



UC San Diego

# Econometrics 120A: Discussion Section

Week 1

---

Natalia Madrid & Lapo Bini

Department of Economics

Some Stuff

---

## About the course organization

- ⇒ **Discussion sections:** on Monday (from 3:00 pm to 7:00 pm)
- ⇒ **Piazza for questions:** available on Canvas
- ⇒ **Problem sets:** the first one is due October 9th
- ⇒ **Office hours:** every day of the week (from 12:30 pm to 1:30 pm)

## About the office hours

### Natalia Madrid

- ⇒ Office hours: Tuesdays, 12:30 pm to 1:30 pm
- ⇒ E-mail: [namadrid@ucsd.edu](mailto:namadrid@ucsd.edu)

### Lapo Bini

- ⇒ Office hours: Monday, 12:30 pm to 1:30 pm
- ⇒ E-mail: [lbini@ucsd.edu](mailto:lbini@ucsd.edu)
- ⇒ Website: [▶ Lapo Bini](#)

# Motivation for the course

## The Nobel Prize: We can all succeed in different ways

"My mind is much slower than most of my colleagues' minds and I discovered that when I was an undergraduate.

I struggled for the first year and a half as an undergraduate at Caltech... but I developed my own ways of learning things. Keeping records of what I was learning, working things out in my own way.

So I advise students, **find your own way**, and experiment in mastering material and your own directions that you can be successful in and everybody is different. We can all succeed in different ways."

Kip Thorne, awarded the Nobel Prize in Physics 2017 for the detection of gravitational waves.

# Chapter 1: How Not to Do Plots

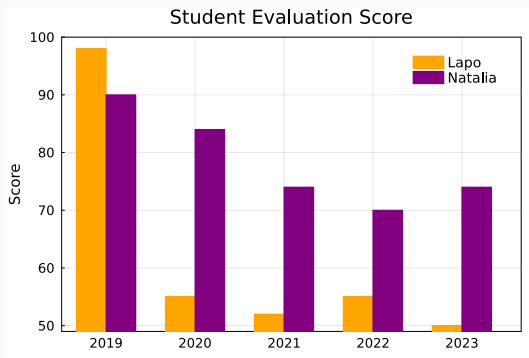
---

# Graphical Questions

## Kind of PS1 Question 1

ECON120TAs have been having problems over the last years, the following graph shows the dynamics of the student evaluation for Lapo and Natalia.

(a) The decline of Lapo has been larger than Natalia. By how much does it change?

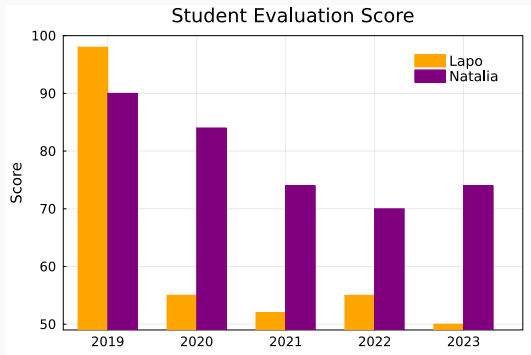


# Graphical Questions

## Kind of PS1 Question 1

ECON120TAs have been having problems over the last years, the following graph shows the dynamics of the student evaluation for Lapo and Natalia.

(b) Does the visual impression of the graph match with your results in (a)? Why or why not?



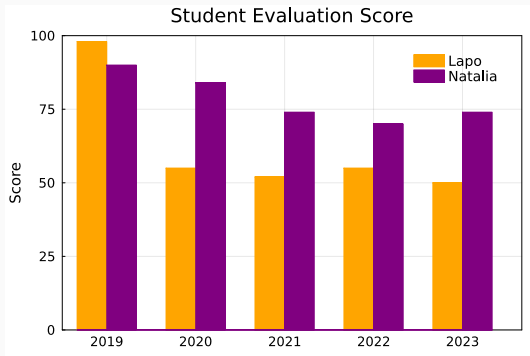


# Graphical Questions

## Kind of PS1 Question 1

ECON120TAs have been having problems over the last years, the following graph shows the dynamics of the student evaluation for Lapo and Natalia.

