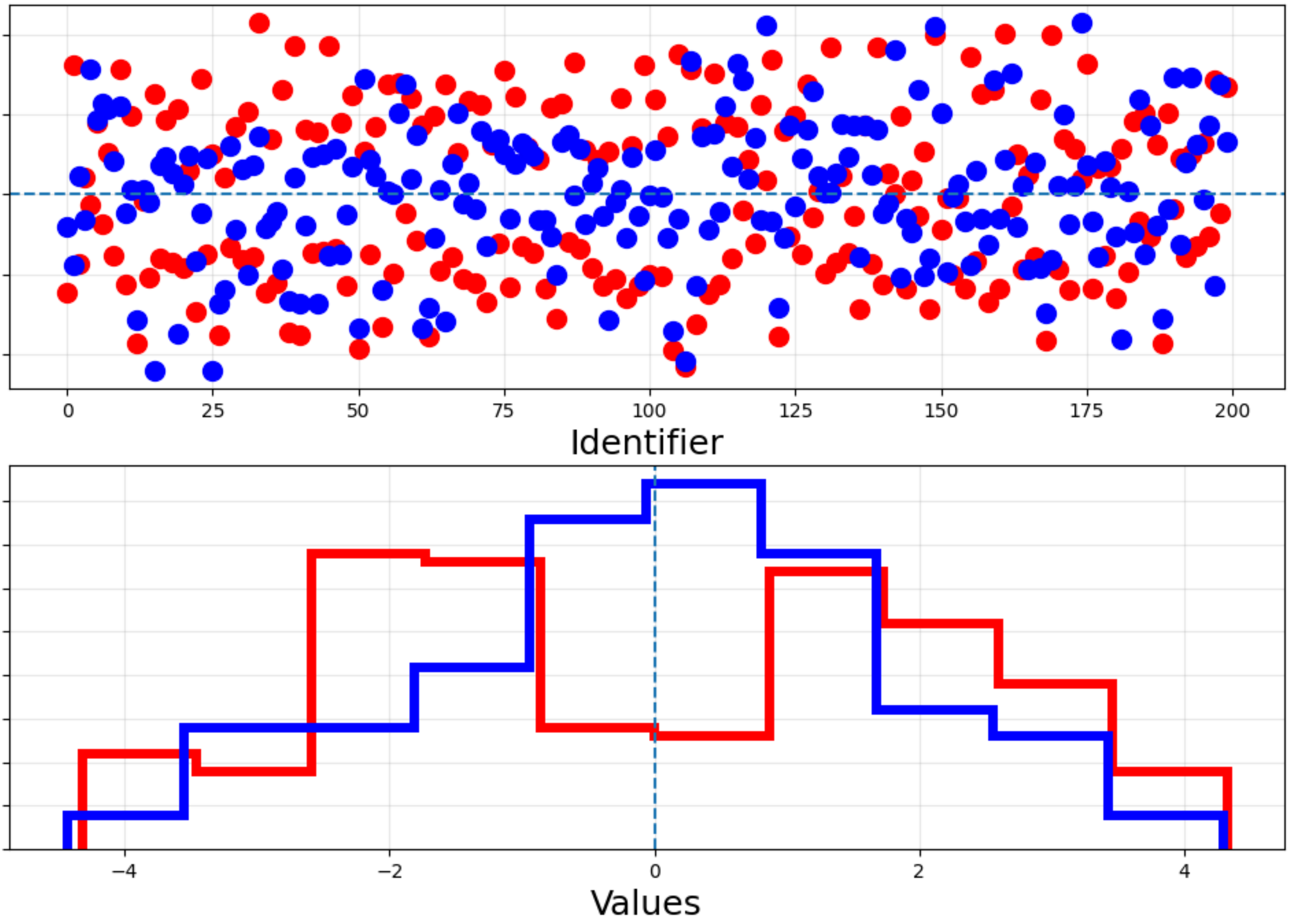
mean $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$



Let's define $d = \sum_{n=1}^{M} (x_n - \mu)$

Looks like it is not a good measure, we are not able to distinguish trivial cases

