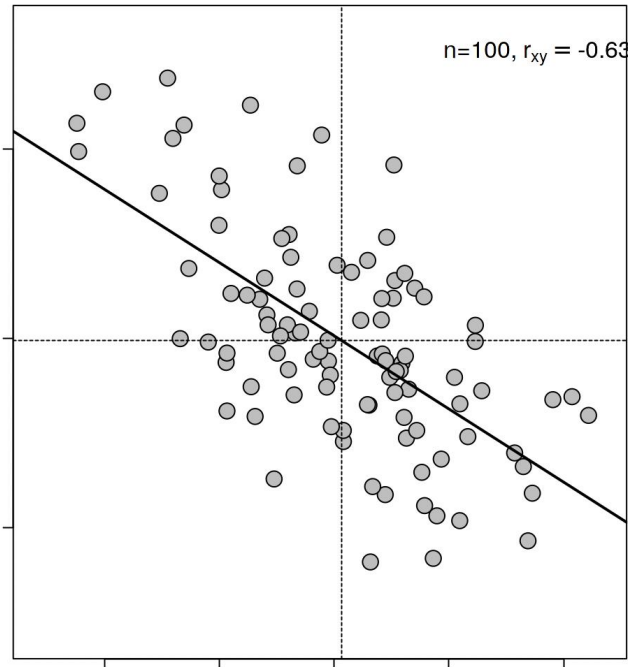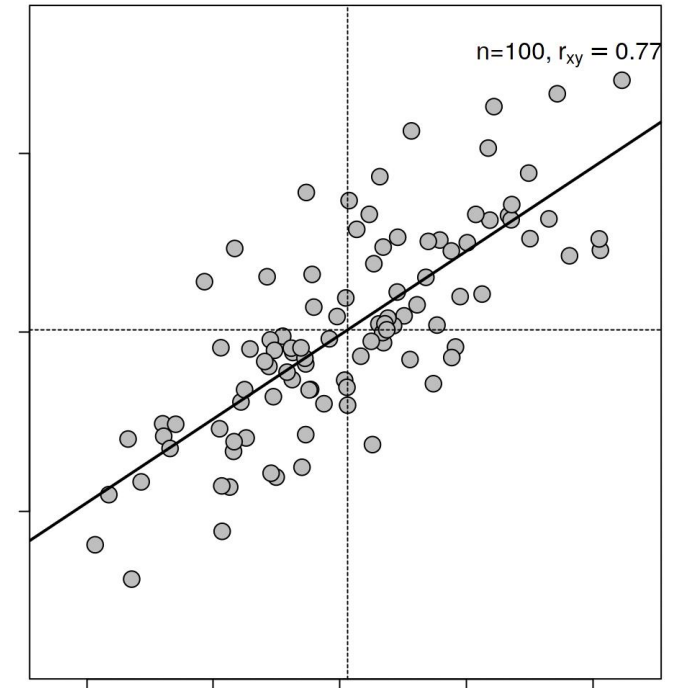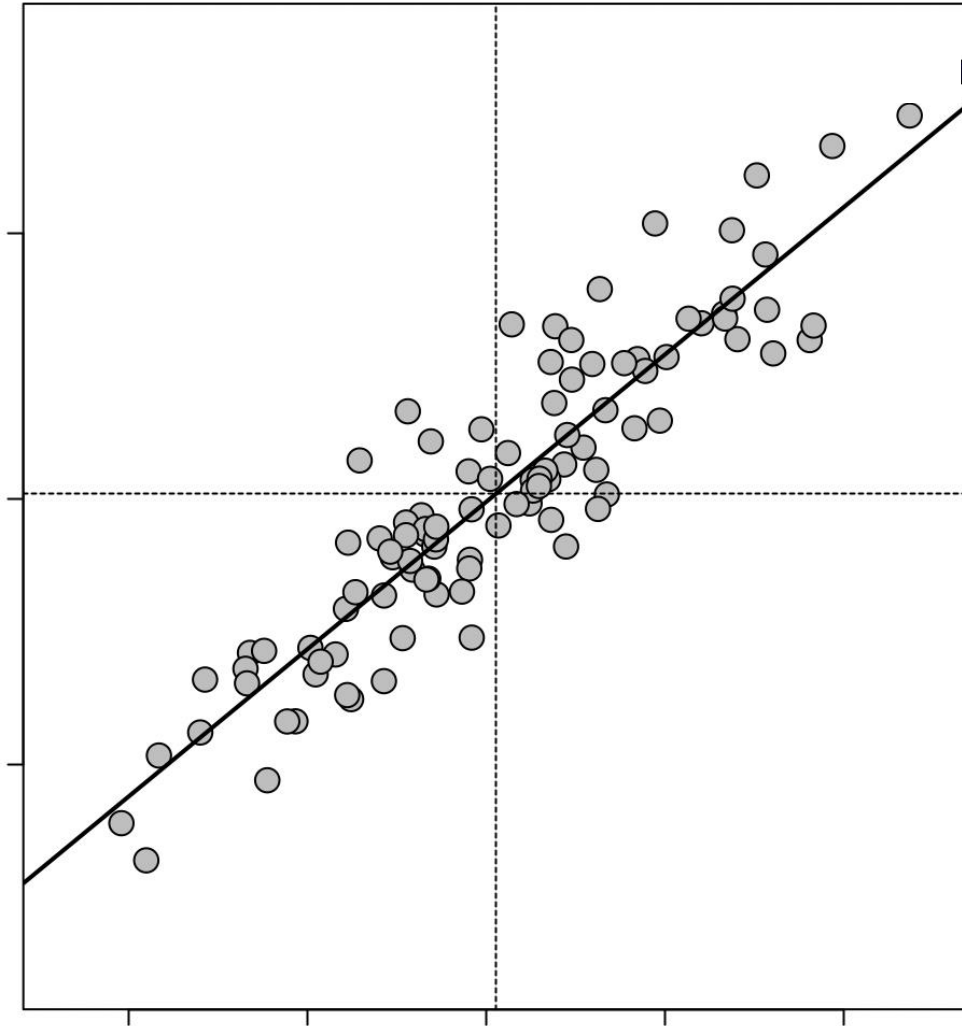# Fundamentals of Econometrics Models