(b) Does the visual impression of this new version of the graph match with your results in (a)?

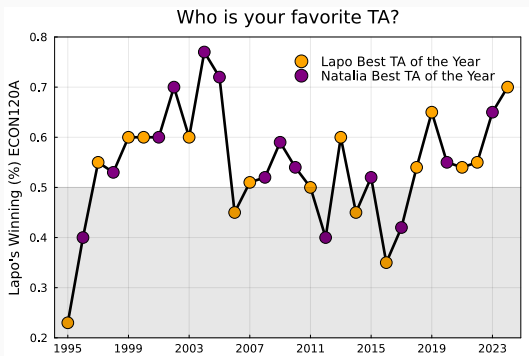


# Graphical Questions

## Kind of PS1 Question 2

The consensus is that even though Lapo is the dominant TA for ECON120A in terms of preferences from the students, this did not seem to amount to much when it came to the "Best TA of the Year" Prize. Assumption: Lapo and Natalia are the only TAs in the department.

(a) How often is Natalia the dominant TA for ECON120A?

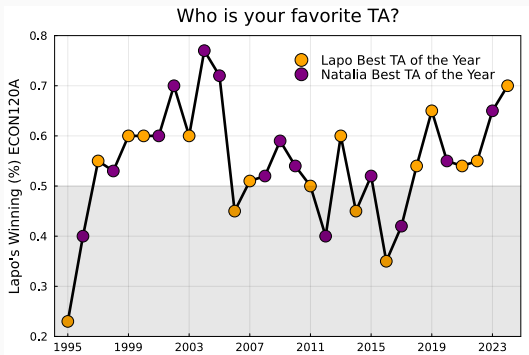


# Graphical Questions

## Kind of PS1 Question 2

The consensus is that even though Lapo is the dominant TA for ECON120A in terms of preferences from the students, this did not seem to amount to much when it came to the "Best TA of the Year" Prize. Assumption: Lapo and Natalia are the only TAs in the department.

(b) Does Lapo dominate the "Best TA of the Year" prize?

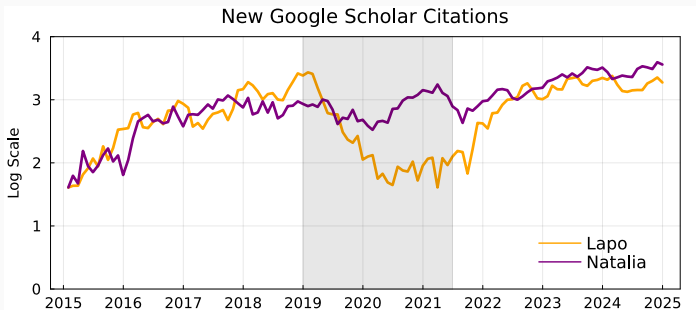


# Graphical Questions

## Kind of PS1 Question 3

The chart shows the number of new Google Scholar citations for Lapo and Natalia from January 2015 to January 2025. The goal is to study the impact of the bankruptcy (January 2019) of the journal where Lapo and Natalia used to publish their papers.

(a) How accurately does the figure portray the downturn visually?

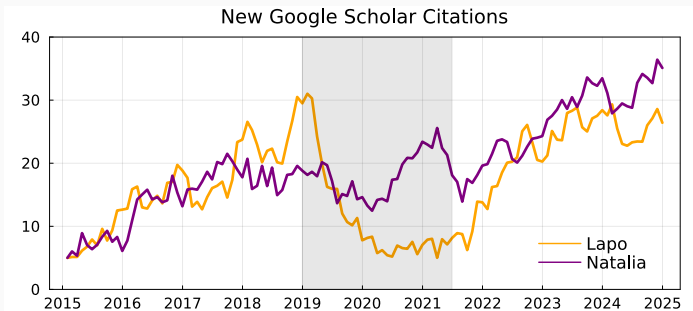


# Graphical Questions

## Kind of PS1 Question 3

The chart shows the number of new Google Scholar citations for Lapo and Natalia from January 2015 to January 2025. The goal is to study the impact of the bankruptcy (January 2019) of the journal where Lapo and Natalia used to publish their papers.

