ECOM20001 Econometrics 1 Assignment 2

Name	**Student ID**
Yixin (Chelsea) Wu	1178379
Lang (Ron) Chen	1181506

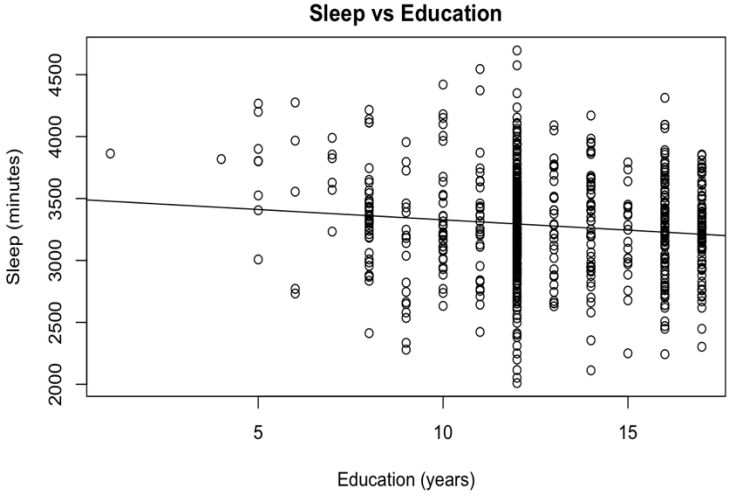
Q1

A typical observation based on the sample means: has 3281.487 minutes sleep (overnight) per week, 116.203 minutes napping per week, 2111.409 minutes working per week, 12.786 years of educational attainment, is 38.851 years old and have been married for 11.75 years. A typical observation also has an 89% chance of having self-reported “Excellent” or “Good” health, a 40% chance of living in a US urban area (SMSA), 22% chance of being part of a union, 13.1% chance of being self-employed, 82.1% chance of being married. They also have a 13.0% chance of having a young child less than 3 years old at home, and a 56.6% chance of being male.

Descriptive Statistics of Data

Statistic	N	Mean	St. Dev.	Min	Max
id	700	350.500	202.217	1	700
sleep	700	3,281.487	413.637	2,010	4,695
nap	700	116.203	224.433	0	2,250
totwrk	700	2,111.409	934.440	0	5,043
educ	700	12.786	2.788	1	17
age	700	38.851	11.351	23	65
gdhlth	700	0.890	0.313	0	1
smsa	700	0.400	0.490	0	1
union	700	0.220	0.415	0	1
selfe	700	0.131	0.338	0	1
marr	700	0.821	0.383	0	1
yrsmarr	700	11.753	11.609	0	43
yngkid	700	0.130	0.337	0	1
male	700	0.566	0.496	0	1

Q2



Q3.

The relationship between sleep and education is negative (-); the relationship between sleep and age is positive (+); the relationship between education and age is negative (-).

Given that $\hat{\beta} \rightarrow \beta + \rho xu \frac{\sigma u}{\sigma x}$ and sign (ρxu) = (+) * (-) = (-); $\hat{\beta} \rightarrow \beta + \rho xu \frac{\sigma u}{\sigma x}$

where ρxu < 0

$\hat{\beta} < \beta$

Where β is the true population value and $\hat{\beta}$ is the estimated regression coefficient for education (with dependent variable sleep)

There is a downwards bias for the slope coefficient on education in the single linear regression of sleep on education due to the effect from the omitted variable bias. Because the single linear regression coefficient is negative, the magnitude of the coefficient with negative omitted variable bias is too large compared to the coefficient of the education coefficient in a regression for sleep including age.

Q4.

	Dependent variable:				
	Sleep in minutes				
	(1)	(2)	(3)	(4)	(5)
educ	−16.66***	−14.47**	−11.90**	−10.39*	−9.95
	(5.58)	(5.79)	(5.94)	(6.05)	(6.06)
age		2.00	1.90	2.03	1.87
		(1.42)	(1.42)	(1.43)	(1.93)
gdhlth			−96.58*	−97.70*	−105.98**
			(51.17)	(51.20)	(51.74)
smsa				−59.93*	−57.11*
				(32.14)	(32.47)
union				−20.66	−21.47
				(38.04)	(38.15)
selfe				−5.20	−7.38
				(46.41)	(46.57)
marr					59.25
					(50.38)
yrsmarr					−0.06
					(2.06)
yngkid					−19.14
					(49.74)
Constant	3,494.53***	3,388.75***	3,445.65***	3,451.40***	3,413.69***
	(73.03)	(104.78)	(108.85)	(110.95)	(126.50)
Observations	700	700	700	700	700
R2	0.01	0.02	0.02	0.03	0.03
Adjusted R2	0.01	0.01	0.02	0.02	0.02
Residual Std. Error	411.32 (df = 698)	411.03 (df = 697)	410.27 (df = 696)	410.06 (df = 693)	410.35 (df = 690)
F Statistic	8.91*** (df = 1; 698)	5.45*** (df = 2; 697)	4.84*** (df = 3; 696)	3.04*** (df = 6; 693)	2.25** (df = 9; 690)
Note:					*p<0.1; **p<0.05; ***p<0.01

