

UC San Diego

Econometrics 120A: Discussion Section

Week 1

Linghui Wu, Anita Sammarini & Lapo Bini

Department of Economics

Introduction to 120A

Introduction to 120A

About the course organization

- ⇨ **Discussion sections**: on Fridays (from 8:00 am to 12:00 pm)
- ⇨ **Problem sets**: the first one is due October 9th
- ⇨ **Office hours**: Each TA holds weekly office hours

About the office hours

Each TA holds weekly office hours.

Linghui

- ⇨ Office hours: Monday, 10:00–11:00 am, Malk Hall 387
- ⇨ E-mail: liw036@ucsd.edu

Anita

- ⇨ Office hours: Wednesday, 1:00–2:00 pm, Malk Hall 387
- ⇨ E-mail: asammarini@ucsd.edu

Lapo

- ⇨ Office hours: Thursday, 1:00–2:00 pm, Malk Hall 487
- ⇨ E-mail: lbini@ucsd.edu

Chapter 1: How Not to Do Plots

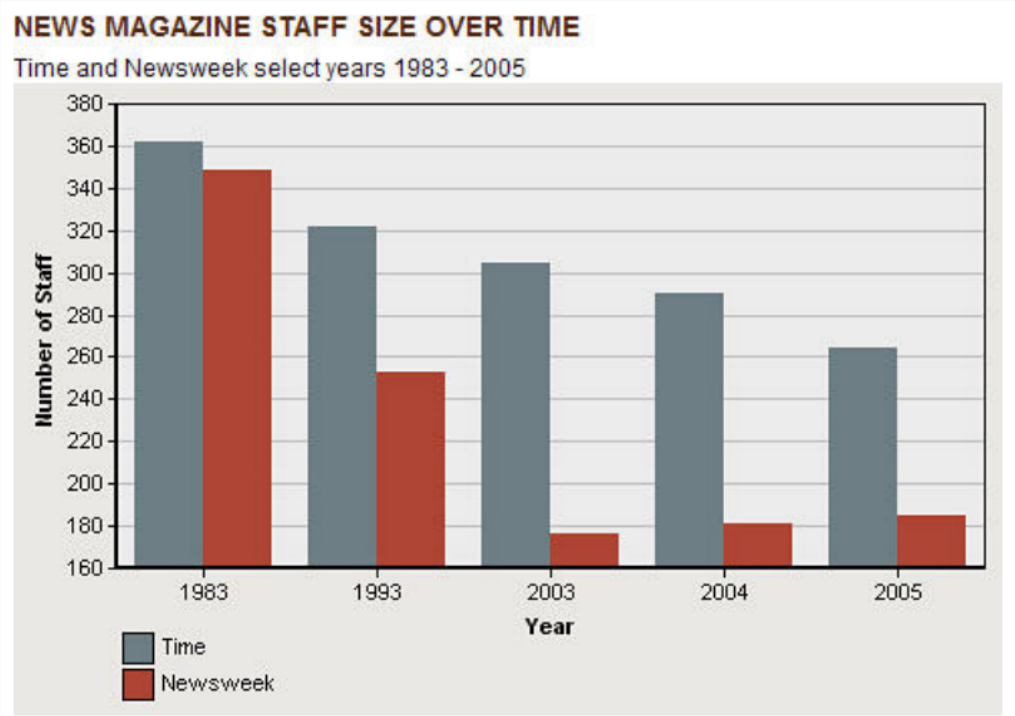
Graphical Questions

PS1 Question 1

Traditional media have been having problems over the last years. The graph shows the effect by showing the decrease in the number of employees in the magazine divisions at Time and Newsweek.

(a) The decline at Newsweek has been larger than at Time. By how much does it change?

