# Ec140 - Core Concepts from Statistics

Fernando Hoces la Guardia
06/23/2022

# Housekeeping

- Updated Syllabus
- Unofficial Course Capture!
- What is the weirdest concept you remember from yesterday?
- Switch to finish yesterday's slides

# This Lecture

- Introduction to Data
- Mean and Expectation
- Variance and Standard Deviation

# What Defines a Data Set?

- Data Set is the collection of any type information (of multiple *Datum*)
- In quantitative analysis we focus on *structured* data sets (unlike, for example, unstructured field notes).
- In econometrics the most commnon way to structure data is in tabular, or rectangular, form.
- A tabular data set is a collection of variables that with information for one or more entities.
- Entities can represent multiple individuals, one individual over time, firms, countries, etc.
- Variables are represented in columns, and observations are represented by rows. (for more on variables The Effect, Ch3)

# Data

# But What Can We Do With Data?

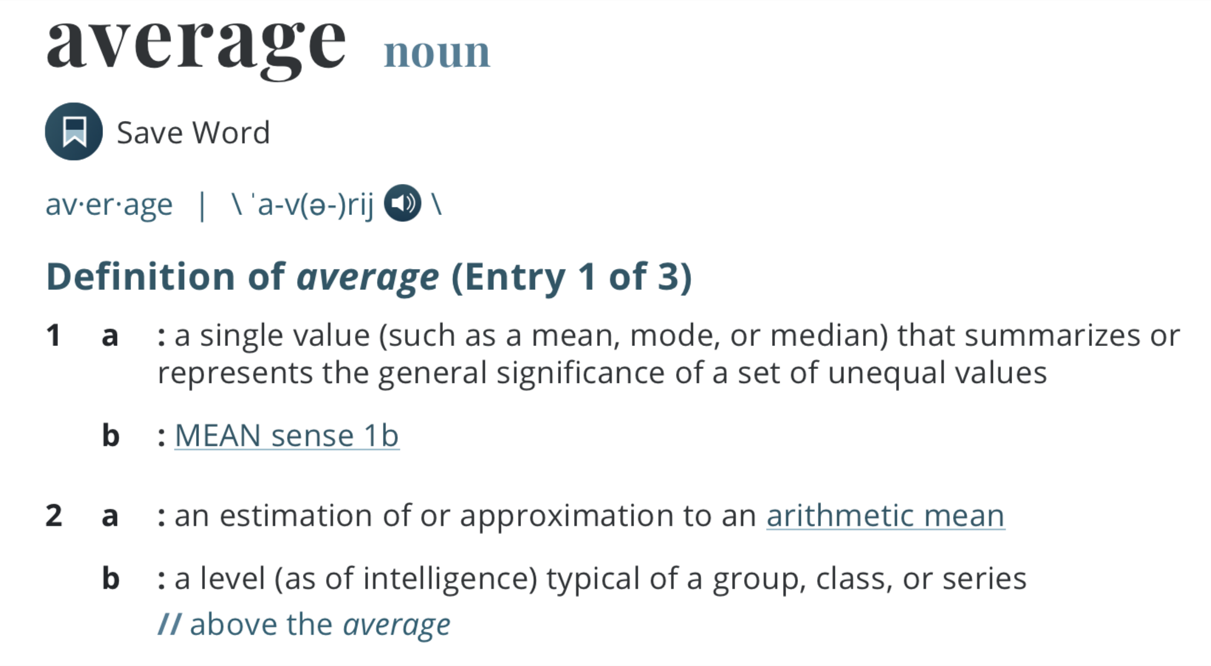
- We summarized it! (see the great short story by J.L. Borges on why summarizing is essential)

- One of the first thing we do when summarizing data is to look at *some type of average*.
  - Wait? *Type* of average? Isn't there just one average? called *the mean*?

# But What Can We Do With Data?

- We summarized it! (see the great short story by J.L. Borges on why summarizing is essential)

- One of the first thing we do when summarizing data is to look at *some type of average*.
  - Wait? *Type* of average? Isn't there just one average? called *the mean*?
- These is also referred as measure of central tendency.
- In this course, we will focus primarily on the mean. **From now on in this course**



**average** noun

Save Word

av·er·age | \ ˈa-v(ə-)rij \

**Definition of *average* (Entry 1 of 3)**

**1** **a** : a single value (such as a mean, mode, or median) that summarizes or represents the general significance of a set of unequal values

**b** : MEAN sense 1b

**2** **a** : an estimation of or approximation to an arithmetic mean

**b** : a level (as of intelligence) typical of a group, class, or series

// above the *average*

# Mean

- The mean is defined by the sum of a set of values divided by the number of values.

