

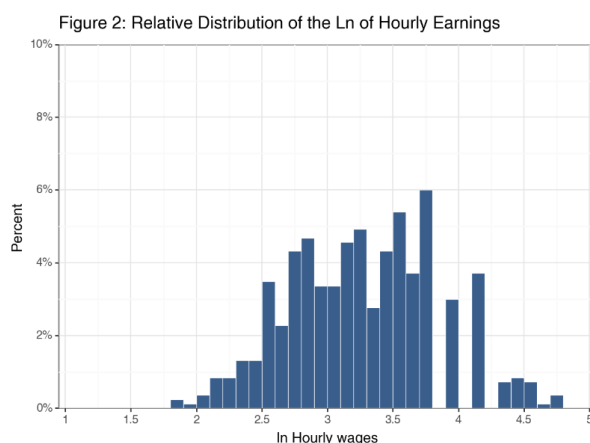
Data Analysis 2 - Assignment 1

Github: https://github.com/mateschieszler/DA2_2023

Occupation: Physicians and surgeons (2010 CENSUS CODE: 3060)

Data on 343 female and 439 male physicians and surgeons. Taking the natural logarithm of earnings is crucial for us as we wish to examine the percentage difference between male and female income.

Distribution of hourly earnings and Ln hourly earnings:



Task 1 - Unconditional Gender Gap

Model 1:

Gender – Level wage

In our data female physicians earn 0,5 dollars less an hour on average compared to their male colleagues. The average hourly wage of male physicians and surgeons is 40.9 dollars an hour. The CI shows that the true value of the difference in earnings is between 3.5 and -4.5. The low R-squared suggests that a different model may bring more definitive results.

OLS Regression Results										
Dep. Variable:	w	R-squared:	0.000	coef	std err	z	P> z 	[0.025	0.975]	
Model:	OLS	Adj. R-squared:	-0.001	Intercept	40.8531	0.926	44.098	0.000	39.037	42.669
Method:	Least Squares	F-statistic:	0.06190	female	-0.5016	2.016	-0.249	0.804	-4.453	3.450
Date:	Sat, 25 Nov 2023	Prob (F-statistic):	0.804	Omnibus:	897.004	Durbin-Watson:			1.842	
Time:	11:07:26	Log-Likelihood:	-3938.5	Prob(Omnibus):	0.000	Jarque-Bera (JB):			129206.131	
No. Observations:	839	AIC:	7881.	Skew:	4.706	Prob(JB):			0.00	
Df Residuals:	837	BIC:	7890.	Kurtosis:	63.062	Cond. No.			2.46	
Df Model:	1									
Covariance Type:	HC1									

Table 1 – Model 1 Regression results

Model 2:

Gender – Ln wage

Female physicians and surgeons earn on average 8% less than their male colleagues in our data. The R-squared is higher than in model one (0.002), meaning the variance in the ln of wages is better explained by this model. The CI suggests that the true value of this wage difference lies between 20% less or 4% more wage for women in this occupation. With a p-value of 0.192 we lack statistically significant evidence to claim that the wage for female and male physicians differs at 0.05 significance level.

OLS Regression Results										
Dep. Variable:	lnw	R-squared:	0.002	coef	std err	z	P> z 	[0.025	0.975]	
Model:	OLS	Adj. R-squared:	0.001	Intercept	3.5277	0.038	93.576	0.000	3.454	3.602
Method:	Least Squares	F-statistic:	1.652	female	-0.0801	0.062	-1.285	0.199	-0.202	0.042
Date:	Sat, 25 Nov 2023	Prob (F-statistic):	0.199	Omnibus:	1082.755	Durbin-Watson:			1.866	
Time:	11:07:26	Log-Likelihood:	-1075.7	Prob(Omnibus):	0.000	Jarque-Bera (JB):	245489.765			
No. Observations:	839	AIC:	2155.	Skew:	-6.487	Prob(JB):	0.00			
Df Residuals:	837	BIC:	2165.	Kurtosis:	85.789	Cond. No.	2.46			
Df Model:	1									
Covariance Type:	HC1									