Vicenç Soler

v.soler@tbs-education.org

~~vincent.soler@tbs-education.org~~

Quick reminder

Is there a correlation?

If so, positive or negative?

Weak or strong?

Can you guess the correlation coefficient (between -1 and 1)?

R = 0.91

Quick reminder

What is *** telling us?

	Dependent variable:			
	Resist	Resistance		Resistance
	(1)	(2)	(3)	(4)
Cement	0.988***			0.846***
	(0.027)			(0.056)
Additives		64.266***		
		(2.682)		
Water			0.196***	
			(0.067)	
Additives				15.160***
				(5.284)
Water				-0.073
				(0.065)
Constant	0.738	97.890***	211.884***	15.925
	(7.150)	(6.694)	(14.240)	(14.846)
Observations	804	804	804	804
R2	0.617	0.417	0.011	0.623
Adjusted R2	0.617	0.417	0.009	0.621
Residual Std. Error	37.778 (df = 802)	46.616 (df = 802)	60.740 (df = 802)	37.563 (df = 800)
F Statistic	1,293.782*** (df = 1; 802)	574.388*** (df = 1; 802)	8.700*** (df = 1; 802)	439.929*** (df = 3; 800)

Note: $*p<0.1$; $**p<0.05$; $***p<0.01$

Quick reminder

```
Call:
lm(formula = Mshare ~ TropPremium + Trop + MMaid + Aldi, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-8.9771 -2.9296 -0.6031  2.3847 16.6914

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.4101     8.2955   1.375    0.172
TropPremium    8.3962     6.9659   1.205    0.231
Trop          -4.1488     6.9982  -0.593    0.554
MMaid         -8.5738     0.7558 -11.343  < 2e-16 ***
Aldi           4.1263     0.7390   5.584 1.57e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.519 on 116 degrees of freedom
Multiple R-squared:  0.5627,    Adjusted R-squared:  0.5476
F-statistic: 37.32 on 4 and 116 DF,  p-value: < 2.2e-16
```

SESSION 4
-DUMMY VARIABLES-

Dummy Variables

Quite often, the data that we analyse include **categorical variables**, which split the sample into groups.

Can you think about examples of categorical variables?

Gender, marital status, citizenship, industrial sector, etc.

To include the groups in the regression equation, we use **dummy variables**.

Dummy Variables

A dummy variable is a variable taking values **0** or **1**

Dummy Variables

Let us start with the simplest case, in which we have only two groups, for instance, female and male managers.

We code these two groups with a dummy D, which takes values as follows: **D = 1 in the male group**, and **D = 0 in the female group**.