(a) What about now?



## Chapter 2: Descriptive Statistics

---

## Measures of Central Tendencies

- ⇒ **Sample Mean:** Given a sample made of  $n$  observations  $\{X_1, \dots, X_n\}$ , the sample mean is computed as:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- ⇒ **Sample Median:** this is the very middle observation in the data, separating the higher half from the lower half of a data sample.
- ⇒ **Sample Mode:** this is the most frequently observed value in the sample.

## Kind of PS1 Question 2.1

The chart shows the number of cars owned by 13 American families and their yearly income in thousands of US dollars.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13
# Cars	1	2	3	2	2	2	1	1	2	3	5	2	3
Incom	25	30	45	27	32	35	17	20	39	45	70	30	50

(a) Compute the average number of cars owned by those families.

$$\bar{X} =$$



# Mean Median Mode

## Kind of PS1 Question 2.1

(b) Let's compute the median number of cars per family. We will rearrange the entries of the previous table.

ID	1	7	8	2	4	5	6	9	12	3	10	13	11
# Cars	1	1	1	2	2	2	2	2	2	3	3	3	5

(c) What is the mode for the number of cars owned?

ID	1	7	8	2	4	5	6	9	12	3	10	13	11
# Cars	1	1	1	2	2	2	2	2	2	3	3	3	5

## Something useful for 2.2.7

We can rewrite the sample mean in terms of frequencies.

ID	1	7	8	2	4	5	6	9	12	3	10	13	11
# Cars	1	1	1	2	2	2	2	2	2	3	3	3	5

# Cars	1	2	3	5
Frequency	3/13	6/13	3/13	1/13

We can easily compute the mode by looking at the highest frequency. The sample mean simplifies too:

$$\bar{X} = \sum_{i=1}^F \omega_i X_i = \frac{3}{13}1 + \frac{6}{13}2 + \frac{3}{13}3 + \frac{1}{13}5$$

## Something useful for 2.2.6

Now we will study the role of big values (outliers) on the mean and median. Let's consider the average number of citations in the Economics Department at UCSD.

Category	Average Citations	(%) Less than Average	(%) with <1000 Citations
Professors	14000	55%	10%
TAs	3000	95%	56%

(a) How is it possible that in each category the average citation is larger than the actual number of citations received by the greater majority of people?

## Something useful for 2.2.6

Category	Average Citations	(%) Less than Average	(%) with <1000 Citations
Professors	14000	55%	10%
TAs	3000	95%	56%

Natalia has 7 citations, Georgy 5, Artur 6:  $\bar{X} = \frac{1}{3}(7 + 5 + 6) = 6$

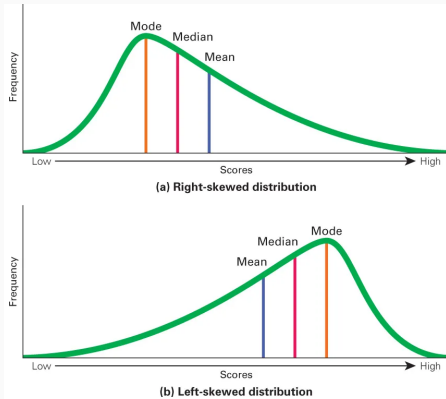
Now we add Lapo with 102 citations:  $\bar{X} = \frac{1}{4}(5 + 5 + 6 + 102) = 30$

(b) What is the median value for each category?

# Mean Median Mode

## Something useful for 2.2.2 and 2.2.6

If we want to report the "typical" number of citations, we should use the median. The distribution of citations needs to be skewed to the right. For skewed distributions, rarely is the sample mean representative, a more reasonable number to report would be the median.