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Intercept (hetero robust SE)	76.0522	107.4112	118.6947	121.6361	137
Educ (HR SE)	5.6679	5.8798	5.7742	5.7703	5.8372
Age (HR SE)		1.3962	1.4030	1.4113	1.9483
gdhlth (HR SE)			57.2708	56.8294	58.3344
smsa (HR SE)				31.6521	32.2227
union (HR SE)				38.5050	38.5720
selfe (HR SE)				46.2244	46.4844
marr (HR SE)					52.9604
yrsmarr (HR SE)					2.1042
yngkid (HR SE)					53.3647
#Obs	700	700	700	700	700
Adjusted R-squared	0.0112	0.01258	0.01620	0.01722	0.01583

Q5

a) Yes, it matches. The coefficient of education changes from -16.663 in regression 1 to -14.470 in regression 2, which means the magnitude of education is too large in terms of magnitude in the initial single regression.

b) The coefficient of education changes from -14.470 in regression 2 to -11.900 in regression 3. As the heteroskedastic robust p-value is 0.03969 for education

coefficient in regression 3, this means that it is statistically significant from 0 at 5% level. This change in the coefficient occurs because of the addition of the “gdhlth” variable which further reduces omitted variable bias. Intuitively, it makes sense that the quality of health potentially has a positive relationship with the number of years of educational attainment and also has a negative relationship with the minutes of sleep overnight per week. This will create a downwards bias in the estimated coefficient of education if the “gdhlth” variable is not included in the regression due to the omitted variable bias in regression 2. When the “gdhlth” variable is included in the regression 3, the increase in the coefficient of education reflects the removal of the omitted variable bias from health condition.

c) In regression 3, the coefficient of education is -11.900 with p-value of 0.0397, which means it is statistically significantly different from 0 at 5% significance level.
In regression 4, the work context is added. The coefficient of education is -10.386 with p-value of 0.0723, which means that it is not statistically significant at 5% significance level.
In regression 5, the household variables are added to the regression on top of the work context. The coefficient of education is -9.954 with p-value of 0.0886, which means that it is not statistically significant at 5% significance level.
Overall, there is not much change in the value of the coefficient of education, but the coefficient becomes statistically insignificant after the addition of work context and work context plus household variables.

d) The sleep changes by -29.862 minutes overnight per week when education increases by 3 years. The 99% confidence interval for this predicted change is [-74.96600, 15.24825] minutes overnight per week.

e) After accounting for heteroskedasticity,

F-statistics	2.3055
Degrees of freedom	9 and 690
p-value	0.01482

The p-value is 0.01482 < 0.05 for the hypothesis that there is no statistically significant relationship amongst the independent variables and 9 dependent variables jointly; hence it indicates that the 9 variables in the regression 5 are jointly statistically significant in explaining the change in amount of sleep.

Q6

	Dependent variable:	
	Sleep/Nap in minutes (1)	nap (2)
educ	-9.95	-4.85
	(6.06)	(3.29)
age	1.87	0.99
	(1.93)	(1.04)
gdhlth	-105.98**	-30.87
	(51.74)	(28.06)
smsa	-57.11*	-0.78
	(32.47)	(17.61)
union	-21.47	-20.30
	(38.15)	(20.69)
selfe	-7.38	-59.03**
	(46.57)	(25.26)
marr	59.25	-45.91*
	(50.38)	(27.33)
yrsmarr	-0.06	0.05
	(2.06)	(1.12)
yngkid	-19.14	-24.18
	(49.74)	(26.98)
Constant	3,413.69***	219.97***
	(126.50)	(68.61)

Heteroskedasticity-robust standard errors	Regression 1	Regression 2
intercept	137.003	75.449
educ	5.837	3.834
age	1.948	1.314
gdhlth	58.334	33.598
smsa	32.222	17.525
union	38.572	19.083
selfe	46.484	20.025
marr	52.960	25.909
yrsmarr	2.104	1.110
yngkid	53.365	24.314

	Regression 1	Regression 2
Adjusted R-squared for model fit	0.0158	0.0165
Number of observations	700	700

magnitude of “educ”	9.95	4.85
P value of coefficient of ‘educ’	0.0886	0.206

The regression model 1 on sleep’s education coefficient has a magnitude of 9.95, with a p-value of 0.0886. The p-value is smaller than the significance level 0.1 (suggesting that we reject the null hypothesis that there is no relationship between education and sleep conditioned on all other included variables in the regression holding constant), hence we conclude that education has a statistically significant effect on sleep at 10% significance level.

