Escuela Superior de Física y Matemáticas

ECONOMETRÍA

Alumno:
Roberto Carlos Santos Alonzo

Actividad 5

Diciembre 2020

Ejercicios

1. El archivo cps12.RData contiene información relativa al salario de trabajadoras(es), su edad y nivel de estudio.

1 Estimar los modelos:

$$log(ahe_i) = \beta_0 + \beta_1 age_i + u_i \quad (1)$$
$$log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + u_i \quad (2)$$

¿Qué modelo se ajusta mejor?

```
load("C:/Users/81799/Downloads/cps12.RData")
library(lmtest)
library(sandwich)
library(car)
library(stargazer)
```

Solución:

1. $log(ahe_i) = \beta_0 + \beta_1 age_i + u_i$

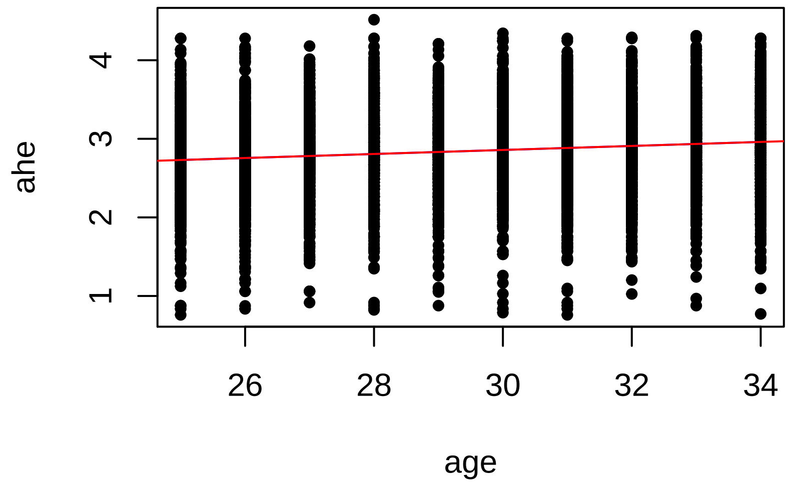
```
ahe <- I(log(cps12$ahe))
age <- cps12$age
modelo1 <- data.frame(ahe,age)

#creo un data frame, el cual todos los datos de ahe le aplica el ln, o sea ln(ahe)

ols.1 <- lm(ahe~age ,data = modelo1);ols.1

##
## Call:
## lm(formula = ahe ~ age, data = modelo1)
##
## Coefficients:
## (Intercept)          age
##     2.09078      0.02558

# Hace la regresión del data frame modelo1(logaritmo(ahe) sobre age)
```

```
coeftest(ols.1, vcov=vcovHC, type="HC1")

##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 2.0907821  0.0642515  32.541 < 2.2e-16 ***
## age         0.0255768  0.0021612  11.835 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
stargazer(ols.1, type="text")

##
```

```
## ===============================================
##                          Dependent variable:
##                      ---------------------------
##                                  ahe
## -----------------------------------------------
## age                           0.026***
##                                (0.002)
##
## Constant                      2.091***
##                                (0.064)
##
## -----------------------------------------------
## Observations                    7,440
## R2                              0.019
## Adjusted R2                     0.018
## Residual Std. Error      0.528 (df = 7438)
## F Statistic          140.529*** (df = 1; 7438)
## ===============================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
plot(modelo1$age, modelo1$ahe,pch=20, xlab="age", ylab="ahe",

     xlim = c(min(modelo1$age),max(modelo1$age)),

     ylim = c(min(modelo1$ahe),max(modelo1$ahe)))

# Hago la grafica con los datos del data frame modelo1

datos=data.frame(age=seq(min(cps12$age),max(cps12$age),len=7440))

#ordeno de menor a mayor, los datos del data frame original

modl1 <- data.frame(predict(ols.1, newdata = datos) ,datos$age)

#Data frame en el cual tiene los datos estimados de la regresión log(ahe) sobre age

#Este es un data frame, de los estimado de age y ahe.

lines(datos$age ,predict(ols.1, newdata = datos) ,col="blue")
abline(ols.1, col = "red")
```




```r
#Hago la linea de regresión conforme a los datos del data frame datos
```

