Does hourly wage influence sleep?

Econometetrics Project - Group 1

Austin Halvorsen, Chase Harward, Suneet Dash

12/8/2020

# Introduction

How does someone’s hourly wage impact the amount of sleep that they get on a given night?

* Determining what role hourly wage plays in sleep is important because it can help us to understand an aspect of sleep that isn’t explored as often as opposed to just looking at hours of work itself.
* Through this question we can even see how different income classes sleep, such as how many hours does someone in the top income percentile sleep as opposed to someone in the middle class or someone making minimum wage.

## Initial Assumptions

* On one side of the question we assumed that people making minimum wages would typically sleep more, as a lot of these people are just entering the labor market and would be younger (high school/college). Also they wouldn’t have to deal with as many high-stress situations as someone who is making a significantly higher amount of money.
* On the other side we considered that someone making minimum wages could be sleeping less, as they might be working a second job or something like that to make their quality of life better. And that someone making a higher wage would be in a comfortable position in life where financial issues weren’t impacting their sleep.

## Ethical Importance

This question poses some interesting ethical discussions.

* We typically look towards financial positions, family status, and where people live as basic indications of a quality of life. Something we don’t often consider is what type of quality of life information we can get from sleep data.
* Another example would be in terms of government programs allocating resources towards low-income workers, can we make a conclusion on the quality of life they have based on the relationship between the hourly wage and sleep?

# Data

* We will be using the dataset “sleep75” found in the Wooldridge package and originally used for the paper “Sleep and the Allocation of Time” by Jeff Biddle and Daniel Hamermesh, published in the Journal of Political Economy in 1998.
* The goal of the original paper was to show a relationship between an increase in time in the labor market and how much sleep someone gets.
* Previous questions involving the “sleep75” data set typically included some analysis on a tradeoff between sleep and time worked, so we aimed to ask a question that hasn’t really been explored with this data.
* This data set includes 706 observations on 34 variables, notably hourly wage, sleep, years of experience, marital status, spouse wage, age, children or no children, type of employment, and quality of health.

## Variables

Our explanatory variables in our model are:

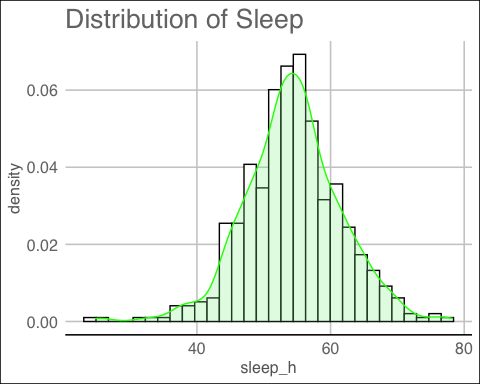
* age = years old of the individual
* exper = combination of age and education
* hrwage = the hourly rate of the individual
* marr = Is the person married or not?
* yrsmarr = time in years the person was married (if married)
* spsepay = amount of pay for spouse if they are paid
* male = is the person male or not
* gdhlth = is the individual in good health or not
* totwrk = amount of time spent working per week

## Variables (con’t)

Our Dependent variable in our study is sleep.

* As part of our manipulation for our data, we changed sleep from “minutes” to “hours” to have a simpler interpretation of our regression.
* Additionally since we were only analyzing individuals who are working, we filtered our data to only include those who are in the labor force and have an hourly wage.

## Distribution



## Summary Stats

sleep\_h age exper hrwage   
 Min. :24.75 Min. :23.00 Min. : 0.00 Min. : 0.350   
 1st Qu.:50.06 1st Qu.:29.00 1st Qu.:10.00 1st Qu.: 2.890   
 Median :54.19 Median :36.00 Median :17.00 Median : 4.380   
 Mean :54.32 Mean :38.32 Mean :19.59 Mean : 5.083   
 3rd Qu.:58.58 3rd Qu.:47.00 3rd Qu.:29.00 3rd Qu.: 6.210   
 Max. :78.25 Max. :65.00 Max. :55.00 Max. :35.510   
 yrsmarr spsepay male gdhlth   
 Min. : 0.00 Min. : 0 Min. :0.0000 Min. :0.0000   
 1st Qu.: 0.00 1st Qu.: 0 1st Qu.:0.0000 1st Qu.:1.0000   
 Median : 8.00 Median : 1000 Median :1.0000 Median :1.0000   
 Mean :11.18 Mean : 5270 Mean :0.5508 Mean :0.8853   
 3rd Qu.:18.25 3rd Qu.: 9395 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :43.00 Max. :50000 Max. :1.0000 Max. :1.0000   
 totwrk marr   
 Min. : 0 Min. :0.0000   
 1st Qu.:1611 1st Qu.:1.0000   
 Median :2300 Median :1.0000   
 Mean :2161 Mean :0.8177   
 3rd Qu.:2700 3rd Qu.:1.0000   
 Max. :6415 Max. :1.0000