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



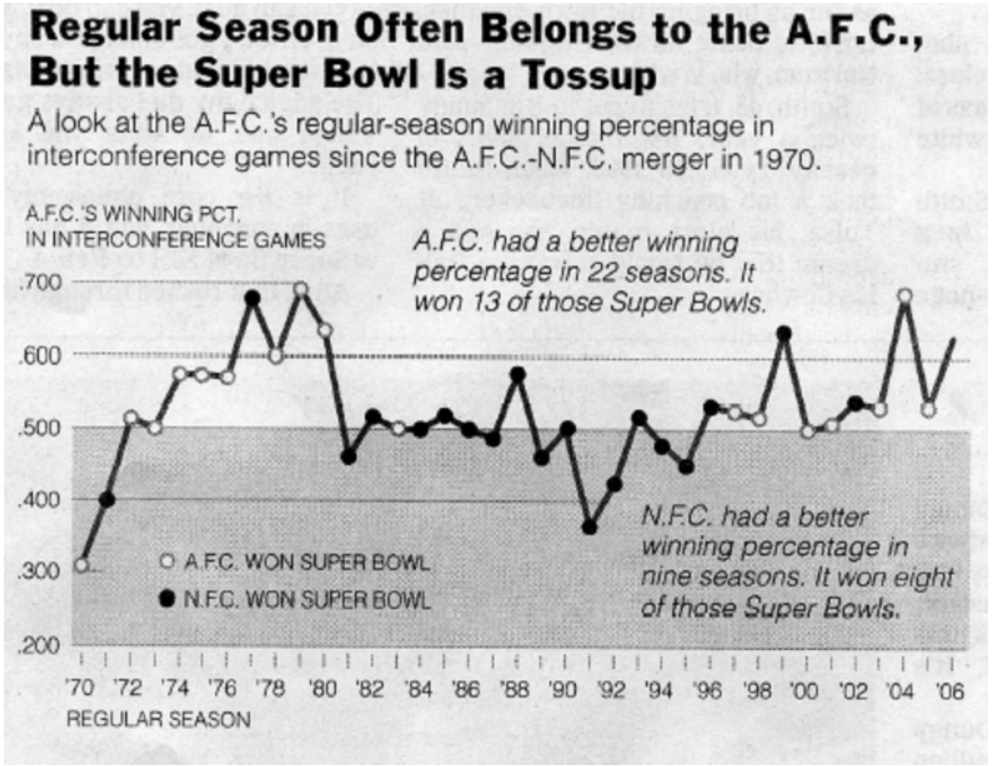
Graphical Questions

PS1 Question 2

Graph from the New York Times, 28 January 2007 (the Sunday before the Superbowl). The line is the AFC's regular season winning percentage against the NFC since 1970: above 0.500 the AFC had the better record. White dots: the AFC won the Superbowl; black dots: the NFC won.

The reporter's claim: the AFC dominates the regular season, but this does not amount to much in the Superbowl.

(a) How often is the NFC the dominant league?