El 1er modelos es de la forma log-lineal.De este modelo tenemos que $\bar{R}^2$ es:

```r
summary(ols.1)$adj.r.squared

## [1] 0.01841109
```

2. $log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + u_i$

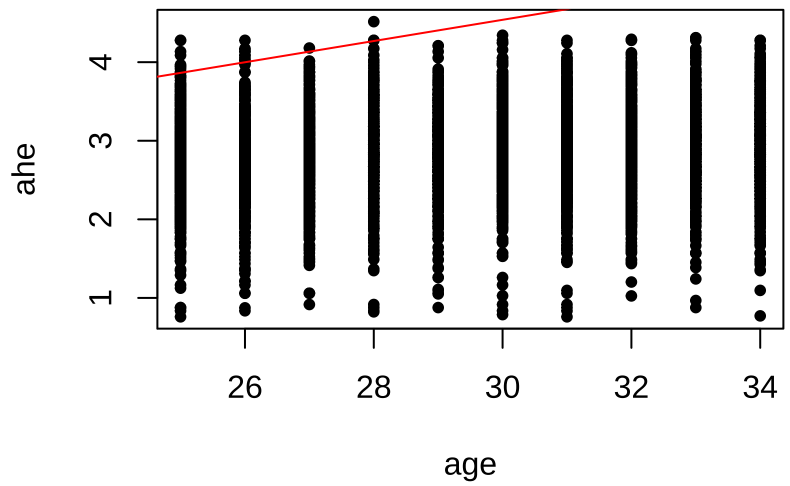
```r
ahe <- I(log(cps12$ahe))
age <- cps12$age
agecua <- I(cps12$age^2)
modelo2 <- data.frame(ahe,age,agecua)

#creo un data frame, el cual todos los datos de ahe le aplica el ln, o sea ln(ahe), y a los

#datos age los eleva al cuadrado (age^2)

ols.2 <- lm(ahe~age+agecua ,data = modelo2);ols.2

##
## Call:
## lm(formula = ahe ~ age + agecua, data = modelo2)
##
## Coefficients:
## (Intercept)          age       agecua
##    0.488087     0.135037    -0.001852

# Hace la regresión del data frame modelo1(logaritmo(ahe) sobre age+age^2)
```

```r
coeftest(ols.2, vcov=vcovHC, type="HC1")

##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept)  0.48808733  0.74140522  0.6583 0.510348
## age          0.13503739  0.05052795  2.6725 0.007545 **
## agecua      -0.00185171  0.00085461 -2.1667 0.030287 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
stargazer(ols.2, type="text")

##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                 ahe
## -----------------------------------------------
## age                          0.135***
##                               (0.050)
##
## agecua                       -0.002**
##                               (0.001)
##
## Constant                       0.488
##                               (0.740)
##
## -----------------------------------------------
## Observations                   7,440
## R2                             0.019
## Adjusted R2                    0.019
## Residual Std. Error      0.528 (df = 7437)
## F Statistic          72.662*** (df = 2; 7437)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```r
plot(modelo2$age, modelo2$ahe,pch=20, xlab="age", ylab="ahe",

     xlim = c(min(modelo2$age),max(modelo2$age)),

     ylim = c(min(modelo2$ahe),max(modelo2$ahe)))

# Hago la grafica con los datos del data frame modelo2

datos=data.frame(age=seq(min(cps12$age),max(cps12$age),len=7440))

#ordeno de menor a mayor, los datos del data frame original

modl2 <- data.frame(predict(ols.2, newdata = datos) ,datos$age)
```

```r
#Data frame en el cual tiene los datos estimados de la regresión log(ahe) sobre age + age^2

#Este es un data frame, de los estimado de age y ahe.

abline(ols.2, col = "red")
```




```r
#Hago la linea de regresión conforme a los datos del data frame datos
```