Let's define $s = \sum_{n=1}^{M} (x_n - \mu)^2$



$\mu_1 \approx 0$ and $\mu_2 \approx 0$

$d_1 \approx 0$ and $\mathrm{d}_2 \approx 0$

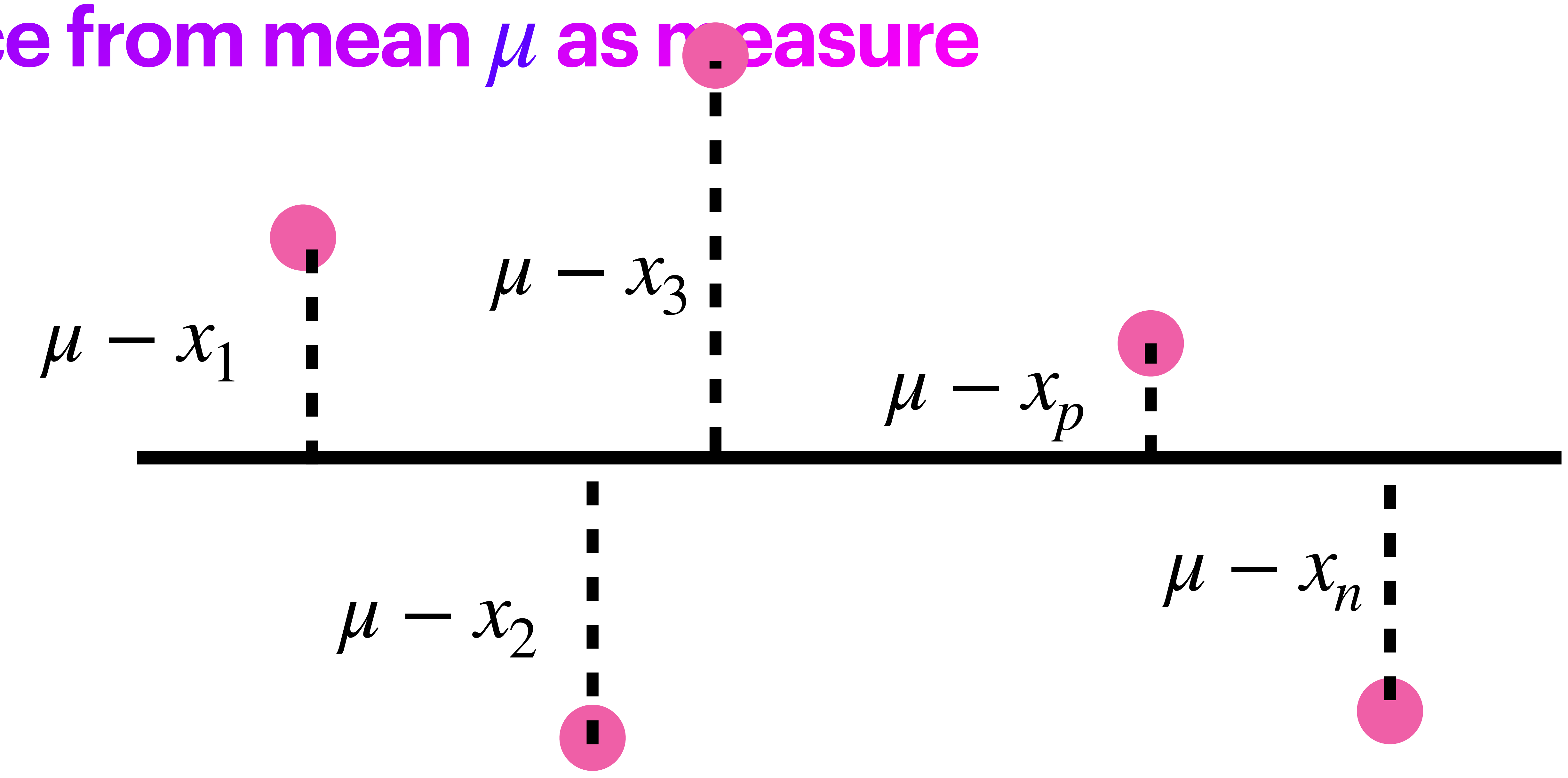
$s_1 \approx 1056$

$s_2 \approx 570$

Let's use distance from mean μ as measure

mean $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$

Let's define $d = \sum_{n=1}^{N} (x_n - \mu)$



Looks like it is not a good measure, we are not able to distinguish trivial cases

Let's define $\sigma^2 = \frac{1}{N}\sum_{n=1}^{N} (x_n - \mu)^2$

σ^2 is called variance of a distribution

Some of the useful measures

mean $\mu = \dfrac{1}{N} \sum_{n=1}^{N} x_n$

Median $m = x_{min} + 0.5 \left(x_{max} - x_{min} \right)$

variance $\sigma^2 = \dfrac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2$

The standard deviation of a set of measurements is equal to the positive square root of the variance.

Back to Class test example