ST 516: Foundations of Data Analytics

Means

Mean (Average) of a Set of Numbers

Recall the definition of the **average** or **mean** of a set of numbers $\{x_1, x_2, \ldots, x_n\}$:

Definition

The (arithmetic) **mean** (or **average**) of a set of numbers $\{x_1, x_2, \ldots, x_n\}$ is

$$\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

That is, we find the mean by adding up all the values in our set, and then dividing by the number of values in the set.

We use the notation $\overline{x}$ to denote the average, so we write

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Mean of a Set of Numbers: Example

Example: What is the mean of the set of numbers $\{1, 4, -3, -10\}$?

$$\overline{x} = \frac{1 + 4 + (-3) + (-10)}{4} = \frac{-8}{4} = -2$$

Mean of a Set of Numbers: Example

Example: What is the mean of the set of numbers $\{1, 4, -3, -10\}$?

$$\overline{x} = \frac{1 + 4 + (-3) + (-10)}{4} = \frac{-8}{4} = -2$$

Example: What is the mean of the set of numbers $\{2, 8, -6, -20\}$?

Notice that this new set of numbers is just 2 times each of the values in the previous set. Can you guess what the mean will be without calculating it?

$$\overline{x} = \frac{2 + 8 + (-6) + (-20)}{4} = \frac{-16}{4} = -4 = 2 \times -2$$

Mean of a Set of Numbers: Example

Example: What is the mean of the set of numbers $\{1, 4, -3, -10\}$?

$$\overline{x} = \frac{1 + 4 + (-3) + (-10)}{4} = \frac{-8}{4} = -2$$

Mean of a Set of Numbers: Example

Example: What is the mean of the set of numbers $\{1, 4, -3, -10\}$?

$$\overline{x} = \frac{1 + 4 + (-3) + (-10)}{4} = \frac{-8}{4} = -2$$

Example: What is the mean of the set of numbers $\{11, 14, 7, 0\}$?

Notice that this new set of numbers is just 10 plus each of the values in the previous set. Can you guess what the mean will be without calculating it?

$$\overline{x} = \frac{11 + 14 + 7 + 0}{4} = \frac{32}{4} = 8 = 10 + -2$$

Properties of Means

The examples on the previous slides illustrated the following property:

If the mean of

$$\{x_1, x_2, \ldots, x_n\}$$

is $\overline{x}$, then the mean of

$$\{ax_1 + b,\ ax_2 + b,\ \ldots, ax_n + b\}$$

is $a\overline{x} + b$.

That is, if we multiply each number in our set by the same value a, and then add to each number the same value b, the mean of the new set is equal to a times the mean of the old set plus b.

Sample Mean

Suppose we perform an experiment where we gather n experimental units (members of a population of interest) and measure a variable X on each unit, so we have observations $X_1, X_2, \ldots, X_n$.

- X_1 is the value of the variable X for the first unit selected
- X_2 is the value of the variable X for the second unit selected
- ... and so on.

The **sample mean** for a the sample of observations $X_1, X_2, \ldots, X_n$ is the mean of the values observed in that particular sample.

We often use the notation $\overline{X}$ to denote the sample mean of a sample of random variables.

Sample Mean: Example

Suppose we randomly select an adult who lives in New York City, and we measure the variable

$$X = \text{Amount (\$) that person spent on taxis during the past year}$$

If we obtain a sample of $n = 30$ randomly selected New Yorkers and measure X for each one, the sample mean is the average amount that those 30 New Yorkers spent on taxis over the last year.

Note that if we repeat the experiment by getting a different sample of 30 New Yorkers, we would get a different sample mean.

The sample mean is a *random variable*: it takes on different values depending on the specific random sample obtained. We can therefore consider the *probability distribution of the sample mean*.

Population Mean

The **population mean** of a random variable X is the mean of the values of that variable for the entire population.

We imagine that we could collect all members of the population of interest, and measure the value of the variable X for each individual member. Then we would take the average of all of those values.

