Escuela Superior de Física y Matemáticas

ECONOMETRÍA

Alumno:
Roberto Carlos Santos Alonzo

Actividad 3

Noviembre 2020

Ejercicios

```r
library(lmtest)
library(sandwich)
```

1. Considerar el modelo de regresión lineal sin intercepto

$$Y_i = \beta X_i + u_i$$

donde u_i $iid(0, \sigma^2), cov(x_i, u_i) = 0, (x_i, y_i)$ son iid y no hay observaciones atípicas

1.1. Obtener el estimador β^* utilizando mínimos cuadrados ordinarios.

El estimador de mínimos cuadrados de β^* es:

$$Y_i = \beta X_i + u_i \longrightarrow u_i = (Y_i - \beta X_i)$$

La función del objetivo de mínimos cuadrados es

$$\sum_{i=1}^{n}(Y_i - \beta X_i)^2$$

El diferencial con respecto de β^*

$$\begin{aligned} \frac{\partial}{\partial \beta}(\sum_{i=1}^{n}(Y_i - \beta X_i)^2) &= -2\sum_{i=1}^{n} X_i(Y_i - \beta X_i) \\ &= -2\sum_{i=1}^{n} Y_i X_i + 2\sum_{i=1}^{n} \beta X_i^2 \end{aligned}$$

Igualando a cero y resuelviendo para el estiamdor de mínimos cuadrados β^*

$$-2\sum_{i=1}^{n} Y_i X_i + 2\sum_{i=1}^{n} \beta X_i^2 = 0$$

Así

$$\sum_{i=1}^{n} \beta X_i^2 = \sum_{i=1}^{n} Y_i X_i$$

Por lo que obtenemos β^*

$$\beta^* = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} \beta X_i^2}$$

1.2. Comprobar que el estimador β^* es lineal en Y_i es insesgado

Para demostrar que β^* es lineal en Y_i consideramos

$$\beta^* = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} \beta X_i^2}$$

Así

$$\beta^* = \sum_{i=1}^{n} a_i Y_i$$

Donde

$$a_i = \frac{X_i}{\sum\limits_{j=1}^{n} X_j^2}$$

Entonces, dado que a_i es independiente de X_i y no de Y_i, entonces β^* es una función lineal de Y_i

Para demostrar que β^* es condicional insesgado que es $E(\beta^*|X_1,\ldots,X_i) = \beta$, consideramos la función

$$E(\beta^*|X_1,\ldots,X_i) = \beta + \frac{\sum\limits_{i=1}^{n} X_i E(u_i|X_1,\ldots,X_n)}{\sum\limits_{j=1}^{n} X_j^2}$$

Entonces

$$E(u_i|X_1,\ldots,X_n) = 0$$

$$\therefore E(\beta^*|X_1,\ldots,X_i) = \beta$$

1.3. La constante del modelo fue eliminada por error cuando debería estar presente.¿Sigue siendo insesgado en el estimador $\hat{\beta}^*$ [Pista: introducir el modelo con intercepto $\hat{\beta}^*$]

Para mostrar que β^* es insesgado, en el estimador tenemos

$$\sum_{i=1}^{n}(X_i - \hat{X}) = 0$$

Las desviaciones sobre la suma media igual a cero es:

$$\sum_{i=1}^{n}(X_i - \hat{X})(Y_i - \hat{Y}) = Y_i(\sum_{i=1}^{n}(X_i - \hat{X})) - \hat{Y}(\sum_{i=1}^{n}(X_i - \hat{X}))$$

$$= \sum_{i=1}^{n} Y_i(X_i - \hat{X})$$

Sustituyendo el modelo de intercepto, tenemos

$$\beta^* = \frac{\sum\limits_{i=1}^{n} Y_i(X_i - \hat{X})}{\sum\limits_{i=1}^{n}(X_i - \hat{X})^2}$$

$$\beta^* = \hat{a}_i Y_i$$

Donde a_i

$$\hat{a_i^*} = \frac{\sum\limits_{i=1}^{n}(X_i - \hat{X})}{\sum\limits_{j=1}^{n}(X_j - \hat{X})^2}$$

Entonces $\hat{a}_i$ depende de $X_1,\ldots,X_n$ y no de $Y_1,\ldots,Y_n$ entonces β^* es un estimador lineal. Por la condición de Gauss-Markov, β^* sigue siendo insesgado.