We wish to interpret the coefficients of an equation

$$Y = a + b\,D$$

Dummy Variables

Let's imagine that Y is "salary" and D is "gender"

$$Y = a + b\,D$$

Since D can only take two values, the equation can be easily read by looking at the two cases separately:

- In the group $D = 0$ (female), the predicted value of Y is a.

- In the group $D = 1$ (male), the predicted value of Y is $a + b$.

TIME TO PRACTICE!

Gender Salary Gap

Norwegian Airlines is one of the biggest Norwegian corporations with an annual turnover of 3.800 million euros and nearly 10,000 employees.

It also has a long history of **respect for diversity**. They have an internal rule against gender, race, and religious discrimination.

Since then, it has been a model for managing diversity in organizations.

Gender Salary Gap

However, in the last years, Norwegian's CEO has observed that **the number of women in senior management positions was very low**, and decided to make an effort to, whenever possible, increase the proportion of women as board members, regional managers, country managers, and heads of department.

Gender Salary Gap

Preparing a meeting in Barcelona, where he will meet with European country managers of the company, he reviews the data on salaries for the Spanish managers. It seems that **Norwegian has done a good job in promoting women in the country**: a relevant proportion of senior managers are woman. But something is weird:

The average salary of female managers is much lower than that of men!

Gender Salary Gap

At the Barcelona headquarters, he meets Spain's country manager, Alfons. Discussing what he has found, the country manager suggests that the salary gap is not due to gender.

According to him, the source of the difference is that, in general, the female managers have been less time with the company, so they had fewer opportunities to get salary increases.

Gender Salary Gap

The data for the analysis (file gender.csv) covers 288 managers.

The salaries are in euros, and there is also data on the tenure, in years.

First of all, check:

Is the gender distribution balanced?
How many female and male managers do they have?

The gender distribution is, indeed, balanced,
with 140 female and 148 male managers.

Gender Salary Gap

However, the CEO wonders...

Second question: How wide is the salary average gap?

Gender Salary Gap

How is the average salary for men?

For women?

What is the gap?

Gender Salary Gap

How wide is the salary gap?

The average annual salary is **$74,420 for the female managers** and **$91,552 for the male managers**.

Indeed, this looks like a clear case of gender discrimination.

Gender Salary Gap

However, Alfons told the CEO that the gap can be explained by the number of years with the company. So the CEO asks you to check it with what you already know about regression analysis.

Calculate the correlation between salary and tenure, and visualize the data with a scatter plot