## Emperical Framework

* Our true model is:

$$sleep = \beta\_0 + \beta\_1 hrwage + \beta\_2 exper + \beta\_3 age + \beta\_4 yrsmarr + \beta\_5 spsepay + \\ \beta\_6 male + \beta\_7 gdhlth + \beta\_8 totwrk + \beta\_9 marr + u$$

* Our estimated model is:

$$\hat{sleep} = 60.143 + 0.0104 hrwage + 0.1074 exper -0.0619 age -0.046 yrsmarr -0.0000403 spsepay \\ + 0.359 male -1.198 gdhlth -0.0026 totwrk + 2.035 marr + u$$

* Our “engine” we used was **OLS**
  1. We have normally distributed data.
  2. Our data is homoskedastic
  3. OLS is BLUE for our data.

The Standard Errors from our model are:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | hrwage | exper | age | yrsmarr | spsepay | male | gdhlth | totwrk | marr |
| 3.073 | 0.09039 | 0.1233 | 0.136 | 0.0396 | 4.92e-05 | 0.772 | 0.9639 | 0.0003494 | 1.034 |

# Classical Linear Model (CLM) Assumptions

1. MLR.1 Linear in Parameters
2. MLR.2 Random Sampling
3. MLR.3 No Perfect Collinearity
4. MLR.4 Zero Conditional Mean
5. MLR.5 Homoskedasticity
6. MLR.6 Normality

## MLR.1 Linear in Parameters

The model in the population can be written as:

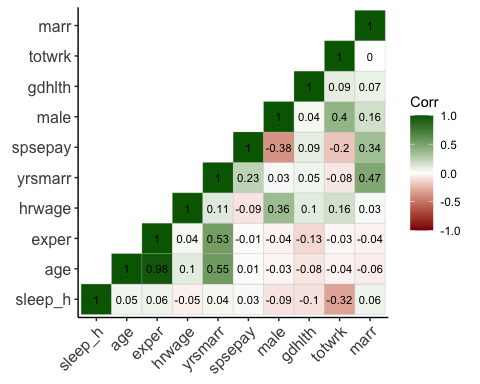
$$sleep\_h = \beta\_0 + \beta\_1 hrwage + \beta\_2 exper + \beta\_3 age + \beta\_4 yrsmarr + \beta\_5 spsepay + \\ \beta\_6 male + \beta\_7 gdhlth + \beta\_8 totwrk + \beta\_9 marr + u$$

* All of our parameters are linear

## MLR.2 Random Sampling

* Data is taken from Wooldridge dataset sleep75, which was randomly sampled with 706 observations

## MLR.3 No Perfect Collinearity



* No evidence of perfect collinearity
* Nothing is perfectly correlated with our dependent variable “sleep”
* No independent variable is constant.

## MLR.4 Zero Conditional Mean

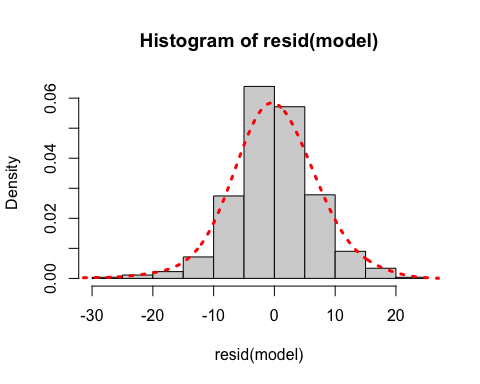
* We included multiple important variables in order to avoid violating the Zero Conditional Mean Assumption.