Let's look at the mean from the "hang out with a friend" exercise.

- Total over N

$$Average(X) = \frac{1 \times 10 + 2 \times 9 + 3 \times 11}{30} = 2.03$$

- One number, **highly informative** for a variable of interest.
- Always important to keep an eye on the units and magnitude (relevant for PS1).

# Mean: Notation (Message to me: draw histogram on the board)

$$Average(X) = \frac{1 \times 10 + 2 \times 9 + 3 \times 11}{30} = 2.03$$

$$Ave(X) = 1 \times \frac{10}{30} + 2 \times \frac{9}{30} + 3 \times \frac{11}{30} = 2.03$$

# Expected Value

- Let’s look at the histogram for the exercise above (drawn in the board) and pretend it is not a sample but the entire population. How can we move from frequencies into probabilities?
- Replace frequencies by probabilities
- The population version of the sample mean is the **expected value**.

# Expected Value: Definition (Discrete)

The expected value of a discrete random variable $X$ is the weighted average of its $k$ values $\{x_1, \ldots, x_k\}$ and their associated probabilities:

$$\begin{aligned}\mathbb{E}(X) &= x_1\,\mathbb{P}(X = x_1) + x_2\,\mathbb{P}(X = x_2) + \cdots + x_k\,\mathbb{P}(X = x_N) \\ &= \sum_x x\,\mathbb{P}(X = x)\end{aligned}$$

- Also known as the population mean.

# Expected Value: Definition (Discrete)

The expected value of a discrete random variable $X$ is the weighted average of its $k$ values $\{x_1, \ldots, x_k\}$ and their associated probabilities:

$$\begin{aligned}\mathbb{E}(X) &= x_1\,\mathbb{P}(X = x_1) + x_2\,\mathbb{P}(X = x_2) + \cdots + x_k\,\mathbb{P}(X = x_k)\\ &= \sum_x x\,\mathbb{P}(X = x) = \sum_x x\,f(x)\end{aligned}$$

- Also known as the population mean. Compare it to the sample mean:

$$\overline{X}_n = \sum_x x \times prop_n(x_1)$$

# Expected Value

## Example

Rolling a six-sided die once can take values $\{1, 2, 3, 4, 5, 6\}$, each with equal probability. What is the expected value of a roll?

$$\mathbb{E}(\mathrm{Roll}) = 1 \times \tfrac{1}{6} + 2 \times \tfrac{1}{6} + 3 \times \tfrac{1}{6} + 4 \times \tfrac{1}{6} + 5 \times \tfrac{1}{6} + 6 \times \tfrac{1}{6} = 3.5.$$

- **Note:** The expected value can be a number that isn't a possible outcome of $X$.

# Expected Value. Definition (Continuous)

- Compare it to the discrete version
- Continuous

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

- Discrete

$$\mathbb{E}(X) = \sum_{x} x f(x)$$

# Expected Value. Definition (Continuous)

- Compare it to the discrete version
- Continuous

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

- Discrete

$$\mathbb{E}(X) = \sum_{x} x f(x)$$

This explanation was inspired by
this lecture from Eddie Woo

# Expected Value. Definition. One Last Thing 1/2

Let's go back to the mean of our exercise:

$$\overline{X}_n = 1 \times \frac{10}{30} + 2 \times \frac{9}{30} + 3 \times \frac{11}{30} = 2.03$$

But now let's switch the values of the random variables to: 10, 20, 30. How should we compute the mean?

$$\overline{g(X)}_n = 10 \times \frac{10}{30} + 20 \times \frac{9}{30} + 30 \times \frac{11}{30} = 20.33$$

# Expected Value. Definition. One Last Thing 2/2

Hence, we can conclude, that for a random variable $X$, any transformation $g(X)$ has a sample aveage:

$$\overline{X}_n = \sum_x g(x) \times prop_n(x_1)$$

And an expectation:

$$\mathbb{E}(X) = \sum_x g(x) f(x)$$

The same idea applies in the case of a continues random variable

# Expected Value: Rules (or Properties)

## Rule 1

For any constant $c$, $\mathbb{E}(c) = c$.