The regression model 2 on nap’s education coefficient has a magnitude of 4.85, with a p-value of 0.206. The p-value is larger than the significance level 0.1 (suggesting that we do not reject the null hypothesis that there is no relationship between education and sleep conditioned on all other included variables in the regression holding constant), hence we conclude that education does not have a statistically significant effect on nap at 10% significance level.

Q7

	Dependent variable:		
	Sleep - full vs male vs female in minutes		
	(1)	(2)	(3)
totwrk	-0.13***	-0.17***	-0.12***
	(0.02)	(0.02)	(0.03)
educ	-9.43	-9.87	-12.48
	(5.79)	(7.47)	(9.23)
age	2.38	2.70	-0.17
	(1.84)	(2.52)	(2.83)
gdhlth	-62.66	-84.59	-25.30
	(49.71)	(68.24)	(74.07)
smsa	-69.35**	-69.26*	-86.93*
	(31.05)	(39.93)	(49.27)
union	-15.03	32.80	-102.69*
	(36.46)	(46.34)	(59.68)
selfe	20.80	77.27	-96.66
	(44.62)	(54.34)	(78.40)
marr	66.41	5.17	89.20
	(48.14)	(70.21)	(69.95)
yrsmarr	-1.59	0.63	-3.60
	(1.97)	(2.69)	(2.95)
yngkid	-19.42	38.04	-151.93*
	(47.52)	(57.28)	(86.46)
Constant	3,638.55***	3,756.92***	3,742.09***
	(123.93)	(168.45)	(198.18)
Observations	700	396	304
R2	0.11	0.16	0.11
Adjusted R2	0.10	0.14	0.08
Residual Std. Error	392.03 (df = 689)	377.83 (df = 385)	404.70 (df = 293)
F Statistic	8.92*** (df = 10; 689)	7.33*** (df = 10; 385)	3.68*** (df = 10; 293)

Heteroskedastic SE	Regression 1	Regression 2	Regression 3
intercept	131.280	185.670	208.156
totwrk	0.017	0.027	0.025
educ	5.590	7.920	8.221
age	1.855	2.390	3.131
gdhlth	56.814	71.873	87.072
smsa	30.909	39.986	46.939
union	37.052	45.585	62.570
selfe	47.301	56.151	91.014
marr	51.020	73.546	75.676
yrsmarr	2.021	2.510	3.351
yngkid	52.816	64.340	94.240

	Regression 1	Regression 2	Regression 3
Adjusted R-squared for fit	0.102	0.138	0.0811
Number of observation	700	396	304

Q8

a) P-value is 1.239 * e-13 < 5%
The coefficient on totwrk is statistically significant.
The predicted change in sleep from working one more hour per day for five days (60*5 = 300 minutes) with a given week: -39.561 minutes change

b)

F-statistics	1.374
Degrees of freedom	1 and 689
p-value	0.242

The p-value is larger than the significance level 5% which implies we don't reject the null that is the coefficient on marr is the same as the coefficient on yngkid. This means that it is unlikely being married and having a young child have a different impact on the minutes of sleep per week.