Del 2do modelo tenemos que $\bar{R}^2$ es:

```r
summary(ols.2)$adj.r.squared

## [1] 0.01890236
```

Si utilizamos el valor de $\bar{R}^2$ para comparar los modelos, podemos observar que la diferencia es de:

```r
summary(ols.2)$adj.r.squared-summary(ols.1)$adj.r.squared

## [1] 0.0004912689
```

$\therefore$ el segundo modelo tiene un $\bar{R}^2$ mayor, con esto concluimos que entonces este modelo se ajusta mejor a los datos.

2. Considerar los siguientes modelos

$$ahe_i = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i \tag{3}$$
$$log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i \tag{4}$$
$$log(ahe_i) = \beta_0 + \beta_1 log(age_i) + \beta_2 female_i + \beta_3 bachelor_i + u_i \tag{5}$$
$$log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 female_i + \beta_4 bachelor_i + u_i \tag{6}$$

La regresión de cada modelo, nos queda:

1.
```r
ols.1 <- lm(ahe ~ age + female + bachelor, data = cps12)
stargazer(ols.1, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                 ahe
## -----------------------------------------------
## age                          0.510***
##                               (0.040)
##
## female                       -3.810***
##                               (0.230)
##
## bachelor                     8.319***
##                               (0.227)
##
## Constant                       1.866
##                               (1.188)
##
## -----------------------------------------------
## Observations                   7,440
## R2                             0.180
## Adjusted R2                    0.180
## Residual Std. Error      9.678 (df = 7436)
## F Statistic          544.495*** (df = 3; 7436)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

2.
```r
ols.2 <- lm(I(log(ahe)) ~ age + female + bachelor, data = cps12)
stargazer(ols.2, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                             I(log(ahe))
## -----------------------------------------------
## age                          0.026***
##                               (0.002)
##
## female                       -0.192***
##                               (0.011)
##
## bachelor                     0.438***
##                               (0.011)
##
## Constant                     1.941***
##                               (0.059)
##
## -----------------------------------------------
## Observations                   7,440
```

```
## R2                              0.196
## Adjusted R2                     0.196
## Residual Std. Error       0.478 (df = 7436)
## F Statistic           605.726*** (df = 3; 7436)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

3.
```r
ols.3 <- lm(I(log(ahe)) ~ I(log(age)) + female + bachelor, data = cps12)
stargazer(ols.3, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                             I(log(ahe))
## -----------------------------------------------
## I(log(age))                 0.753***
##                              (0.057)
##
## female                      -0.192***
##                              (0.011)
##
## bachelor                    0.438***
##                              (0.011)
##
## Constant                      0.150
##                              (0.194)
##
## -----------------------------------------------
## Observations                  7,440
## R2                            0.197
## Adjusted R2                   0.196
## Residual Std. Error       0.478 (df = 7436)
## F Statistic           606.413*** (df = 3; 7436)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

4.
```r
ols.4 <- lm(I(log(ahe)) ~ age+I(age^2)+ female + bachelor, data = cps12)
stargazer(ols.4, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                             I(log(ahe))
## -----------------------------------------------
## age                          0.104**
##                              (0.046)
##
## I(age2)                      -0.001*
##                              (0.001)
```

```
##
## female                        -0.192***
##                                (0.011)
##
## bachelor                      0.437***
##                                (0.011)
##
## Constant                        0.792
##                                (0.670)
##
## ----------------------------------------------
## Observations                    7,440
## R2                              0.197
## Adjusted R2                     0.196
## Residual Std. Error      0.478 (df = 7435)
## F Statistic          455.156*** (df = 4; 7435)
## ==============================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