Gender Salary Gap

R = 0,3

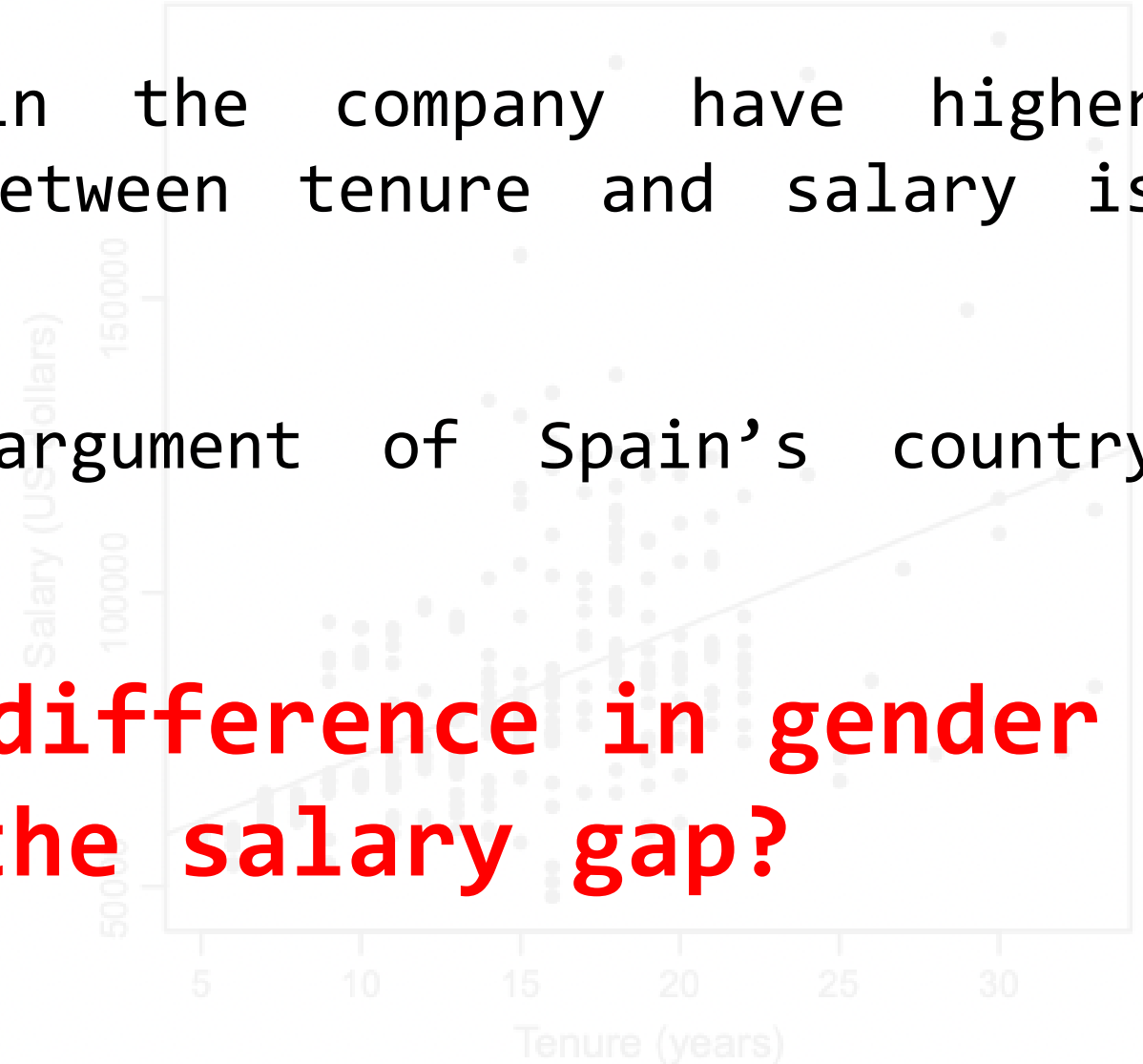**How do you interpret these results?**
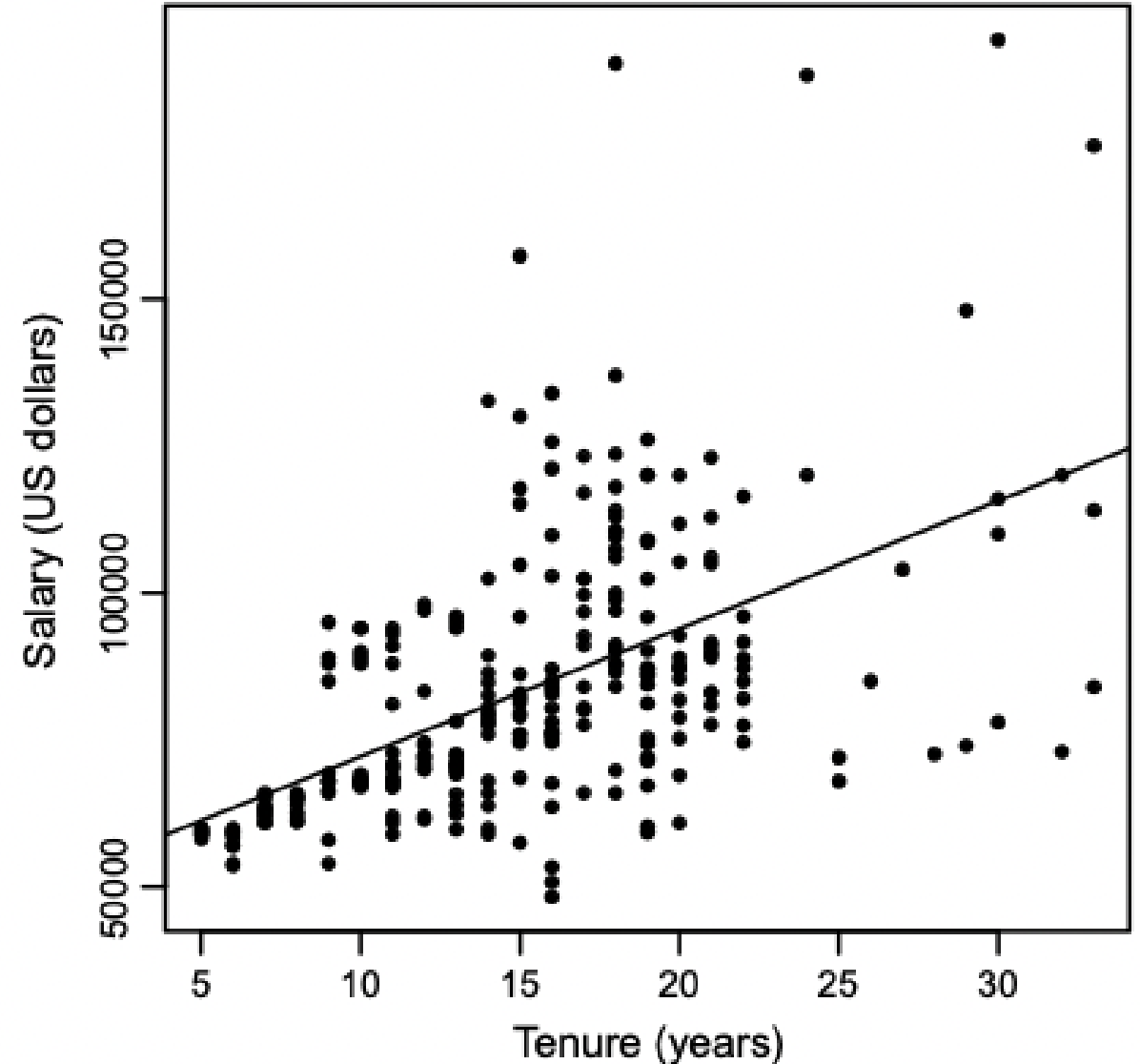
Gender Salary Gap