c) The p-value on the totwrk coefficient is 1.557e-09 and 3.218e-06 respectively in regression 1 and 2. This means that the number of minutes working per week has a statistically significant effect on sleep at a 5% significance level. If working one more hour per day for 5 days within a week, this will result in 49.69 minutes less of sleep for males, and 36.224 minutes less of sleep for females. In simplistic terms, males working more hours will on average lose more sleep time than a female working exactly the same number of extra hours.

```
setwd('/Users/tg.chenny/Desktop/1. University/1. Undergraduate/23. Econometrics 1/Asmt/Asmt 2')
data = read.csv('./as2_sleep.csv')

library(stargazer)

##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

library(AER)

## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival

library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Q1.

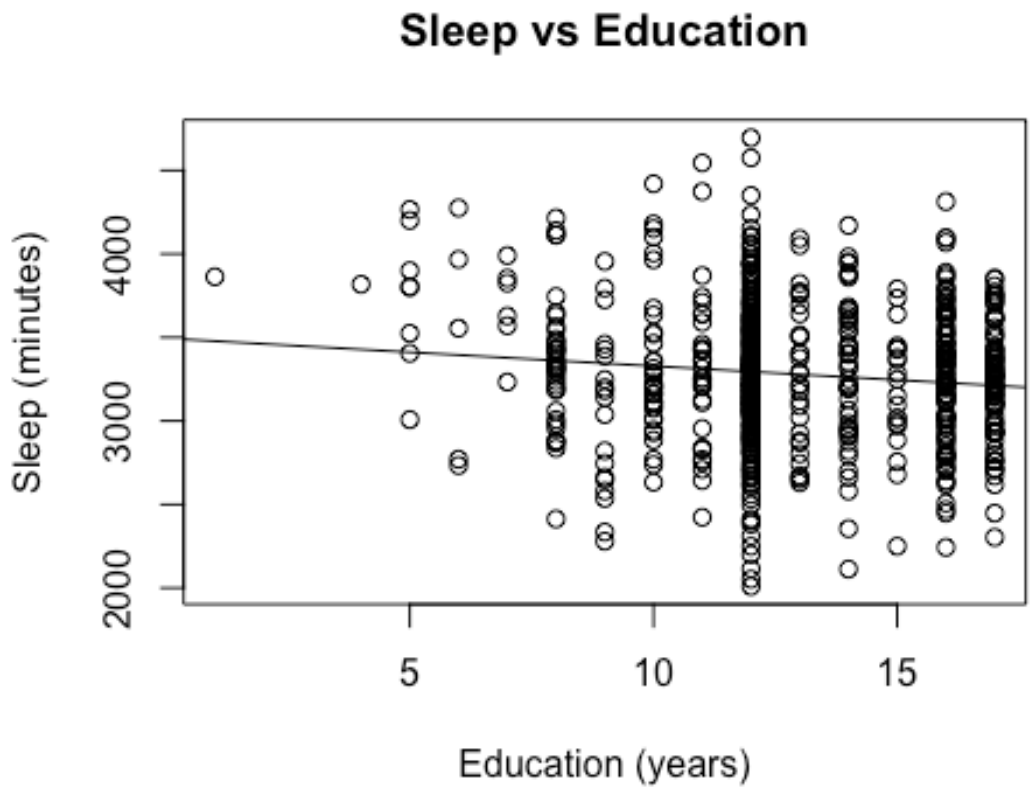
```
# get summary for data
stargazer(data,
          summary.stat = c("n", "mean", "sd", "min", "max"),
          type="text", title="Descriptive Statistics of Data",
          out="sumstats1.txt")

##
## Descriptive Statistics of Data
## ===========================================
## Statistic  N    Mean    St. Dev.  Min   Max
## -------------------------------------------
## id        700  350.500  202.217    1    700
## sleep     700 3,281.487 413.637  2,010 4,695
## nap       700  116.203  224.433    0   2,250
## totwrk    700 2,111.409 934.440    0   5,043
## educ      700  12.786    2.788     1    17
## age       700  38.851    11.351   23    65
## gdhlth    700   0.890    0.313     0     1
```

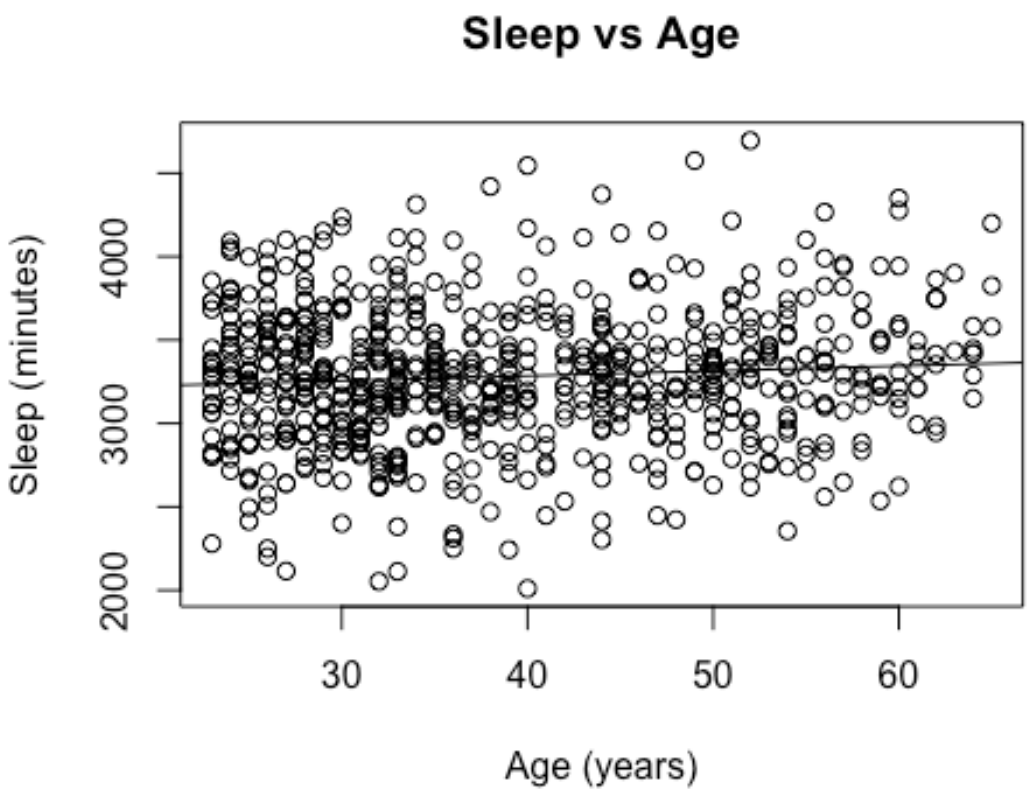
```
## smsa      700   0.400    0.490     0     1
## union     700   0.220    0.415     0     1
## selfe     700   0.131    0.338     0     1
## marr      700   0.821    0.383     0     1
## yrsmarr   700  11.753    11.609    0    43
## yngkid    700   0.130    0.337     0     1
## male      700   0.566    0.496     0     1
## ---------------------------------------------
```

Q2.