2.1. Si la variable *age* aumenta de 25 a 26, en cuanto cambia el ingreso, evaluar para cada modelo

2.1.1. $ahe_i = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 25 a 26 en age.
Entonces $\Delta\hat{Y} = f(age = 26, female, bachelor) - f(age = 25, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 26 + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 25 + \hat{\beta}_2 + \hat{\beta}_3) = \hat{\beta}_1(26 - 25)$$
$$\Delta\hat{Y} = \hat{\beta}_1$$

```
DY.1 <-ols.1$coef[2]; DY.1

##      age
## 0.510286
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 26 y una con un promedio de edad de 25 es de 0.510286

2.1.2. $log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 25 a 26 en age.
Entonces $\Delta\hat{Y} = f(age = 26, female, bachelor) - f(age = 25, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 26 + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 25 + \hat{\beta}_2 + \hat{\beta}_3) = \hat{\beta}_1(26 - 25)$$
$$\Delta\hat{Y} = \hat{\beta}_1$$

```
DY.2 <-ols.2$coef[2]; DY.2

##        age
## 0.02551788
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 26 y una con un promedio de edad de 25 es de 0.02551788

2.1.3. $log(ahe_i) = \beta_0 + \beta_1 log(age_i) + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 25 a 26 en age.
Entonces $\Delta\hat{Y} = f(age = 26, female, bachelor) - f(age = 25, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 log(26) + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 log(25) + \hat{\beta}_2 + \hat{\beta}_3)$$
$$\Delta\hat{Y} = \hat{\beta}_1 [log(26) - log(25)]$$

```
DY.3 <- ols.3$coef[2]*(log(26)-log(25)); DY.3

## I(log(age))
##  0.02953087
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 26 y una con un promedio de edad de 25 es de 0.02953087

2.1.4. $log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 female_i + \beta_4 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 26 a 25 en age.
Entonces $\Delta\hat{Y} = f(age = 26, female, bachelor) - f(age = 25, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 26 + \beta_2 \cdot \hat{(26)}^2 + \hat{\beta}_3 + \hat{\beta}_4 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 25 + \beta_2 \cdot \hat{(25)}^2 + \hat{\beta}_3 + \hat{\beta}_4)$$
$$\Delta\hat{Y} = \hat{\beta}_1 + 51 \cdot \hat{\beta}_2$$

```
DY.4 <- ols.4$coef[2]+51*ols.4$coef[3]; DY.4

##        age
## 0.03629504
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 26 y una con un promedio de edad de 25 es de 0.03629504

2.2. Si *age* aumenta de 33 a 34, ¿ cuál es la variación en el ingreso?, analizar para cada modelo

2.2.1. $ahe_i = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 33 a 34 en age.
Entonces $\Delta\hat{Y} = f(age = 34, female, bachelor) - f(age = 33, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 34 + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 33 + \hat{\beta}_2 + \hat{\beta}_3) = \hat{\beta}_1(34 - 33)$$
$$\Delta\hat{Y} = \hat{\beta}_1$$

```
DY.11 <-ols.1$coef[2]; DY.11

##      age
## 0.510286
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 34 y una con un promedio de edad de 33 es de 0.510286

2.2.2. $log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 33 a 34 en age.
Entonces $\Delta\hat{Y} = f(age = 34, female, bachelor) - f(age = 33, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 34 + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 33 + \hat{\beta}_2 + \hat{\beta}_3) = \hat{\beta}_1(34 - 33)$$
$$\Delta\hat{Y} = \hat{\beta}_1$$

```
DY.22 <-ols.2$coef[2]; DY.22

##        age
## 0.02551788
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 34 y una con un promedio de edad de 33 es de 0.02551788