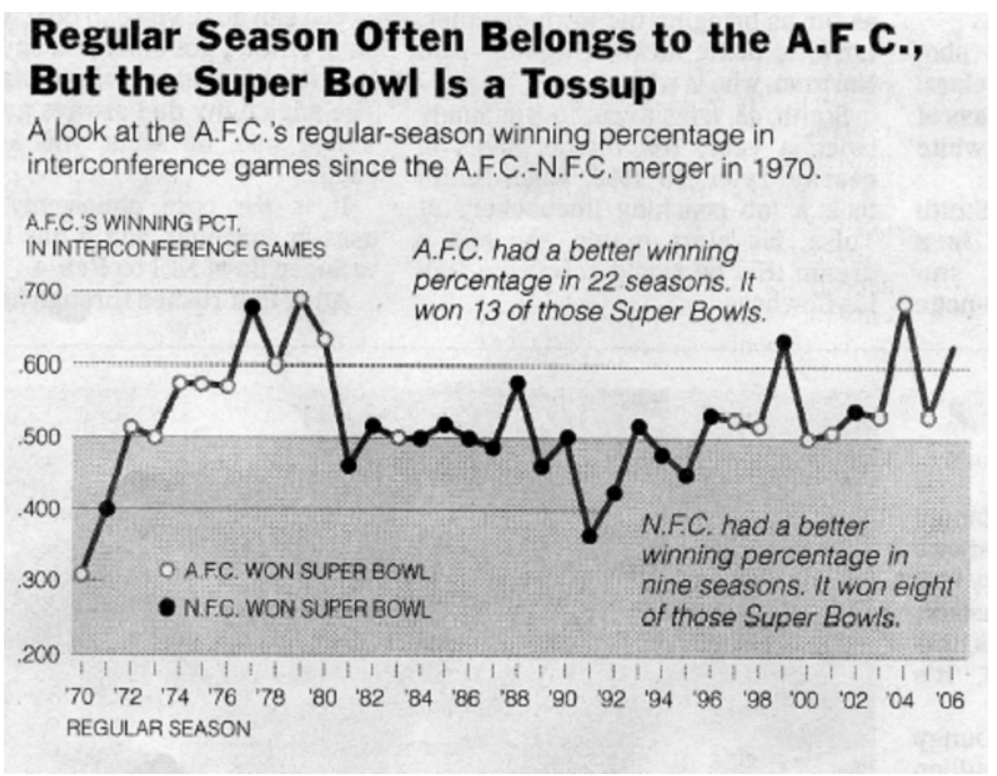
Graphical Questions

PS1 Question 2

Graph from the New York Times, 28 January 2007 (the Sunday before the Superbowl). The line is the AFC's regular season winning percentage against the NFC since 1970: above 0.500 the AFC had the better record. White dots: the AFC won the Superbowl; black dots: the NFC won.

The reporter's claim: the AFC dominates the regular season, but this does not amount to much in the Superbowl.

(b) Just by looking at the graph (no computations), what do you think the average winning percentage of the NFC is?



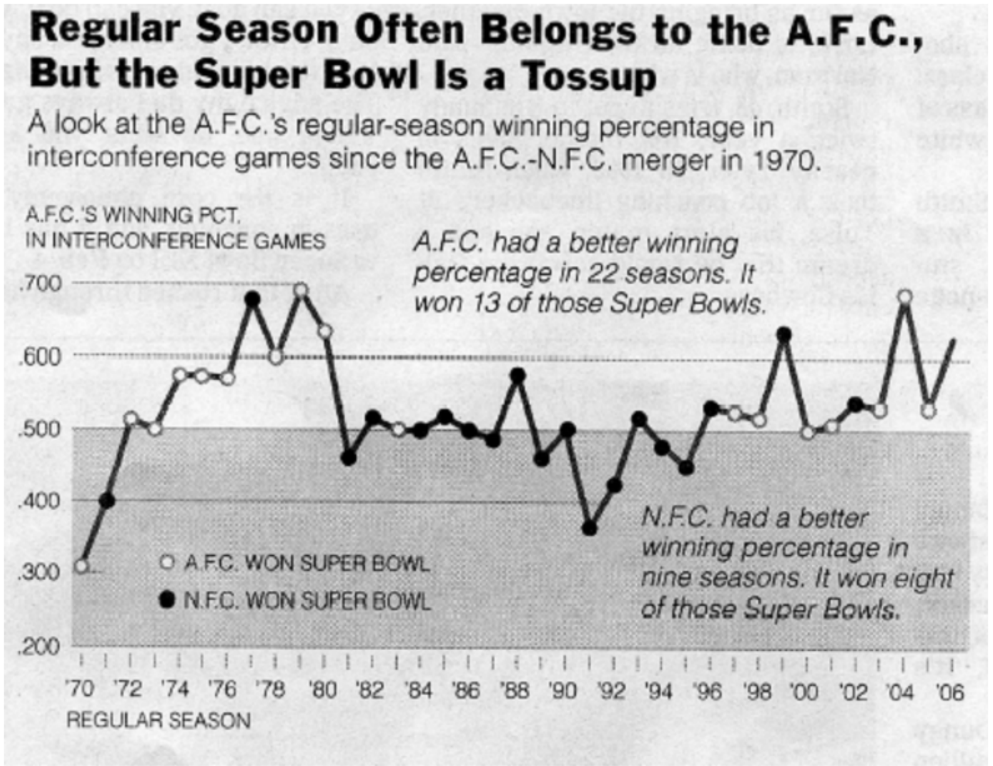
Graphical Questions

PS1 Question 2

Graph from the New York Times, 28 January 2007 (the Sunday before the Superbowl). The line is the AFC's regular season winning percentage against the NFC since 1970: above 0.500 the AFC had the better record. White dots: the AFC won the Superbowl; black dots: the NFC won.