```
# plot sleep vs education
model2_1 = lm(sleep~educ, data=data)
plot(data$educ, data$sleep, main = 'Sleep vs Education', ylab = 'Sleep (minutes)', xlab = 'Education (years)')
abline(model2_1)
```




```
# plot sleep vs age
model2_2 = lm(sleep~age, data=data)
plot(data$age, data$sleep, main = 'Sleep vs Age', ylab = 'Sleep (minutes)', xlab = 'Age (years)')
abline(model2_2)
```




```r
# plot education vs age
model2_3 = lm(educ~age, data=data)
plot(data$age, data$educ, main = 'Education vs Age', ylab = 'Education (years)', xlab = 'Age (years)')
abline(model2_3)
```

Q4.

```
# create the five regressions
model4_1 = lm(sleep ~ educ, data=data)
model4_2 = lm(sleep ~ educ + age, data=data)
model4_3 = lm(sleep ~ educ + age + gdhlth, data=data)
model4_4 = lm(sleep ~ educ + age + gdhlth + smsa + union + selfe, data=data)
model4_5 = lm(sleep ~ educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid, data=data)

se1 = coef(summary(model4_1))[, "Std. Error"]
se2 = coef(summary(model4_2))[, "Std. Error"]
se3 = coef(summary(model4_3))[, "Std. Error"]
se4 = coef(summary(model4_4))[, "Std. Error"]
se5 = coef(summary(model4_5))[, "Std. Error"]

stargazer(model4_1, model4_2, model4_3, model4_4, model4_5, type="text",
          se=list(se1, se2, se3, se4, se5),
          digits=2,
          dep.var.labels = c("Sleep in minutes"),
          covariate.labels=
            c("educ",
              "age",
              "gdhlth",
              "smsa",
              "union",
              "selfe",
              "marr",
              "yrsmarr",
              "yngkid",
              "Constant"),
          out="reg_output.txt")

##
##
```

```
================================================================================================================
===============
##                                                    Dependent variable:
##                     ---------------------------------------------------------------------------------------
------------------
##                                                      Sleep in minutes
##                             (1)                 (2)                 (3)                 (4)
(5)
## ----------------------------------------------------------------------------------------------------------
------------------
## educ                     -16.66***           -14.47**            -11.90**            -10.39*
-9.95
##                           (5.58)              (5.79)              (5.94)              (6.05)
(6.06)
##
## age                                            2.00                1.90                2.03
1.87
##                                               (1.42)              (1.42)              (1.43)
(1.93)
##
## gdhlth                                                            -96.58*             -97.70*
-105.98**
##                                                                   (51.17)             (51.20)
(51.74)
##
## smsa                                                                                  -59.93*
-57.11*
##                                                                                       (32.14)
(32.47)
##
## union                                                                                 -20.66
-21.47
##                                                                                       (38.04)
(38.15)
##
## selfe                                                                                  -5.20
-7.38
##                                                                                       (46.41)
(46.57)
##
## marr
59.25
##
(50.38)
##
## yrsmarr
-0.06
##
(2.06)
##
## yngkid
-19.14
##
(49.74)
##
## Constant                 3,494.53***         3,388.75***         3,445.65***         3,451.40***
3,413.69***
##                           (73.03)             (104.78)            (108.85)            (110.95)
(126.50)
##
## ----------------------------------------------------------------------------------------------------------
------------------
## Observations                700                 700                 700                 700
700
## R2                          0.01                0.02                0.02                0.03
0.03
## Adjusted R2                 0.01                0.01                0.02                0.02
0.02
## Residual Std. Error   411.32 (df = 698)   411.03 (df = 697)   410.27 (df = 696)   410.06 (df = 693)
410.35 (df = 690)
## F Statistic         8.91*** (df = 1; 698) 5.45*** (df = 2; 697) 4.84*** (df = 3; 696) 3.04*** (df = 6; 693)
2.25** (df = 9; 690)
##
```

```
================================================================================================================================
===============
## Note:                                                                                                          *p<0.1;
**p<0.05; ***p<0.01

# get the heteroskedastic standard errors and the adjusted r-squares
paste("model4_1")

## [1] "model4_1"

coeftest(model4_1, vcov = vcovHC(model4_1, "HC1"))