2.2.3. $log(ahe_i) = \beta_0 + \beta_1 log(age_i) + \beta_2 female_i + \beta_3 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 33 a 34 en age.
Entonces $\Delta\hat{Y} = f(age = 34, female, bachelor) - f(age = 33, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 log(34) + \hat{\beta}_2 + \hat{\beta}_3 - (\hat{\beta}_0 + \hat{\beta}_1 log(33) + \hat{\beta}_2 + \hat{\beta}_3)$$
$$\Delta\hat{Y} = \hat{\beta}_1 \left[log(34) - log(33)\right]$$

```
DY.33 <- ols.3$coef[2]*(log(34)-log(33)); DY.33

## I(log(age))
##  0.02247751
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 34 y una con un promedio de edad de 33 es de 0.02247751

2.2.4. $log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 female_i + \beta_4 bachelor_i + u_i$

Cambio en ahe debido a un cambio de 33 a 34 en age.
Entonces $\Delta\hat{Y} = f(age = 34, female, bachelor) - f(age = 33, female, bachelor)$
Por el modelo, entonces tenemos que:

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 34 + \beta_2 \cdot \hat{(34)}^2 + \hat{\beta}_3 + \hat{\beta}_4 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 33 + \beta_2 \cdot \hat{(33)}^2 + \hat{\beta}_3 + \hat{\beta}_4)$$
$$\Delta\hat{Y} = \hat{\beta}_1 + 67 \cdot \hat{\beta}_2$$

```
DY.44 <- ols.4$coef[2]+67*ols.4$coef[3]; DY.44

##        age
## 0.01504019
```

es decir, la diferencia esperada en el salario de los trabajadoras(es) entre una persona con un promedio de edad de 34 y una con un promedio de edad de 33 es de 0.01504019

3. Explicar los siguientes incisos:

3.1. Prefieres la regresión c) o la regresión b)

Si utilizamos el valor de $\bar{R}^2$ para comparar los modelos, tenemos que:

- $\bar{R}^2$ del modelo c) (ols.3)

```
summary(ols.3)$adj.r.squared

## [1] 0.1962391
```

- $\bar{R}^2$ del modelo b) (ols.2)

```
summary(ols.2)$adj.r.squared

## [1] 0.19606
```

Como el $\bar{R}^2$ del modelo c) es mayor que el $\bar{R}^2$ del modelo b),por lo tanto, en este caso prefiero la regresión del modelo c) ya que está más ajustado a los datos

3.2. Prefieres la regresión d) o la regresión b)

Si utilizamos el valor de $\bar{R}^2$ para comparar los modelos, tenemos que:

- $\bar{R}^2$ del modelo d) (ols.4)

```
summary(ols.4)$adj.r.squared

## [1] 0.1962726
```

- $\bar{R}^2$ del modelo b) (ols.2)

```
summary(ols.2)$adj.r.squared

## [1] 0.19606
```

Como el $\bar{R}^2$ del modelo d) es mayor que el $\bar{R}^2$ del modelo b),por lo tanto en este caso prefiero la regresión del modelo d) ya que está más ajustado a los datos (aunque age no sea estadísticamente significativo)

3.3. Prefieres la regresión d) o la regresión c)

- $\bar{R}^2$ del modelo d) (ols.4)

```
summary(ols.4)$adj.r.squared

## [1] 0.1962726
```

- $\bar{R}^2$ del modelo c) (ols.3)

```r
summary(ols.3)$adj.r.squared

## [1] 0.1962391
```

Como el $\bar{R}^2$ del modelo d) es mayor que el $\bar{R}^2$ del modelo c),por lo tanto en este caso prefiero la regresión del modelo d) ya que está más ajustado a los datos (aunque age no sea estadísticamente significativo)

4. El modelo de regresión básico es:

$$log(ahe_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + u_i$$

las variables *binarias* o *dummys* son:

$$female = \begin{cases} 1 & si\ es\ mujer \\ 0 & otro\ caso \end{cases}$$

$$bachelor = \begin{cases} 1 & si\ tiene\ bachelor \\ 0 & si\ tiene\ high\ school \end{cases}$$

4.1. Calcular las diferencias salariales de hombres con *bachelor*, mujeres con *bachelor* y hombres con *high school* respecto a la categoría básica *mujeres con high school*.