## Not-so-exciting examples

$\mathbb{E}(5) = 5$.

$\mathbb{E}(1) = 1$.

$\mathbb{E}(4700) = 4700$.

# Expected Value

## Rule 2

For any constants $a$ and $b$, $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$.

## Example

Suppose $X$ is the high temperature in degrees Celsius in Eugene during August. The long-run average is $\mathbb{E}(X) = 28$. If $Y$ is the temperature in degrees Fahrenheit, then $Y = 32 + \frac{9}{5}X$. What is $\mathbb{E}(Y)$?

- $\mathbb{E}(Y) = 32 + \frac{9}{5}\,\mathbb{E}(X) = 32 + \frac{9}{5} \times 28 = 82.4$.

# Expected Value

## Rule 3: Linearity

If $\{a_1, a_2, \ldots, a_n\}$ are constants and $\{X_1, X_2, \ldots, X_n\}$ are random variables, then

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1 \, \mathbb{E}(X_1) + a_2 \, \mathbb{E}(X_2) + \cdots + a_n \, \mathbb{E}(X_n).$$

In English, the expected value of the sum = the sum of expected values.

# Expected Value

## Rule 3

The expected value of the sum = the sum of expected values.

## Example

Suppose that a coffee shop sells $X_1$ small, $X_2$ medium, and $X_3$ large caffeinated beverages in a day. The quantities sold are random with expected values $\mathbb{E}(X_1) = \mathbf{43}$, $\mathbb{E}(X_2) = \mathbf{56}$, and $\mathbb{E}(X_3) = \mathbf{21}$. The prices of small, medium, and large beverages are $\mathbf{1.75}$, $\mathbf{2.50}$, and $\mathbf{3.25}$ dollars. What is expected revenue?

$$
\begin{aligned}
\mathbb{E}(1.75X_1 + 2.50X_2 + 3.35X_n) &= 1.75\,\mathbb{E}(X_1) + 2.50\,\mathbb{E}(X_2) + 3.25\,\mathbb{E}(X_3) \\
&= 1.75(43) + 2.50(56) + 3.25(21) \\
&= 283.5
\end{aligned}
$$

# Expected Value

## Caution

Previously, we found that the expected value of rolling a six-sided die is $\mathbb{E}(\mathrm{Roll}) = \mathbf{3.5}$.

- If we square this number, we get $[\mathbb{E}(\mathrm{Roll})]^2 = \mathbf{12.25}$.

**Is $[\mathbb{E}(\mathrm{Roll})]^2$ the same as $\mathbb{E}\left(\mathrm{Roll}^2\right)$?**

**No!**

# Expected Value

## Caution

Except in special cases, the transformation of an expected value **is not** the expected value of a transformed random variable.