##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 3494.5329    76.0522 45.9491 < 2.2e-16 ***
## educ         -16.6628     5.6679 -2.9399  0.003392 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model4_1)$adj.r.squared

## [1] 0.01119575

paste("model4_2")

## [1] "model4_2"

coeftest(model4_2, vcov = vcovHC(model4_2, "HC2"))

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3388.7500   107.4112 31.5493  < 2e-16 ***
## educ         -14.4689     5.8798 -2.4608  0.01411 *
## age            2.0008     1.3962  1.4330  0.15230
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model4_2)$adj.r.squared

## [1] 0.01258113

paste("model4_3")

## [1] "model4_3"

coeftest(model4_3, vcov = vcovHC(model4_3, "HC1"))

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3445.6540   118.6947 29.0296  < 2e-16 ***
## educ         -11.8999     5.7742 -2.0609  0.03969 *
## age            1.9032     1.4030  1.3565  0.17537
## gdhlth       -96.5830    57.2708 -1.6864  0.09216 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model4_3)$adj.r.squared

## [1] 0.01619736

paste("model4_4")

## [1] "model4_4"
```

```
coeftest(model4_4, vcov = vcovHC(model4_4, "HC1"))

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3451.3987   121.6361 28.3748  < 2e-16 ***
## educ         -10.3855     5.7703 -1.7998  0.07233 .
## age            2.0342     1.4113  1.4413  0.14995
## gdhlth       -97.7009    56.8294 -1.7192  0.08602 .
## smsa         -59.9327    31.6521 -1.8935  0.05871 .
## union        -20.6622    38.5050 -0.5366  0.59171
## selfe         -5.1959    46.2244 -0.1124  0.91053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model4_4)$adj.r.squared

## [1] 0.01721718

paste("model4_5")

## [1] "model4_5"

coeftest(model4_5, vcov = vcovHC(model4_5, "HC1"))

##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept) 3413.694713  137.003083 24.9169  < 2e-16 ***
## educ          -9.953959    5.837230 -1.7053  0.08860 .
## age            1.866201    1.948274  0.9579  0.33846
## gdhlth      -105.984385   58.334354 -1.8168  0.06967 .
## smsa         -57.106643   32.222649 -1.7723  0.07679 .
## union        -21.473167   38.571985 -0.5567  0.57791
## selfe         -7.383275   46.484433 -0.1588  0.87385
## marr          59.245963   52.960432  1.1187  0.26366
## yrsmarr       -0.064546    2.104273 -0.0307  0.97554
## yngkid       -19.139604   53.364684 -0.3587  0.71996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model4_5)$adj.r.squared

## [1] 0.01582727
```

Q5. d)

```
# calculated estimated 3 unit change in sleep when education increases by 3
-9.953959 * 3

## [1] -29.86188

# CI
3*(-9.952959 + c(-1, 1) * qnorm(1-(0.01/2), 0, 1) * 5.83723)

## [1] -74.96600  15.24825

# calculate heteroskedastic F-test
waldtest(model4_5, vcov = vcovHC(model4_5, "HC1"))

## Wald test
##
## Model 1: sleep ~ educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr +
##     yngkid
## Model 2: sleep ~ 1
##   Res.Df Df      F  Pr(>F)
## 1    690
## 2    699 -9 2.3055 0.01482 *
```

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

Q6.

```
# fit 2 linear regressions
model6_1 = lm(sleep ~ educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid, data = data)
model6_2 = lm(nap ~ educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid, data = data)

se6_1 = coef(summary(model6_1))[, "Std. Error"]
se6_2 = coef(summary(model6_2))[, "Std. Error"]

stargazer(model6_1, model6_2, type="text",
          se=list(se6_1, se6_2),
          digits=2,
          dep.var.labels = c("Sleep/Nap in minutes"),
          covariate.labels=
            c("educ",
              "age",
              "gdhlth",
              "smsa",
              "union",
              "selfe",
              "marr",
              "yrsmarr",
              "yngkid",
              "Constant"),
          out="reg_output2.txt")
```

```
##
## ===============================================================
##                                      Dependent variable:
##                                  ----------------------------
##                                  Sleep/Nap in minutes    nap
##                                          (1)             (2)
## ---------------------------------------------------------------
## educ                                    -9.95           -4.85
##                                         (6.06)         (3.29)
##
## age                                      1.87           0.99
##                                         (1.93)         (1.04)
##
## gdhlth                                -105.98**        -30.87
##                                        (51.74)         (28.06)
##
## smsa                                   -57.11*          -0.78
##                                        (32.47)         (17.61)
##
## union                                   -21.47         -20.30
##                                        (38.15)         (20.69)
##
## selfe                                    -7.38        -59.03**
##                                        (46.57)         (25.26)
##
## marr                                     59.25          -45.91*
##                                        (50.38)         (27.33)
##
## yrsmarr                                  -0.06           0.05
##                                         (2.06)         (1.12)
##
## yngkid                                  -19.14         -24.18
##                                        (49.74)         (26.98)
##
## Constant                              3,413.69***      219.97***
##                                         (126.50)        (68.61)
##
## ---------------------------------------------------------------
## Observations                             700             700
## R2                                       0.03           0.03
## Adjusted R2                              0.02           0.02
## Residual Std. Error (df = 690)          410.35         222.57
## F Statistic (df = 9; 690)               2.25**         2.31**
```