The Greek letter μ is often used to denote the population mean.

Population Mean: Example

Suppose we randomly select an adult who lives in New York City, and we measure the variable

X = Amount (\$) that person spent on taxis during the past year

The population mean μ of X could be found by

- Collecting the amounts that *each adult* in New York City spent on taxis over the last year
- Adding up *all* of these amounts
- Dividing that sum by the *total number* of adults in New York City

Population Mean

The population mean is also referred to as the **expectation** or **expected value** of the variable.

We often use the notation

$$\mu = \mathsf{E}(X)$$

to denote the expected value (population mean) of the variable X.

Population Mean/Expected Value

Another way we could think of the population mean is as the long-run average over many repeat observations of the variable X:

- Suppose we randomly sample a value of X many, many times, where each observed value is independent of the others
- As we get more and more observations, the average of our sample of observations will get closer and closer to the population mean μ.

This is another expression of the Law of Large Numbers:

As the sample size n gets larger and larger, the sample mean of n independent observations gets closer and closer to the population mean μ.

Population Mean/Expected Value

Why does it make sense to call μ (the population mean) the expected value or expectation of X? Recall the taxi expenditure example:

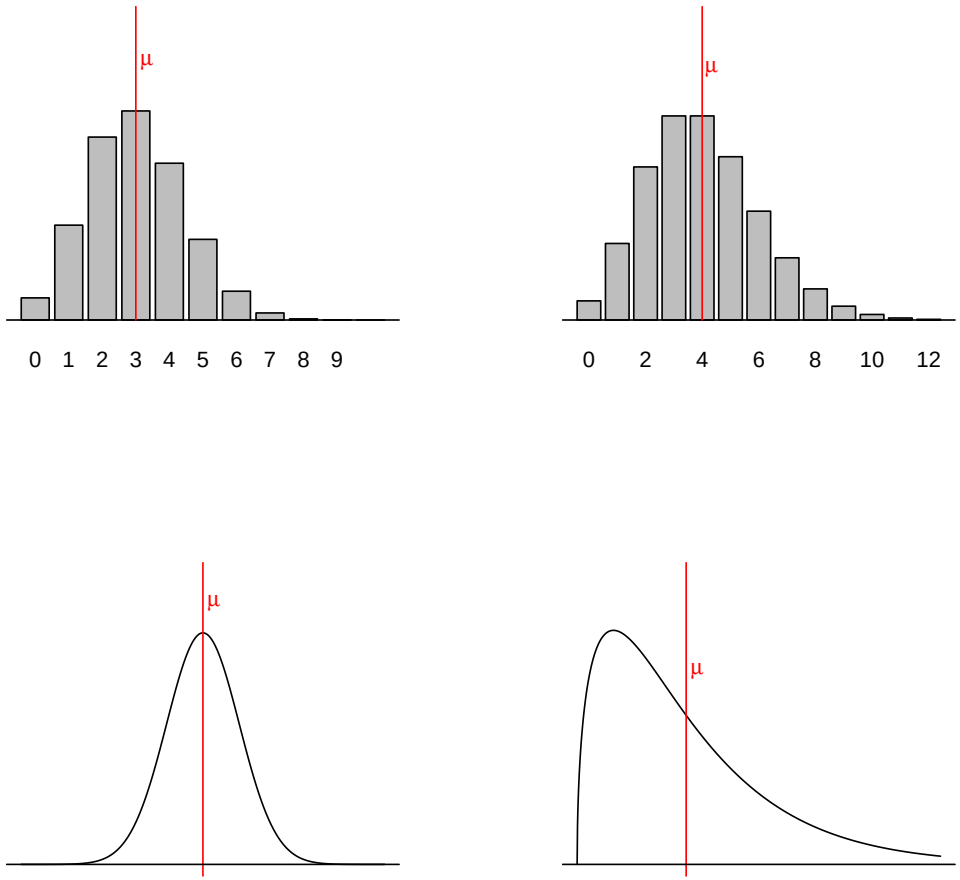
Randomly sample an adult who lives in New York City, and measure the variable

$X =$ Amount that person spent on taxis (\$) during the past year

Now imagine that a generous billionaire decides to 'sponsor' 100 people's taxi costs for last year. If she chooses 100 people at random, how much should she *expect* to spend?