Table 2 – Model 2 Regression results

Task 2 - Variance on Level of Education

Model 3:

Dependent variable: lnw			
	(1)	(2)	(3)
female	-0.080 (0.062)	-0.058 (0.062)	-0.076 (0.062)
ed_Pro		0.294*** (0.075)	0.085 (0.064)
ed_PhD		0.264*** (0.080)	
ed_MA			-0.123 (0.100)
Constant	3.528*** (0.038)	3.273*** (0.069)	3.489*** (0.060)
Observations	839	839	839
R ²	0.002	0.014	0.007
Adjusted R ²	0.001	0.010	0.003
Residual Std. Error	0.873 (df=837)	0.869 (df=835)	0.872 (df=835)
F Statistic	1.652 (df=1; 837)	7.006*** (df=3; 835)	3.044** (df=3; 835)

Table 3 - Model 3 Regression results

The Baseline model is identical to Model 2 from Task 1. Controlling for professional education and Ph.D. degrees narrows the wage difference to 5.8% less (7.6% for Master's) on average for women. However, this coefficient remains statistically insignificant for us. Notably, this model tells us about individuals with professional education earn on average 29.4% more than those with a Master's degree. Similarly, individuals with Ph.D. degrees see a 26.4% increase in earnings, both coefficients are statistically significant. Introducing Master's education shows that individuals with a Master's degree in our data earn on average 12.3% less than those with a Ph.D. Controlling for education also improves the model's explanatory power as we see an increased R-squared of 0.014.

Model 4:

This model looks at age and the ln of wages. It shows for each year of age, females can expect to get 1,8% higher salary at the 99% confidence level. However, for males, at the same confidence level this value is 2.4%. The interaction coefficient also shows that there is a 0.6% differential effect of age on the natural logarithm of wages for females compared to males. However, since the interaction coefficient is not statistically significant, we cannot do inference about gender differences.

Dependent variable: lnw			
	Women	Men	All
	(1)	(2)	(3)
female			0.197
			(0.215)
age	0.018***	0.024***	0.024***
	(0.003)	(0.003)	(0.003)
female x age			-0.006
			(0.005)
Constant	2.688***	2.491***	2.491***
	(0.155)	(0.149)	(0.149)
Observations	343	496	839
R ²	0.041	0.098	0.074
Adjusted R ²	0.038	0.096	0.070
Residual Std. Error	0.902 (df=341)	0.798 (df=494)	0.842 (df=835)
F Statistic	27.083*** (df=1; 341)	54.119*** (df=1; 494)	28.033*** (df=3; 835)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5 - Model 4 Regression results

Dependent variable: lnw			
	(1)	(2)	(3)
female	-0.048	-0.028	-0.044
	(0.061)	(0.060)	(0.061)
age	0.021***	0.021***	0.021***
	(0.002)	(0.002)	(0.002)
ed_Pro		0.280***	0.083
		(0.069)	(0.061)
ed_PhD		0.249***	
		(0.074)	
ed_MA			-0.120
			(0.092)
Constant	2.585***	2.351***	2.549***
	(0.114)	(0.118)	(0.133)
Observations	839	839	839
R ²	0.073	0.083	0.077
Adjusted R ²	0.070	0.079	0.073
Residual Std. Error	0.842 (df=836)	0.838 (df=834)	0.841 (df=834)
F Statistic	42.311*** (df=2; 836)	30.767*** (df=4; 834)	23.513*** (df=4; 834)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4 - Model 5 Regression results

Model 5:

This model combines Model 3 and Model 4. This now better describes the variance in the dependent variable as this model demonstrates increased values for R-squared. In our data females with a Professional or Ph.D. degree earn 2.8% less than males compared to those who only have Master's degree. This coefficient is 4.4% when looking at female Master's degrees compared to those who have a Ph.D.

Summary:

In conclusion, our analysis reveals that among Physicians and Surgeons in our dataset, the wage gap diminishes progressively with higher levels of education. However, this conclusion is specific to our data, as our coefficients are not statistically significant at the 0.05 significance level. Nonetheless, we can infer that for each year of age males