DA2 Assignment 1

Alima Dzhanybaeva, Hanna Asipovich

2022-11-24

Introduction

For our assignment we chose to compare the earnings of employees for the category **occ2012=2100** (Lawyers, Judges, magistrates, and other judicial workers). After applying additional filters on age, weekly hours (> 20 hours), and educational levels (grade92 >= 44) we obtained 965 observations. Out of those, 370 are female and 595 are male. Additionally, we created binary variable *female* (that is 1 if the individual is a woman and 0 if a man) and generated an hourly wage variable.

Simple Linear Regression Analysis

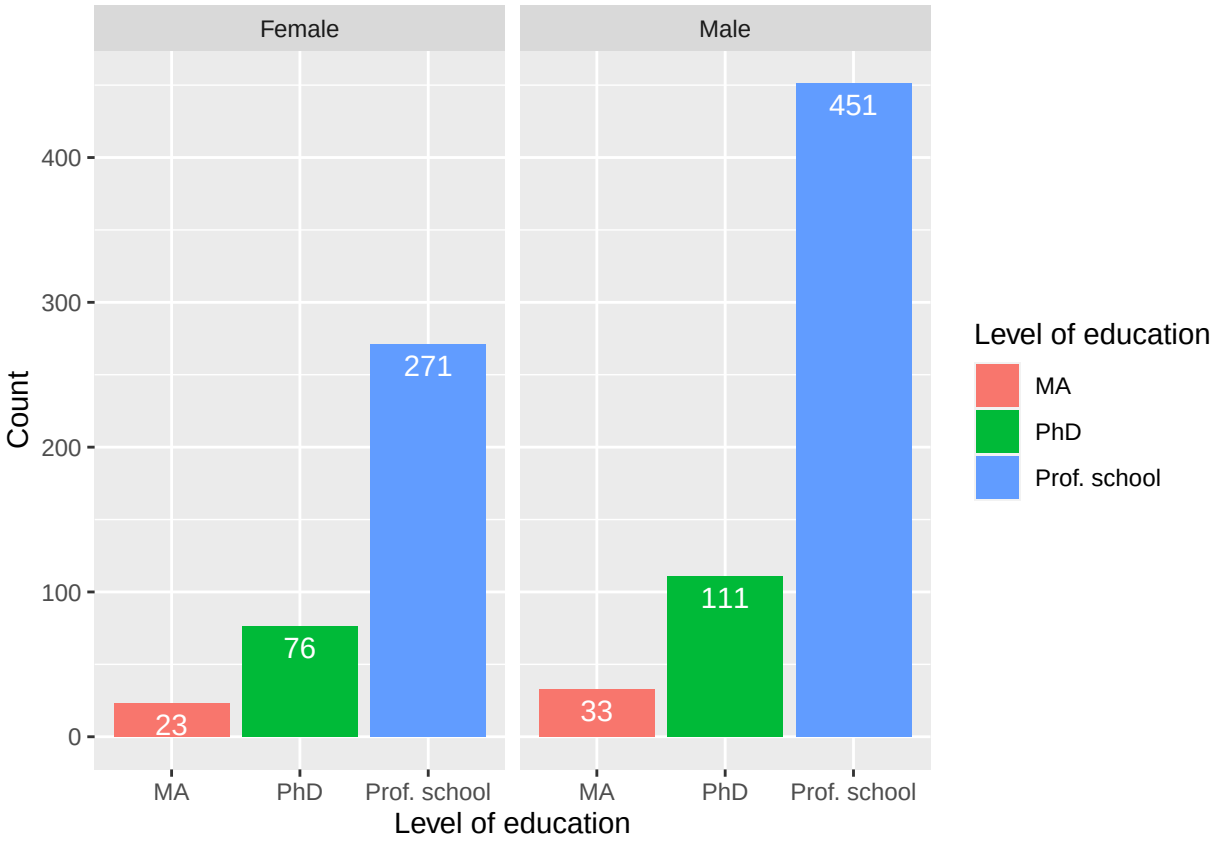
For our simple regression analysis, we applied log transformation on hourly wages and used it as the dependent variable. As an explanatory variable, we used *female* variable. For **Model 1** we used simple linear regression and for **Model 2** we applied linear regression with robust standard errors. However, there is no practical difference between these two models.

	Model 1	Model 2
(Intercept)	3.6777**	3.6777**
	(0.0187)	(0.0183)
female	−0.0918**	−0.0918**
	(0.0302)	(0.0306)
Num.Obs.	965	965
R2	0.010	0.010
RMSE	0.46	0.46

* p < 0.05, ** p < 0.01

Based on the obtained results, the explanatory variable *female* is statistically significant even at 0.01 significance level, and women employed as Lawyers, Judges, magistrates, and other judicial workers on average earn 9.2% less than males with the same occupation. We can be 95% confident that the average difference between the hourly earnings of female employees in the chosen category versus male ones was from -15% to -3%.

Multiple regression analysis



We took 3 levels of education background: MA, PhD, and Professional school. **Model 1** shows the results for simple regression with robust standard errors where we had loged hourly wage as the dependent variable and *female* as the independent variable. For **Model 2** we ran robust regression on loged hourly wages with the base category "Professional school", as it has the most observations; the category "Phd" was used as the base category in **Model 3**.

	Model 1	Model 2	Model 3
(Intercept)	3.6777**	3.6942**	3.6462**
	(0.0183)	(0.0201)	(0.0347)
female	−0.0918**	−0.0900**	−0.0900**
	(0.0306)	(0.0305)	(0.0305)
ed_MA		−0.1349*	−0.0869
		(0.0609)	(0.0672)
ed_PhD		−0.0480	
		(0.0372)	
ed_Profess			0.0480
			(0.0372)
Num.Obs.	965	965	965
R2	0.010	0.015	0.015
RMSE	0.46	0.45	0.45

* $p < 0.05$, ** $p < 0.01$

As we can see, gender has proven to always be statistically significant. However, when we introduced additional variables for education level, the coefficient for *female* variable became lower in absolute value. Therefore, we can conclude that grade92 also explains the difference in earnings.

Gender

Both **Model 2** and **Model 3** showed that women of the chosen category, on average, earn 9% less than men working in the same occupation.

Educational level

ed_MA is statistically different from zero at 0.05 significance level in **Model 2** and insignificant in **Model 3**. *ed_Phd* and *ed_Profess* are both statistically insignificant in **Model 2** and **Model 3** respectively. This may come from the prevalence of people with 'Professional school' degree in the chosen occupation. Despite this fact, we still interpret obtained coefficients for these variables.

Model 2: On average, in comparison with workers whose highest educational level is 'Professional school' people with a Master's degree earn 13.5% less, and individuals with Phd earn 4.8% less.

Model 3: In comparison with workers with PhD people with a Master's degree, on average, earn 8.7%% less and individuals whose highest educational level is 'Professional school' earn 4.8% more.

Additionally, it can be seen from the regressions, that the models have a weak fit, so they are not performing particularly well when explaining the variation in y. This is due to the fact that we only used two variables.

Summary

We analyzed the pattern of relationship between hourly earnings, gender, and level of education for lawyers, judges and other legal professionals. Our guess was that females employed in this category earn on average less than men. This proved to be true. Further, we ran linear regression models with a few transformations on the level of education.