

Assignment: Regression Analysis

Eldar Asimov

11/29/22

Contents

Question One	1
Question Two	2
Question Three	4
Appendix	6
Appendix 1: Total Work vs Sleep	6

Question One

The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990) to study the trade-off between time spent sleeping and working and to look at other factors affecting sleep:

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + u$$

where **sleep** and **totwrk** (total work) are measured in minutes per week and **educ** and **age** are measured in years.

- a. If adults trade off sleep for work, what is the sign of β_1 ?

The expected sign should be negative because an increase in sleep implies reduced hours of work and vice versa.

- b. What signs do you think β_2 and β_3 will have?

I hypothesise that β_2 is positive because skilled labour often performed by better educated individuals affords them more sleep time unlike for unskilled laborers.

For β_3 will have a negative sign indicating an inverse relationship between sleep and age. Older working class people may have less time to sleep.

- a. Using the data in SLEEP75.csv, the estimated equation is;

$$\text{sleep} = 3638.25 - 0.148\text{totwrk} - 11.13\text{educ} + 2.20\text{age},$$

$$n = 706$$

$$R^2 = 0.113$$

If someone works five more hours per week, by how many minutes is sleep predicted to fall? Is this a large trade-off?

If a person works five more hours per week, she or he sleeps approximately 44.4 LESS minutes.

- d. Discuss the sign and magnitude of the estimated coefficient on educ.

There is an inverse relationship between education and sleep (the negative sign). This means that more educated people tend to get less sleep on average, all else remaining the same. Specifically, a one year rise in education corresponds to 11.13 minutes of reduced sleep holding the other variables in the model constant.

- e. Would you say totwrk, educ, and age explain much of the variation in sleep? What other factors might affect the time spent sleeping? Are these likely to be correlated with totwrk?

The variables totwrk, educ, and age may not explain much of the variation in sleep going by the adjusted R^2 of 0.113. These three variables explain about 11.3% of the variation in sleep patterns. The other factors that may affect sleep quantity include the health of the individual, sleeping environment, and the use of drugs, including mild drugs like caffeine.

Question Two

Using the same data of previous problem (SLEEP75.csv), we obtain the estimated equation.

$$\text{sleep} = 3840.83 - 0.163\text{totwrk} - 11.71\text{educ} - 8.70\text{age} + 0.128\text{age}^2 + 87.75\text{male},$$

The standard errors for the coefficients are 235.11, 0.018, 5.86, 11.21, .134, and 34.33. The number of observations, n is 706, $R^2 = 0.123$ and adjusted $R^2 = 0.117$

The variable sleep is total minutes per week spent sleeping at night, totwrk is total weekly minutes spent working, educ and age are measured in years, and male is a gender dummy.

- a. All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

We can compute the p-value as follows:

$$p = \frac{x - \mu}{\delta^2}$$

```
my_coefficients <- c(3840.83, -0.163, -11.71, -8.7, 0.128, 87.75)
my_std_errors <- c(235.11, 0.018, 5.86, 11.21, .134, 34.33)

t_stat <- my_coefficients / my_std_errors

pt(t_stat, df = 706 - length(my_coefficients) - 1)
```

```
[1] 1.000000e+00 6.690128e-19 2.303588e-02 2.189784e-01 8.301026e-01
[6] 9.946018e-01
```

There is insufficient evidence that men sleep more than women. The evidence against the hypothesis that men sleep more than women is very high with a p-value of 0.995.

- b. Is there a statistically significant trade-off between working and sleeping? What is the estimated trade-off?

There is a statistically significant relationship between working and sleeping. Each additional minute of work corresponds to 0.163 less minutes of sleep.

- c. What other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping?

We could omit the variable age and age² are rerun the regressions. I have done so in the example below and indeed the regression remains significant going by the F-statistic. Hence, the removal of age did not significantly alter the regression interpretation. Hence we conclude that age did not significantly contribute to the model.

Call:

```
lm(formula = sleep ~ totwrk + educ + yngkid, data = sleep)
```

Residuals:

Min	1Q	Median	3Q	Max
-2441.43	-234.49	17.81	259.53	1329.69

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3756.13826	81.36283	46.165	<2e-16 ***
totwrk	-0.14945	0.01671	-8.946	<2e-16 ***
educ	-13.47707	5.71538	-2.358	0.0186 *
yngkid	-2.12600	47.44038	-0.045	0.9643

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 420 on 702 degrees of freedom

Multiple R-squared: 0.1104, Adjusted R-squared: 0.1066

F-statistic: 29.05 on 3 and 702 DF, p-value: < 2.2e-16

Question Three

Use the data in SLEEP75.csv 1 for this exercise. The equation of interest is;

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + u$$

- Estimate this equation separately for men and women and report the results in the usual form. Are there notable differences in the two estimated equations?

Table 1: Summary Statistics of Regression Analysis for Sleep Patterns

Dependent variable:		

Sleep: Male vs Female		
	(1)	(2)

Total work	-0.181*** (0.024)	-0.145*** (0.028)
Education	-13.036* (7.405)	-10.742 (9.611)
Age	3.387* (1.990)	-0.089 (2.426)
Age Squared	58.841	-98.502

	(58.669)	(92.674)
Young Kid	3,719.639*** (154.484)	3,680.998*** (183.635)

Observations	400	306
R2	0.156	0.090
Adjusted R2	0.148	0.077
Residual Std. Error	401.817 (df = 395)	438.234 (df = 301)
F Statistic	18.263*** (df = 4; 395)	7.398*** (df = 4; 301)
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

- b. Compute the Chow test for equality of the parameters in the sleep equation for men and women. Use the form of the test that adds male and the interaction terms *male * totwrk* ... *male * yngkid* and uses the full set of observations. What are the relevant df for the test?

The null hypothesis for the Chow test is that there is no structural break in the data, meaning no difference in the models for females and males.

M-fluctuation test

```
data: regression_chow_test
f(efp) = 1.1696, p-value = 0.7505
```

Should you reject the null at the 5% level?

We fail to reject the null hypothesis and conclude that there is no difference between the models involving males and females.

Appendix

Appendix 1: Total Work vs Sleep

