Documentation

The data

We are using a subset of the CPS-earnings dataset filtering for Human Resources professionals coded as Human resources managers (0136), Human resources workers (0630) and Human resources assistants, except payroll and timekeeping (5360). This subset includes 1131 observations from which 861 are female and 270 are male.

The weekly wages show a distribution with a long right tail. The weekly earnings are first converted into hourly earnings using the weekly working hours column to make the figures comparable. After this, the ln of hourly earnings is taken so that we may interpret our findings as a relative change in hourly earnings.

The ln of hourly wages shows a distribution which is close to normal as seen in Figure 1.

The regression

We are regressing the ln of the hourly wage on gender. Based on this, female HR employees earn 15.7% less, on average, than their male counterparts in this data. The R squared is 0.018 meaning that gender explains less than 2% of the variation in earnings in this data.

The coefficient interval of the slope is [-0.232;-0.083][1] which does not include zero. This means that we can be 95% confident that the average earning difference between male and female HR employees is between -23.2% and -0.83%. Therefore, we can rule out with 95% confidence that their average earnings (female and male) are the same in the population (p-value <0.05)[2].

Education

We include education levels into the regression by first filtering out everything below high school level due the low number of observations (10 in total). We then form 4 groups: (1) High school, (2) college (everything above high school but below masters), (3) MA & professional degrees, and (4) PhD. We create two regressions: one where the comparison group is high school (the lowest education level in the data), and another where it is the MA & professional degrees group. For comparison, we also show the regression only including gender too. The coefficients are summarized in Table 1.

In column 2, Female HR employees earn 14.0% less, on average, than their male counterparts with identical levels of education. College, MA or professional and PhD degree holders of the same gender are expected to earn 13.7%, 31.6% and 67.2% more respectively than those with only a high school degree. All covariates are significant at 99%. The model's R squared (0.054) is higher than that of column 1 (0.017) suggesting that including the factor of education increased our model's goodness of fit. Furthermore, in this data women appear to be more likely to be in lower-earning categories than in higher-earner as the female coefficient for column 2 is lower in absolute value than for column 1 (-0.14 vs -0.154).

When using the MA or professional degree group as comparison, high school and college degree holders of the same gender are expected to earn 31.6% and 17.9% less, while PhD holders are expected to earn 35.6% more than those with an MA or professional degree. All covariates are significant at 99% except for PhD which is only significant at 95%.

The confidence intervals of different levels of education do not contain 0, which means that we can rule out with a 95% confidence that their average earnings (females with different level of education) are the same in the population.

[1] Upper bound: -0.1571+1.96*0.038=-0.083;
Lower bound: -0.1571-1.96*0.038=-0.232

[2] P value is also < 0.01 therefore the statement also holds true with 99% confidence.

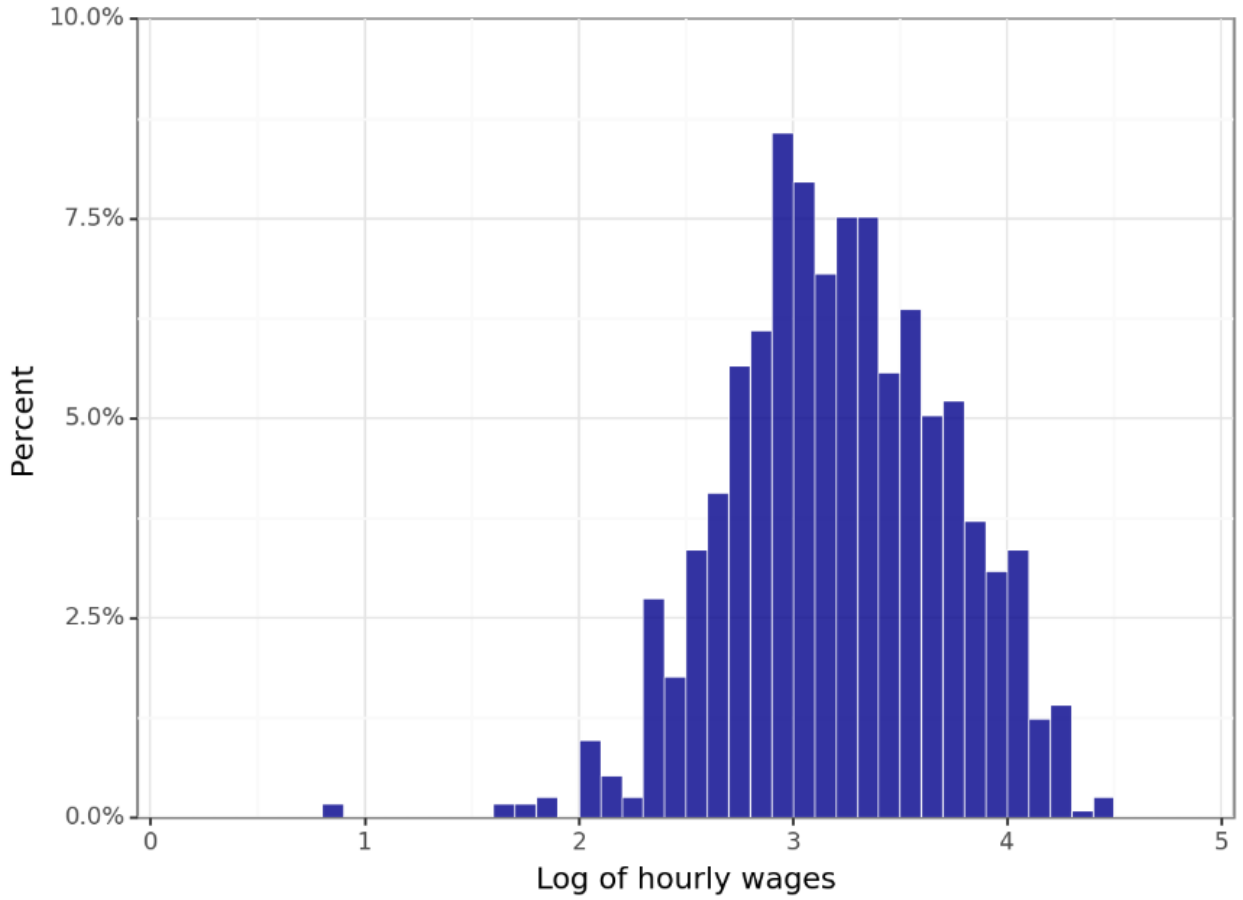

Figure 1: Histogram of (natural log of hourly wages)

	Dependent variable: lnw		
	(1)	(2)	(3)
female	-0.154***	-0.140***	-0.140***
	(0.038)	(0.038)	(0.038)
ed_HS			-0.316***
			(0.055)
ed_college		0.137***	-0.179***
		(0.044)	(0.042)
ed_MA_Prof		0.316***	
		(0.055)	
ed_Phd		0.672***	0.356**
		(0.145)	(0.145)
Constant	3.336***	3.170***	3.486***
	(0.035)	(0.050)	(0.049)
Observations	1121	1121	1121
R^2	0.017	0.054	0.054
Adjusted R^2	0.016	0.051	0.051
Residual Std. Error	0.500 (df=1119)	0.492 (df=1116)	0.492 (df=1116)
F Statistic	16.112*** (df=1; 1119)	15.053*** (df=4; 1116)	15.053*** (df=4; 1116)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1: Regression comparison table