Managers with more time in the company have higher salaries. The correlation between tenure and salary is positive and weak, $R = 0.3$

This hardly supports the argument of Spain's country manager. So…

May it be that the difference in gender accounts for the salary gap?

Gender Salary Gap

Let's check it with regression analysis!

And R result

Gender Salary Gap

Now, check the regression:

Salary = ? + ? Tenure + ? Male

R = ?

Gender Salary Gap

Now, check the regression:

Salary = 50,725 + 1,884 Tenure + 7,906 Male

R = 0.320

How do we interpret these results?

Gender Salary Gap

Salary = 50,725 + 1,884 Tenure + 7,906 Male

How do we interpret these results?

Is there a gender gap?

It has to be interpreted as the **change in salary due to a change in the variable MALE, holding the other variables (TENURE) constant.**

So, it tells us that, **for the same number of years in the company, the female managers get, on average, $7,906 less.**

So, even if TENURE explains, in part, the salary gap, there is still one part of that which could be attributed to gender.

Gender Salary Gap

And statistically, we improve the R result

Gender Salary Gap

Salary = 50,725 + 1,884 Tenure + 7,906 Male

How do we interpret these results?

Is there a gender gap?

QUESTIONS?

It has to be interpreted as the change in salary due to a change in the variable MALE, holding the other variables (TENURE) constant.

Remember:

There are no dumb questions, nobody is born knowing!

So, it tells us that, for the same number of years in the company, the female managers get, on average, $7,907 less.

So, even if TENURE explains, in part, the salary gap, there is still one part of that which could be attributed to gender.
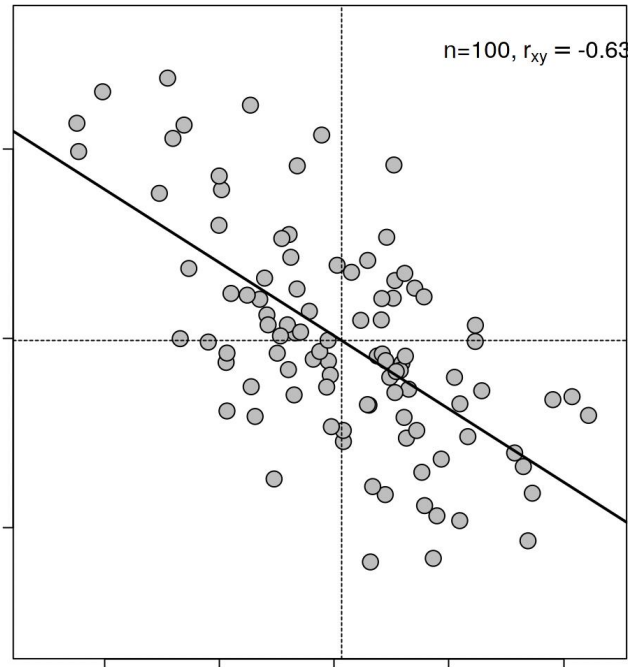
Regarding next class

Is there any of the concepts explained up to now not clear enough?

Fundamentals of Econometrics Models

Vicenç Soler

v.soler@tbs-education.org

~~vincent.soler@tbs-education.org~~

n=100, r_{xy} = -0.63

n=100, r_{xy} = 0.77