The reporter's claim: the AFC dominates the regular season, but this does not amount to much in the Superbowl.

(c) Do these results indicate that the NFC dominates the Superbowl?



Chapter 2: Descriptive Statistics

Mean Median Mode

Measures of Central Tendencies

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

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

$$\bar{X} = \sum_{i=1}^{F} \omega_i X_i$$

⇨ **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.

PS1 Question 2.1.1

A childcare facility is eligible for a county grant if the average age of the children is below 9. The ages of the children are:

Child	1	2	3	4	5	6	7	8	9	10	11
Age	8	5	9	10	9	12	7	12	13	7	8

(a) Compute the mean, do they qualify?

$\bar{X} =$

PS1 Question 2.1.1

(b) What if the rule used the median age? We rearrange the ages from smallest to largest.

Child	2	7	10	1	11	3	5	4	6	8	9
Age	5	7	7	8	8	**9**	9	10	12	12	13

Descriptive Statistics using Frequencies

PS1 Question 2.2.7

A summary of economic forecasters gave the following predictions of the inflation rate next year.

Forecast	Proportion of Forecasters
2 to 4	0.12
4 to 6	0.60
6 to 8	0.23
8 to 10	0.05

(a) What is the average forecast? Suppose the inflation rate turned out to be 3.9%, how well does the average compare?

$$\bar{X} = \sum_{i=1}^{F} \omega_i X_i =$$

Variance & Standard Deviation

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 $\{X_1, \ldots, X_n\}$, it is:

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

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

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

⇨ **Standard deviation**: positive square root of the variance $S_X = \sqrt{S_X^2}$.

Descriptive Statistics using Frequencies

PS1 Question 2.2.7

A summary of economic forecasters gave the following predictions of the inflation rate next year.

Forecast	Proportion of Forecasters
2 to 4	0.12
4 to 6	0.60
6 to 8	0.23
8 to 10	0.05

(b) What is the standard deviation?

$$S_X = \sqrt{\sum_{i=1}^{F} \omega_i (X_i - \bar{X})^2} =$$

Covariance

Definition

⇨ **Covariance**: it is a measure of the relationship between two variables and to what extent they change together.

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

⇨ $S_{XY} > 0$: the two variables tend to move in the same direction.

$S_{XY} < 0$: they tend to move in opposite directions.

⇨ The covariance depends on the units of X and Y: it tells us the **direction** of the relationship, but nothing about how **strong** it is. That is why we introduce the correlation.