```
## ============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01
```

```
# get the heteroskedastic standard errors and the adjusted r-squares
paste("model6_1")
```

```
## [1] "model6_1"
```

```
coeftest(model6_1, vcov = vcovHC(model6_1, "HC1"))
```

```
##
## t test of coefficients:
##
##               Estimate  Std. Error t value Pr(>|t|)
## (Intercept) 3413.694713  137.003083 24.9169  < 2e-16 ***
## educ          -9.953959    5.837230 -1.7053  0.08860 .
## age            1.866201    1.948274  0.9579  0.33846
## gdhlth      -105.984385   58.334354 -1.8168  0.06967 .
## smsa         -57.106643   32.222649 -1.7723  0.07679 .
## union        -21.473167   38.571985 -0.5567  0.57791
## selfe         -7.383275   46.484433 -0.1588  0.87385
## marr          59.245963   52.960432  1.1187  0.26366
## yrsmarr       -0.064546    2.104273 -0.0307  0.97554
## yngkid       -19.139604   53.364684 -0.3587  0.71996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model6_1)$adj.r.squared
```

```
## [1] 0.01582727
```

```
paste("model6_2")
```

```
## [1] "model6_2"
```

```
coeftest(model6_2, vcov = vcovHC(model6_2, "HC1"))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 219.969211  75.448964  2.9155 0.003667 **
## educ         -4.851031   3.834116 -1.2652 0.206217
## age           0.992939   1.313537  0.7559 0.449951
## gdhlth      -30.869309  33.597909 -0.9188 0.358528
## smsa         -0.784175  17.524827 -0.0447 0.964322
## union       -20.302084  19.083461 -1.0639 0.287765
## selfe       -59.026426  20.024713 -2.9477 0.003310 **
## marr        -45.908078  25.908869 -1.7719 0.076851 .
## yrsmarr       0.046395   1.109580  0.0418 0.966660
## yngkid      -24.177402  24.313611 -0.9944 0.320378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model6_2)$adj.r.squared
```

```
## [1] 0.01654254
```

Q7.

```
# get male only and female only data
maledata = data %>% filter(male == 1)
femaledata = data %>% filter(male == 0)

# fit three linear regressions
model7_1 = lm(sleep ~ totwrk + educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid, data=data)
model7_2 = lm(sleep ~ totwrk + educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid,
data=maledata)
model7_3 = lm(sleep ~ totwrk + educ + age + gdhlth + smsa + union + selfe + marr + yrsmarr + yngkid,
data=femaledata)
```

```r
se7_1 = coef(summary(model7_1))[, "Std. Error"]
se7_2 = coef(summary(model7_2))[, "Std. Error"]
se7_3 = coef(summary(model7_3))[, "Std. Error"]

stargazer(model7_1, model7_2, model7_3, type="text",
          se=list(se7_1, se7_2, se7_3),
          digits=2,
          dep.var.labels = c("Sleep - full vs male vs female in minutes"),
          covariate.labels=
            c("totwrk",
              "educ",
              "age",
              "gdhlth",
              "smsa",
              "union",
              "selfe",
              "marr",
              "yrsmarr",
              "yngkid",
              "Constant"),
          out="reg_output2.txt")
```

```
##
## ================================================================================================
##                                          Dependent variable:
##                      ---------------------------------------------------------------------------
##                                  Sleep - full vs male vs female in minutes
##                                (1)                    (2)                    (3)
## ------------------------------------------------------------------------------------------------
## totwrk                      -0.13***               -0.17***               -0.12***
##                              (0.02)                 (0.02)                 (0.03)
##
## educ                         -9.43                  -9.87                  -12.48
##                              (5.79)                 (7.47)                 (9.23)
##
## age                           2.38                   2.70                   -0.17
##                              (1.84)                 (2.52)                 (2.83)
##
## gdhlth                       -62.66                 -84.59                 -25.30
##                             (49.71)                (68.24)                (74.07)
##
## smsa                        -69.35**               -69.26*                -86.93*
##                             (31.05)                (39.93)                (49.27)
##
## union                        -15.03                 32.80                  -102.69*
##                             (36.46)                (46.34)                (59.68)
##
## selfe                        20.80                  77.27                  -96.66
##                             (44.62)                (54.34)                (78.40)
##
## marr                         66.41                   5.17                   89.20
##                             (48.14)                (70.21)                (69.95)
##
## yrsmarr                      -1.59                   0.63                   -3.60
##                              (1.97)                 (2.69)                 (2.95)
##
## yngkid                       -19.42                 38.04                  -151.93*
##                             (47.52)                (57.28)                (86.46)
##
## Constant                 3,638.55***            3,756.92***            3,742.09***
##                            (123.93)               (168.45)               (198.18)
##
## ------------------------------------------------------------------------------------------------
## Observations                  700                    396                    304
## R2                            0.11                   0.16                   0.11
## Adjusted R2                   0.10                   0.14                   0.08
## Residual Std. Error   392.03 (df = 689)      377.83 (df = 385)      404.70 (df = 293)
## F Statistic         8.92*** (df = 10; 689) 7.33*** (df = 10; 385) 3.68*** (df = 10; 293)
## ================================================================================================
## Note:                                                              *p<0.1; **p<0.05; ***p<0.01
```