## MLR.5 Homoskedasticity

studentized Breusch-Pagan test  
  
data: model  
BP = 12.358, df = 9, p-value = 0.1939

* We can prove that the error has the same variance across all values of the explanatory variables
* According to the results of our Breusch-Pagan test, we have a p value of .1939, and so we fail to reject the null hypothesis

## MLR.6 Normality



* Normal distribution of our data is shown here.

# Results

Buckle Up…

## Regression Results

Call:  
lm(formula = sleep\_h ~ hrwage + exper + age + yrsmarr + spsepay +   
 male + gdhlth + totwrk + marr, data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-29.8233 -4.1655 -0.1986 4.1023 21.2346   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 60.1430865 3.0727328 19.573 < 2e-16 \*\*\*  
hrwage 0.0103861 0.0903905 0.115 0.9086   
exper 0.1074192 0.1232935 0.871 0.3840   
age -0.0618677 0.1359932 -0.455 0.6493   
yrsmarr -0.0459626 0.0395951 -1.161 0.2462   
spsepay -0.0000403 0.0000492 -0.819 0.4131   
male 0.3586507 0.7721346 0.464 0.6425   
gdhlth -1.1976270 0.9639335 -1.242 0.2146   
totwrk -0.0026290 0.0003494 -7.525 2.32e-13 \*\*\*  
marr 2.0352069 1.0340334 1.968 0.0496 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.787 on 522 degrees of freedom  
Multiple R-squared: 0.1222, Adjusted R-squared: 0.1071   
F-statistic: 8.077 on 9 and 522 DF, p-value: 2.981e-11

* Initial view of results: only two of our variables are statistically significant!
* is low at 12.22%, while adjusted is also low at 10.71%
* This tells us that the variation in our explanatory variables explain very little of the variation in our explained variable (sleep)
* We shouldn’t focus too much on the of the regression, but this could be a sign that other factors we haven’t analyzed are affecting sleep.

## Coefficients

Call:  
lm(formula = sleep\_h ~ hrwage + exper + age + yrsmarr + spsepay +   
 male + gdhlth + totwrk + marr, data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-29.8233 -4.1655 -0.1986 4.1023 21.2346   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 60.1430865 3.0727328 19.573 < 2e-16 \*\*\*  
hrwage 0.0103861 0.0903905 0.115 0.9086   
exper 0.1074192 0.1232935 0.871 0.3840   
age -0.0618677 0.1359932 -0.455 0.6493   
yrsmarr -0.0459626 0.0395951 -1.161 0.2462   
spsepay -0.0000403 0.0000492 -0.819 0.4131   
male 0.3586507 0.7721346 0.464 0.6425   
gdhlth -1.1976270 0.9639335 -1.242 0.2146   
totwrk -0.0026290 0.0003494 -7.525 2.32e-13 \*\*\*  
marr 2.0352069 1.0340334 1.968 0.0496 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.787 on 522 degrees of freedom  
Multiple R-squared: 0.1222, Adjusted R-squared: 0.1071   
F-statistic: 8.077 on 9 and 522 DF, p-value: 2.981e-11

* On average, in 1975, a one dollar increase in someone’s hourly wage is associated with an increase of 0.0104 hours of sleep per week
* Standard error is for hrwage is .0904
* This would equate to, on average, only 37.44 seconds extra sleep per week for a one dollar increase in hourly wage
* Rest of our coefficients show similar results in that they appear to, on average, have a small impact on sleep
* Only exception to this is totwrk and also marr, which we can see are statistically significant at varying significance levels

## F and t statistics

Call:  
lm(formula = sleep\_h ~ hrwage + exper + age + yrsmarr + spsepay +   
 male + gdhlth + totwrk + marr, data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-29.8233 -4.1655 -0.1986 4.1023 21.2346   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 60.1430865 3.0727328 19.573 < 2e-16 \*\*\*  
hrwage 0.0103861 0.0903905 0.115 0.9086   
exper 0.1074192 0.1232935 0.871 0.3840   
age -0.0618677 0.1359932 -0.455 0.6493   
yrsmarr -0.0459626 0.0395951 -1.161 0.2462   
spsepay -0.0000403 0.0000492 -0.819 0.4131   
male 0.3586507 0.7721346 0.464 0.6425   
gdhlth -1.1976270 0.9639335 -1.242 0.2146   
totwrk -0.0026290 0.0003494 -7.525 2.32e-13 \*\*\*  
marr 2.0352069 1.0340334 1.968 0.0496 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.787 on 522 degrees of freedom  
Multiple R-squared: 0.1222, Adjusted R-squared: 0.1071   
F-statistic: 8.077 on 9 and 522 DF, p-value: 2.981e-11