Population Mean Examples

Here are some examples of population distributions with their corresponding population means:



Population Mean Examples

For certain population distribution families (like the Binomial, Poisson, Normal), we know the population mean: it is a function of the parameters that we specify to describe the distribution.

- If a random variable X has the Binomial(n, p) distribution, the population mean of X is

$$\mu = \mathsf{E}(X) = np$$

- If a random variable X has the Poisson(λ) distribution, the population mean of X is

$$\mu = \mathsf{E}(X) = \lambda$$

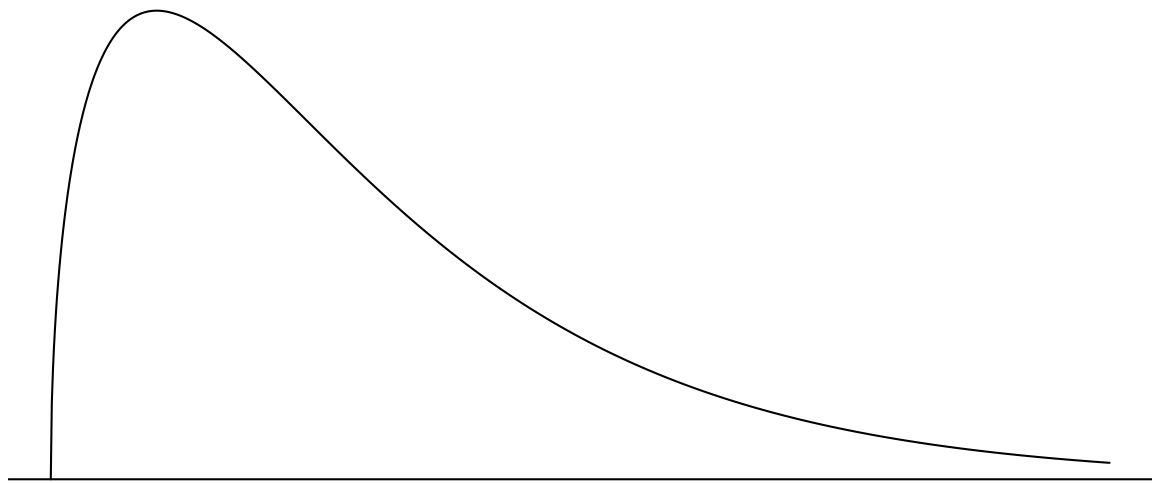
- If a random variable X has the Normal(μ, σ^2) distribution, the population mean of X is

$$\mathsf{E}(X) = \mu$$

  (This is why we use the notation 'μ' to denote the first parameter of a normal distribution!)