```r
# get the heteroskedastic standard errors and the adjusted r-squares
paste("model7_1")
```

```
## [1] "model7_1"

coeftest(model7_1, vcov = vcovHC(model7_1, "HC1"))

##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 3638.55434  131.28032 27.7159 < 2.2e-16 ***
## totwrk        -0.13187    0.01743 -7.5656 1.239e-13 ***
## educ          -9.42615    5.59042 -1.6861   0.09222 .
## age            2.38314    1.85523  1.2845   0.19938
## gdhlth       -62.66081   56.81368 -1.1029   0.27045
## smsa         -69.35014   30.90852 -2.2437   0.02517 *
## union        -15.03156   37.05240 -0.4057   0.68510
## selfe         20.79807   47.30139  0.4397   0.66030
## marr          66.41315   51.01990  1.3017   0.19345
## yrsmarr       -1.58715    2.02063 -0.7855   0.43245
## yngkid       -19.42472   52.81554 -0.3678   0.71315
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model7_1)$adj.r.squared

## [1] 0.1017385

paste("model7_2")

## [1] "model7_2"

coeftest(model7_2, vcov = vcovHC(model7_2, "HC1"))

##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept) 3756.920040  185.669970 20.2344 < 2.2e-16 ***
## totwrk        -0.165631    0.026765 -6.1884 1.557e-09 ***
## educ          -9.870582    7.919909 -1.2463   0.21341
## age            2.702842    2.390033  1.1309   0.25881
## gdhlth       -84.592773   71.872770 -1.1770   0.23993
## smsa         -69.259073   39.985988 -1.7321   0.08406 .
## union         32.799689   45.584805  0.7195   0.47225
## selfe         77.268597   56.151181  1.3761   0.16960
## marr           5.167284   73.545652  0.0703   0.94402
## yrsmarr        0.627117    2.509684  0.2499   0.80281
## yngkid        38.035853   64.340398  0.5912   0.55476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model7_2)$adj.r.squared

## [1] 0.1381759

paste("model7_3")

## [1] "model7_3"

coeftest(model7_3, vcov = vcovHC(model7_3, "HC1"))

##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept) 3742.090191  208.156090 17.9773 < 2.2e-16 ***
## totwrk        -0.120748    0.025431 -4.7480 3.218e-06 ***
## educ         -12.476643    8.220623 -1.5177   0.13016
## age           -0.167554    3.130949 -0.0535   0.95736
## gdhlth       -25.300032   87.072287 -0.2906   0.77159
## smsa         -86.926197   46.938753 -1.8519   0.06505 .
## union       -102.687892   62.570389 -1.6412   0.10184
```

```
## selfe        -96.664290   91.014349 -1.0621   0.28908
## marr          89.197668   75.676452  1.1787   0.23949
## yrsmarr       -3.596078    3.351312 -1.0730   0.28414
## yngkid      -151.925296   94.239788 -1.6121   0.10801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model7_3)$adj.r.squared

## [1] 0.08114583
```

Q8. A)

```
# get the estimate for working 300 more minutes per week
60*5*-0.13187

## [1] -39.561

# get the heteroskedastic linear hypothesis test
linearHypothesis(model7_1, "marr=yngkid", var=vcovHC(model7_1, "HC1"))

## Linear hypothesis test
##
## Hypothesis:
## marr - yngkid = 0
##
## Model 1: restricted model
## Model 2: sleep ~ totwrk + educ + age + gdhlth + smsa + union + selfe +
##     marr + yrsmarr + yngkid
##
##   Res.Df       RSS Df Sum of Sq      F Pr(>F)
## 1    690 106102808
## 2    689 105891658  1    211150 1.3739 0.2416

# get the estimates for working 300 more mintues per week for male and female
-0.165631 * 60 * 5

## [1] -49.6893

-0.120748 * 60 * 5

## [1] -36.2244

-0.165631 * 60 * 5 + 0.120748 * 60 * 5

## [1] -13.4649
```