* The F statistics show us unsurprising results: it is statistically significant
* We ran iterations of our model that weren’t even jointly significant, so we made sure to include enough variables to fix this
* The T statistics are almost all very small with the exception of totwrk and marr

## Hypothesis Testing

Call:  
lm(formula = sleep\_h ~ hrwage + exper + age + yrsmarr + spsepay +   
 male + gdhlth + totwrk + marr, data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-29.8233 -4.1655 -0.1986 4.1023 21.2346   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 60.1430865 3.0727328 19.573 < 2e-16 \*\*\*  
hrwage 0.0103861 0.0903905 0.115 0.9086   
exper 0.1074192 0.1232935 0.871 0.3840   
age -0.0618677 0.1359932 -0.455 0.6493   
yrsmarr -0.0459626 0.0395951 -1.161 0.2462   
spsepay -0.0000403 0.0000492 -0.819 0.4131   
male 0.3586507 0.7721346 0.464 0.6425   
gdhlth -1.1976270 0.9639335 -1.242 0.2146   
totwrk -0.0026290 0.0003494 -7.525 2.32e-13 \*\*\*  
marr 2.0352069 1.0340334 1.968 0.0496 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.787 on 522 degrees of freedom  
Multiple R-squared: 0.1222, Adjusted R-squared: 0.1071   
F-statistic: 8.077 on 9 and 522 DF, p-value: 2.981e-11

Null Hypothesis is that hourly wage has no effect on the number of hours of sleep someone gets in a week

Alternative Hypothesis is that hourly wage does have an effect on the amount of sleep that someone gets in a week

* With a p-value of .9086, hrwage is only statistically significant as low as the 90.86% significance level
* We fail to reject the Null Hypothesis
* Take note: totwrk and marr are statistically significant

## Confidence Interval

2.5 % 97.5 %  
(Intercept) 54.1066445680 6.617953e+01  
hrwage -0.1671877597 1.879600e-01  
exper -0.1347932873 3.496317e-01  
age -0.3290289178 2.052934e-01  
yrsmarr -0.1237478438 3.182264e-02  
spsepay -0.0001369586 5.635889e-05  
male -1.1582223382 1.875524e+00  
gdhlth -3.0912927371 6.960387e-01  
totwrk -0.0033152840 -1.942665e-03  
marr 0.0038287326 4.066585e+00

* There is a 95% probability that the true population mean lies within our confidence interval here
* Unfortunately (but unsurprising, 0 lies within our 95% confidence interval), further providing evidence to reject the null hypothesis

# Conclusion

Does one’s hourly wage have some relationship with the amount of sleep that they get in a given week?

## The Downside

Even after exploring multiple potential transformations of our original model, adding more explanatory variables, and close analysis, our conclusion is that someone’s hourly wage has little to no effect on the amount of sleep that someone gets in a given week.

Moreover, the only variables that we could conclude had **ANY** statistically significant effect on sleep were how much someone works in a week and their marital status. We were originally discouraged by these results, as it made us feel like our original research question was a boring one.

## The Upside

These results, however, raise more interesting questions that call for further investigation. While knowing someone’s hourly wage might not help us that much in predicting how much sleep that they get in a week, perhaps the reason for this is that our preferences are more individualistic in a way that cannot be captured simply by how much money that someone makes.

Two different CEOs may have wildly different preferences on sleeping, and we could see a similarly complex situation at lower hourly wages too. So if we can make any conclusion from these results, it’s that we need to know **MORE.**

## Further Investigation

* Some other aspects of our data that would like to investigate would be researching individual preferences. Specifically, how does someone perceive their wage. Is it fair to them or not? Does this perception have a larger influence on sleep than wage?
* Are there other interactions that we could investigate that go into sleeping more than just wage?

## Thank You