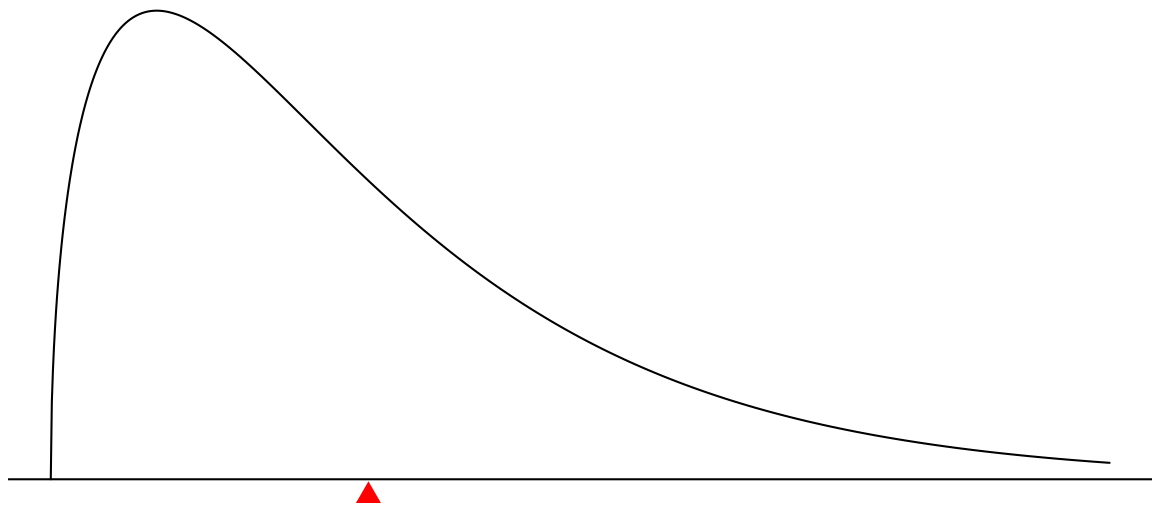
Population Mean Interpretation

The population mean is the *balancing point* of the population distribution: imagine that you cut out the shape of the population distribution from a piece of cardboard, and you were trying to balance that shape on a single point of the x-axis.



Population Mean Interpretation

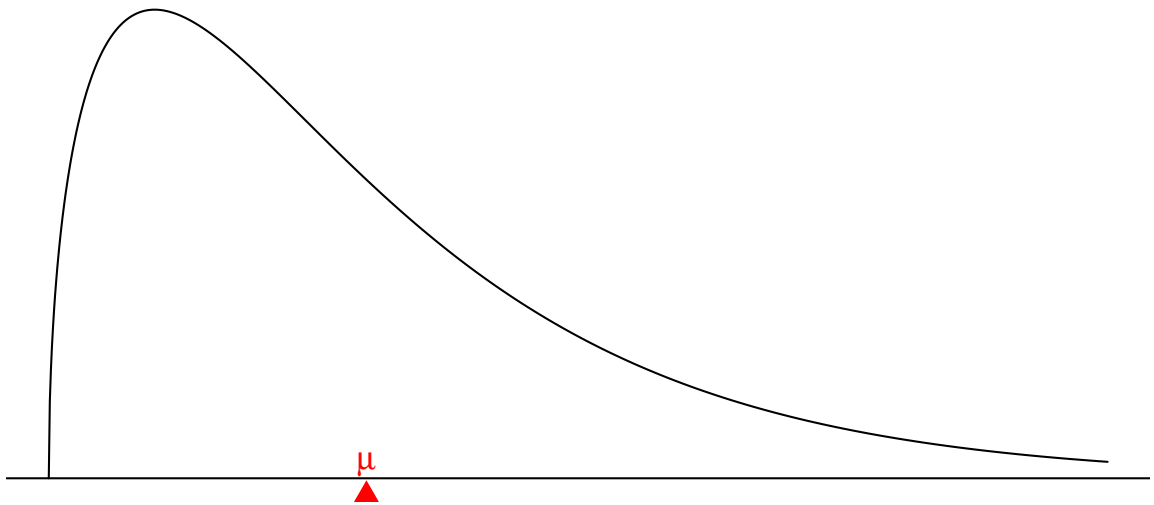
The population mean is the *balancing point* of the population distribution: imagine that you cut out the shape of the population distribution from a piece of cardboard, and you were trying to balance that shape on a single point of the x-axis.



Population Mean Interpretation

The population mean is the *balancing point* of the population distribution: imagine that you cut out the shape of the population distribution from a piece of cardboard, and you were trying to balance that shape on a single point of the x-axis.

The distribution would balance perfectly right at the population mean μ.



Population Mean Interpretation

The population mean is a *location parameter*: It tells us about *typical* values from that population distribution.

There are other location parameters that we might also consider:

- The **population median** is the "middle" value of the population values: 50% of the population has a larger value than the median, and 50% of the population has a smaller value than the median.
- The **population mode** is the most common value in the population.

You will see more about these other location parameters in an upcoming lecture.