For some function $g(\cdot)$, it is typically the case that

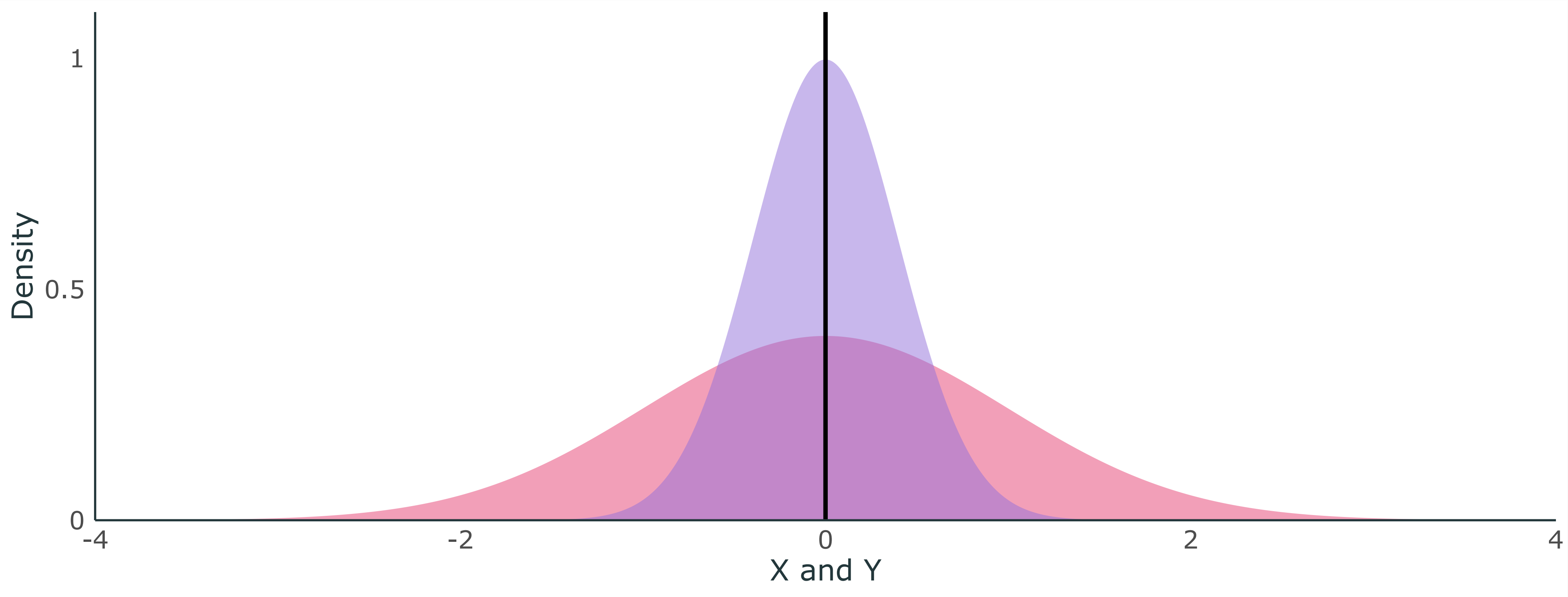
$$g\left(\mathbb{E}(X)\right) \neq \mathbb{E}(g(X)).$$

# Activity 1

- Let's watch another Stat 110's video. Then get together in groups of 3 and discuss:
  - Don't worry about the law of large numbers yet
  - How does the random variables becomes continuous?
  - How does linearity help with computations?

# Variance

Random variables $X$ and $Y$ share the same population mean, but are distributed differently.



# Variance

How tightly is a random variable distributed about its mean?

- Let $\mu = \mathbb{E}(X)$.
- Describe the distance of $X$ from its population mean $\mu$ as the squared difference: $(X - \mu)^2$.

**Variance** tells us how far $X$ deviates from $\mu$, *on average*:

$$\mathbf{Var}(X) \equiv \mathbb{E}\big((X - \mu)^2\big) = \sigma^2$$

- $\sigma^2$ is shorthand for variance.

# Variance

## Rule 1

$\mathbf{Var}(X) = \mathbf{0} \iff X$ is a constant.

- If a random variable never deviates from its mean, then it has zero variance.
- If a random variable is always equal to its mean, then it's a (not-so-random) constant.

# Variance

## Rule 2

For any constants $a$ and $b$, $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$.

## Example

Suppose $X$ is the high temperature in degrees Celsius in Eugene during August. If $Y$ is the temperature in degrees Fahrenheit, then $Y = 32 + \frac{9}{5}X$. What is $\mathrm{Var}(Y)$?

- $\mathrm{Var}(Y) = (\frac{9}{5})^2 \,\mathrm{Var}(X) = \frac{81}{25} \,\mathrm{Var}(X)$.

# Standard Deviation

**Standard deviation** is the positive square root of the variance:

$$\mathrm{sd}(X) = +\sqrt{\mathrm{Var}(X)} = \sigma$$

- $\sigma$ is shorthand for standard deviation.

# Standard Deviation

## Rule 1

For any constant $c$, $\mathrm{sd}(c) = 0$.

## Rule 2

For any constants $a$ and $b$, $\mathrm{sd}(aX + b) = |a| \, \mathrm{sd}(X)$.

# Standardizing a Random Variable

When we're working with a random variable $X$ with an unfamiliar scale, it is useful to **standardize** it by defining a new variable $Z$:

$$Z \equiv \frac{X - \mu}{\sigma}.$$

$Z$ has mean $0$ and standard deviation $1$. How?

- First, some simple trickery: $Z = aX + b$, where $a \equiv \frac{1}{\sigma}$ and $b \equiv -\frac{\mu}{\sigma}$.
- $\mathbb{E}(Z) = a\,\mathbb{E}(X) + b = \mu\frac{1}{\sigma} - \frac{\mu}{\sigma} = 0$.
- $\mathrm{Var}(Z) = a^2\mathrm{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1$.