

## Module 2 – Regression and Prediction

# CASE STUDY ACTIVITY TUTORIAL

### Case Study 2: Gender Wage Gap



# Regression 1.6: Case Study: Gender Wage Gap

V. Chernozhukov  
MIT

June 12, 2016

## Regression 1.6: Case Study: Gender Wage Gap.

Here we consider the case study of the gender wage gap in the U.S.

# The Inference Problem

What is the difference in predicted wages between men and women with the same job-relevant characteristics?

## └ The Inference Problem

What is the difference in predicted wages between men and women with the same job-relevant characteristics?

Here we ask and answer the following question:  
What is the difference in predicted wages between men and women with the same job-relevant characteristics?

## CPS 2012 Data: Summary

	Male means	Female means
Wage	16.12	14.72
Experience	13.58	13.04
College graduate	0.35	0.41
Some college	0.30	0.35
High school graduate	0.34	0.24
Midwest	0.28	0.29
South	0.24	0.26
West	0.22	0.20
Northeast	0.26	0.26

## └ CPS 2012 Data: Summary

	Male means	Female means
Wage Experience	16.12	14.72
College graduate	13.58	13.04
Some college	0.35	0.41
High school graduate	0.30	0.35
Midwest	0.34	0.24
South	0.28	0.29
West	0.24	0.26
Northeast	0.22	0.20

Here we work with the U.S. Current Population Survey Data, for the year 2012.

We first take a look at the following descriptive statistics for the subsamples of single men and single women, with educational attainment equal to high-school, some college, or college.

The mean hourly wage is 16\$ for men and about 15\$ for women, so the difference is 1\$, without controlling for job-relevant characteristics.

If we take a look at some of these characteristics, we see that on average men have more experience, but women are more likely to have college degrees or some college education. Geographical distribution of men and women is similar.

# Specifications

- › We estimate the linear regression model:

$$Y = \beta_1 D + \beta_2' W + \varepsilon. \quad (1)$$

- ›  $D$  is the indicator of being a female (1 if female and 0 otherwise).  $W$ 's are controls.
- › Basic model:  $W$ 's consist of education and regional indicators, experience, experience squared, and experience cubed.
- › Flexible model:  $W$ 's consist of controls in the basic model plus all of their two-way interactions.



## Specifications

- We estimate the linear regression model:

$$Y = \beta_1 D + \beta_2 W + \varepsilon \quad (1)$$

- $D$  is the indicator of being a female (1 if female and 0 otherwise).  $W$ 's are controls.
- Basic model:  $W$ 's consist of education and regional indicators, experience, experience squared, and experience cubed.
- Flexible model:  $W$ 's consist of controls in the basic model plus all of their two-way interactions.

We next proceed to answer our question.

We estimate the linear regression model that we have introduced.

Our outcome  $Y$  is hourly wage, and our target regressor  $D$  is the indicator of being a female, and controls  $W$ 's are the job-relevant characteristics. We consider two specifications for the controls:

In the Basic model, our  $W$ 's consist of education and regional indicators, experience, experience squared, and experience cubed.

In the Flexible model, our  $W$ 's consist of controls in the basic model and all of their two-way interactions.

# Results

	Estimate	Std Error	Confidence Interval
	$\hat{\beta}_1$	$\sqrt{\hat{V} / n}$	95%
Basic reg	-1.83	0.42	[-2.66 -0.99]
Flex reg	-1.88	0.42	[-2.71 -1.05]

- › The estimated gender gap in hourly wage is about -2\$ with a confidence interval that ranges from about -2.7\$ to -1\$.
- › This means that women get paid 2\$ less per hour on average than men, controlling for experience, education, and geographical region.

## Results

	Estimate	Std Error	Confidence Interval
	$\beta_1$	$\sqrt{V / n}$	95%
Basic reg	-1.83	0.42	[-2.66 -0.99]
Flex reg	-1.88	0.42	[-2.71 -1.05]

- The estimated gender gap in hourly wage is about -2\$ with a confidence interval that ranges from about -2.7\$ to -1\$.
- This means that women get paid 2\$ less per hour on average than men, controlling for experience, education, and geographical region.

We now go to the empirical results.

In the following table we see the estimated regression coefficient, its standard error, and the 95% confidence interval, for both the basic and flexible regression model.

The results for basic and flexible regression models are in a very close agreement:

The estimated gender gap in hourly wage is about -2 dollars. This means that women get paid 2 dollars less per hour on average than men, controlling for experience, education, and geographical region.

The 95% confidence interval ranges from -2.7 to -1 dollars.

We conclude that the difference in the hourly wage for men and women, who have the same recorded characteristics, is both statistically and economically significant.

# Conclusions

- › We applied the previous results to learn about the gender wage gap.
- › The gender wage gap may partly reflect **discrimination** against women in the labor market.
- › The gender wage gap may partly reflect a **selection effect**, namely that women are relatively more likely to take on occupations that pay somewhat less (for example, school teaching).

## └ Conclusions

- We applied the previous results to learn about the gender wage gap.
- The gender wage gap may partly reflect **discrimination** against women in the labor market.
- The gender wage gap may partly reflect a **selection effect**, namely that women are relatively more likely to take on occupations that pay somewhat less (for example, school teaching).

To sum it up... we applied the ideas we discussed so far to learn about the gender wage gap.

The gender wage gap may partly reflect genuine discrimination against women in the labor market....

It may also partly reflect the so-called selection effect, namely that women are more likely to end up in occupations that pay somewhat less (for example, school teachers).

This segment concludes our discussion of the classical linear regression.

Next we will briefly discuss other types of regression.

## Module 2 – Regression and Prediction

# THANK YOU

Case Study 2: Gender Wage Gap



PROFESSIONAL EDUCATION

Digital Programs