```r
ols.5 <- lm(I(log(ahe)) ~ age+I(age^2), data = cps12);ols.5

##
## Call:
## lm(formula = I(log(ahe)) ~ age + I(age^2), data = cps12)
##
## Coefficients:
## (Intercept)          age     I(age^2)
##    0.488087     0.135037    -0.001852
```

El modelo incluyendo las variables female y bachelor.

```r
ols.6 <- lm(I(log(ahe)) ~age+I(age^2)+female+bachelor + female:bachelor, data = cps12);ols.6

##
## Call:
## lm(formula = I(log(ahe)) ~ age + I(age^2) + female + bachelor +
##     female:bachelor, data = cps12)
##
## Coefficients:
##     (Intercept)              age         I(age^2)           female
##        0.803741         0.104322        -0.001332        -0.242373
##        bachelor  female:bachelor
##        0.400446         0.089857
```

Iteración entre variables cualitativos: female y bachelor:
female: (mujer=1,hombre=0) bachelor: (bachelor=1, High School=0)

female=1 y bachelor=1 Mujer con bachelor
female=1 y bachelor=0 Mujer con High School
female=0 y bachelor=1 Hombre con Bachelor
female=0 y bachelor=0 Hombre con High School

$$log(ahe) = \hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 age^2 + \hat{\beta}_3 female + \hat{\beta}_4 bachelor + \hat{\beta}_5 female : bachelor$$

DIFERENCIAS SALARIALES

4.1.1. Hombres con bachelor vs. Mujeres con High School:

$$E(log(ahe)|female = 0, bachelor = 1, \ldots) - E(log(ahe)|female = 1, bachelor = 0, \ldots)$$
$$= \hat{\beta}_4 - \hat{\beta}_3$$

```
ols.6$coef[5]-ols.6$coef[4]

##  bachelor
## 0.6428195
```

Los hombres con *bachelor* ganan un 64.28195 % más de lo que ganan las mujeres que cuentan con *High School*

4.1.2. Mujeres con bachelor vs. Mujeres con High School

$$E(log(ahe)|female = 1, bachelor = 1, \ldots) - E(log(ahe)|female = 1, bachelor = 0, \ldots)$$
$$= \hat{\beta}_4 + \hat{\beta}_5$$

```
ols.6$coef[5]+ols.6$coef[6]

##  bachelor
## 0.4903034
```

Las mujeres con *bachelor* ganan 49.03034 % más que las mujeres con *High School*

4.1.3. Hombres con High School vs. Mujeres con High School:

$$E(log(ahe)|female = 0, bachelor = 0, \ldots) - E(log(ahe)|female = 1, bachelor = 0, \ldots)$$
$$= -\hat{\beta}_3$$

```
-ols.6$coef[4]

##    female
## 0.2423732
```

Los hombres con *High School* ganan un 24.23732 % más que las mujeres con el mismo nivel de estudios.

4.2. ¿ Existe diferencial salarial entre hombres y mujeres con la misma edad?

En este caso, utilizamos el modelo de regresión :

$$log(ahe_i) = \hat{\beta}_0 + \hat{\beta}_1 age_i + \hat{\beta}_3 age_i^2 + \hat{\beta}_3 female$$

```
ols.7 <- lm(log(ahe) ~ age + I(age^2) +female, data = cps12)
```

$$\begin{aligned} &E(ahe|female = 0, age, \ldots) - E(log(ahe)|female = 1, age, \ldots) \\ &= -\hat{\beta}_3 \end{aligned}$$

```
-ols.7$coef[4]

##    female
## 0.1268194
```

Los hombres ganan 12.68194 % más que las mujeres de la misma edad.