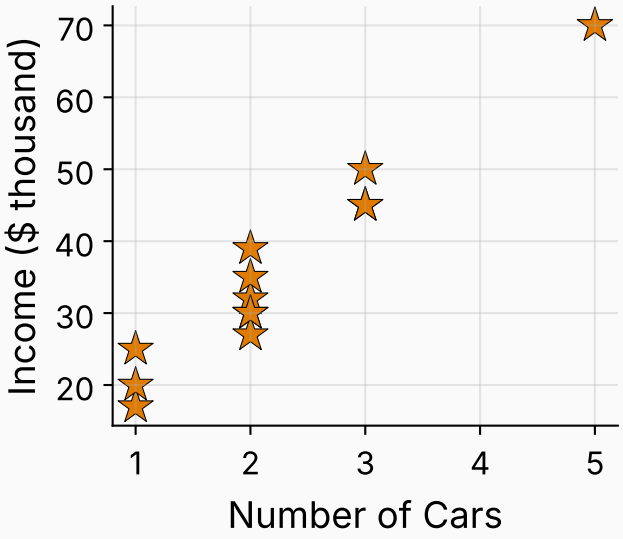
Correlation

Definition

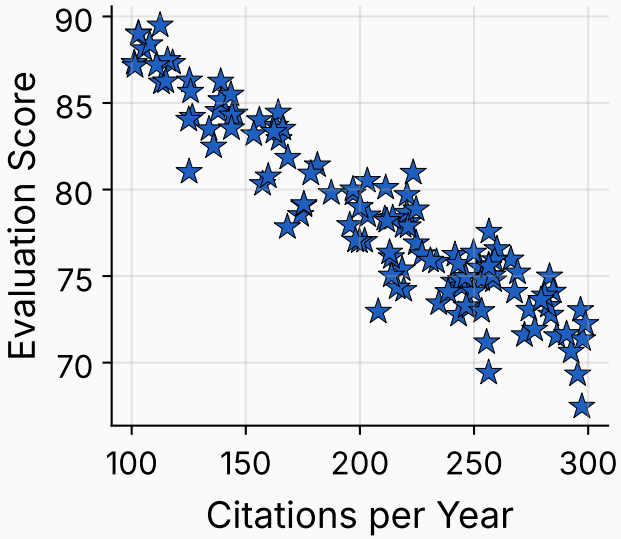
Correlation measures the extent to which two variables are **linearly** related. It is always between -1 and 1:

$$\rho_{XY} = \frac{S_{XY}}{S_X S_Y}$$

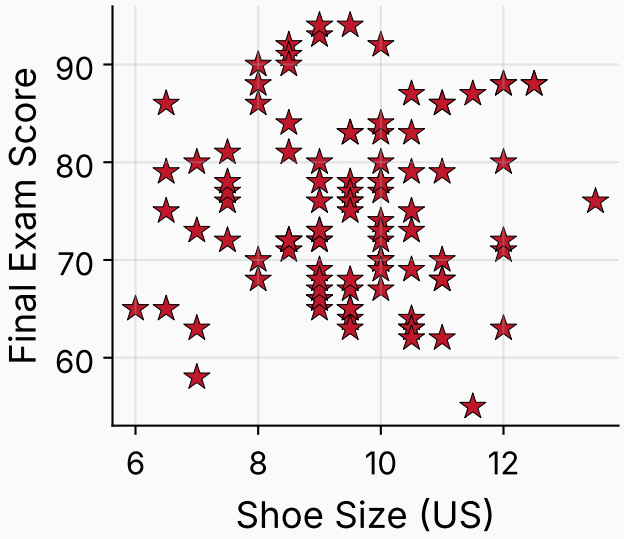
Positive Relationship
ρ = 0.97



Negative Relationship
ρ = -0.94



No Relationship
ρ = 0.03



Covariance & Correlation

PS1 Question 2.2.3

The following summary statistics are provided for data on sales (y) and advertising (x) for a company:

Statistic	Value
$\bar{x}$	4
$\bar{y}$	2
sample variance of x	1
sample variance of y	0.25
covariance of x and y	0.4

(a) Does it appear that it could be true that advertising improves sales? Be clear on why or why not using the numbers above.

Covariance & Correlation

PS1 Question 2.2.3

Statistic	Value
$\bar{x}$	4
$\bar{y}$	2
sample variance of x	1
sample variance of y	0.25
covariance of x and y	0.4

(b) Can you say anything about the strength of the relationship between advertising and sales? Be clear on what numbers you use to support your argument.

$$\rho_{XY} = \frac{S_{XY}}{S_X S_Y} =$$

PS1 Question 2.2.8

Consider the following three observations on the level of bacteria at beaches and the number of ear infections reported to lifeguards. Note that these numbers are relative to the mean, i.e. the mean has already been subtracted out.

Level of Bacteria (minus the mean)	**# Ear Infections (minus the mean)**
-0.025	2
0.05	3
-0.025	-5

(a) What is the covariance between the two variables?

$S_{XY} =$

PS1 Question 2.2.8

Level of Bacteria (minus the mean)	# Ear Infections (minus the mean)
-0.025	2
0.05	3
-0.025	-5

(b) What does this sample tell you about the relationship between the level of bacteria and ear infections (be careful and complete)?