# Variance, Standard Deviation & Covariance

## Definitions

- ⇒ **Variance:** it is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value. Given the sample  $\{X_1, \dots, X_n\}$ , it is:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

If instead of single observations we have the frequency, it becomes:

$$\sigma^2 = \sum_{i=1}^f \omega_i (x_i - \bar{X})^2$$

- ⇒ **Standard deviation:** positive square root of the variance  $\sigma = \sqrt{\sigma^2}$ .
- ⇒ **Covariance:** it is a measure of the relationship between two variables and to what extent they change together.

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}_i)$$

## Kind of PS1 Question 2.2.7

The students of ECON120A are asked to grade Lapo and Natalia on a scale 0 – 100. The students provided the following ranges:

Grade Range	Proportion of Students
[60; 70)	0.01
[70; 80)	0.1
[80; 90]	0.35
[90; 100]	0.54

(a) What is the average grade?

# Descriptive Statistics using Frequencies

## Kind of PS1 Question 2.2.7

The students of ECON120A are asked to grade Lapo and Natalia on a scale 0 – 100. The students provided the following ranges:

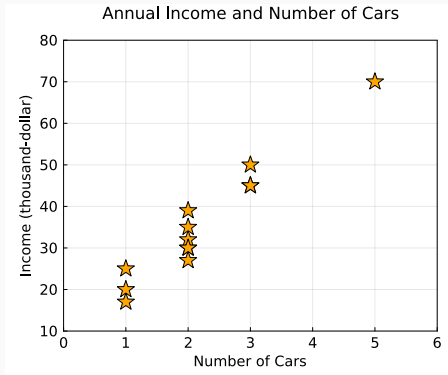
Grade Range	Proportion of Students
[60; 70)	0.01
[70; 80)	0.1
[80; 90]	0.35
[90; 100]	0.54

(a) What is the variance of the grade?



# Covariance

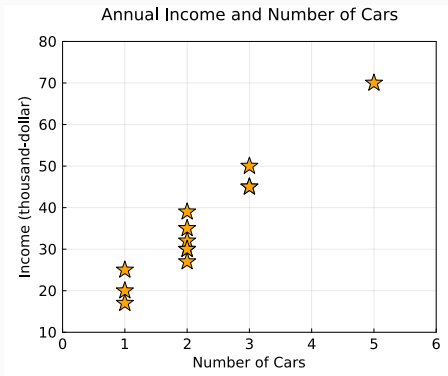
## Relationship between Cars and Income



What is the sign of the covariance between annual income and number of cars owned?

# Covariance

## Relationship between Cars and Income

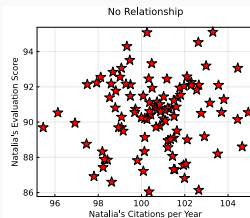
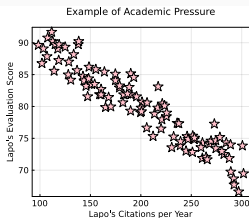
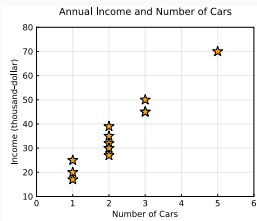


The covariance is **15.057692307692308**. We cannot conclude anything about how strong this relationship is, that is why we are going to introduce the notion of **correlation**.

# Correlation

⇒ **Correlation** is a statistical measure that expresses the extent to which two variables are linearly related:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$



Correlations are: 0.97, -0.94, 0.07 respectively.

# Covariance & Correlation

## Kind of PS1 Questions 2.2.3 and 2.2.8

We now study the relationship between the ECON120A TAs' student evaluation scores ( $X$ ) and their performance on the TOEFL English standardized test ( $Y$ ).

ID	Student Evaluation	TOEFL score
Lapo	80	100
Natalia	92	115
Artur	93	118
Georgy	86	110

Let's compute the correlation between the two variables. We will need  $\bar{X}$ ,  $\bar{Y}$ ,  $\sigma_X$ ,  $\sigma_Y$  and  $\sigma_{XY}$ .