1.4. Calcular la varianza del estimador β^*, comprobar que $var(\beta^*) \leq var(\hat{\beta})$. Pista: $\sum X_i^2 \geq \sum (X_i - \hat{X})^2$

$$var(\beta^*|X_1, \ldots, X_n) = \frac{\sum\limits_{i=1}^{n} X_i^2 var(u_i|X1, \ldots, X_n)}{\sum\limits_{j=1}^{n} X_j^2}$$

Por lo que tenemos

$$var(\beta^*|X_1, \ldots, X_n) = \frac{\mu_u^2}{\sum\limits_{j=1}^{n} X_j^2}$$

1.5. Que elegir, un estimador sesgado pero con una varianza mínima o un insegado con una varianza mayor.

Ya que la varianza es $\frac{\mu_u^2}{\sum\limits_{j=1}^{n} X_j^2}$ va hacer un estimador sesgado con una varianza mínima.

2. Demostrar:

2.1. En el modelo de regresión lineal simple si $\hat{\beta} = 0$, entonces $R^2 = 0$

Solución:

Tenemos que

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \ldots \alpha)$$
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \ldots \gamma)$$

si $\hat{\beta}_1 = 0$, sustituyendo en $\alpha)$ y γ nos queda

$$\bar{Y} = \hat{\beta}_0$$
$$\hat{Y}_i = \hat{\beta}_0$$

Con esto nos queda que

$$\bar{Y} = \hat{Y}_i \longrightarrow \bar{Y} - \hat{Y}_i = 0 \ldots \lambda$$

Ahora, tenemos que la **suma explicada(SE)** y la **suma tota(ST)**:

$$SE = \sum_{i=1}^{n} (\bar{Y} - \hat{Y}_i)^2$$

$$ST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Calculando el rato entre la suma explicada y la suma total, tenemos que:

$$R^2 = \frac{SE}{ST} = \frac{\sum\limits_{i=1}^{n} (\bar{Y} - \hat{Y}_i)^2}{\sum\limits_{i=1}^{n} (Y_i - \bar{Y})^2}$$

sustituyendo $\ldots \lambda)$ en R^2 tenemos que

$$R^2 = \frac{\sum\limits_{i=1}^{n} (\bar{Y} - \hat{Y}_i)^2}{\sum\limits_{i=1}^{n} (Y_i - \bar{Y})^2} = \frac{\sum\limits_{i=1}^{n} (0)^2}{\sum\limits_{i=1}^{n} (Y_i - \bar{Y})^2}$$

$$\therefore R^2 = 0$$

2.2. Demostrar que el R^2 de la regresión de y contra x es el valor cuadrado de la correlación muestral ebtre y y x. Es decir, r_{XY}^2

Solución:

Como es y contra x, tenemos que

$$X_i = \bar{Y}_i + \bar{u}_i \longrightarrow \bar{u}_i = X_i - \bar{Y}_i$$

$$SR = \sum_{i=1}^{n} (\hat{u}_i)^2$$

$$SR = \sum_{i=1}^{n} (\hat{X}_i - \bar{X})^2$$

$$X_i = \hat{\beta}_0 + \hat{\beta}_1 Y_i \ldots \gamma)$$

$$\bar{X} = \hat{\beta}_0 + \hat{\beta}_1 \bar{Y} \longrightarrow \hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{Y} \ldots \lambda)$$

Sustituyendo $\lambda)$ en $\gamma)$

$$\hat{X}_i = \bar{X} - \hat{\beta}_1 \bar{Y} + \hat{\beta}_1 Y_i \ldots \phi$$

Sustityuendo ϕ en SR tenemos que:

$$SR = \sum_{i=1}^{n} (\bar{X} - \hat{\beta}_1 \bar{Y} + \hat{\beta}_1 Y_i - \bar{X})^2$$

$$SR = \sum_{i=1}^{n} (\hat{\beta}_1 (Y_i - \bar{Y}))^2$$

$$SR = \hat{\beta}_1{}^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \ldots \omega)$$

pero, tenemos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \ldots \varpi)$$

sustituyendo $\varpi)$ en $\omega)$ nos queda:

$$SR = (\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$SR = \frac{(\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X}))^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

Ahora, tenemos que:

$$ST = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Recordando que $R^2 = \frac{SR}{ST}$, tenemos que:

$$R^2 = (\frac{\frac{(\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X}))^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{(\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X}))^2}{(\sum_{i=1}^{n}(Y_i-\bar{Y})^2)(\sum_{i=1}^{n}(X_i-\bar{X})^2)}$$

$$R^2 = [\frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sqrt{\frac{1}{n-1}(\sum_{i=1}^{n}(Y_i-\bar{Y}}\sqrt{\frac{1}{n-1}(\sum_{i=1}^{n}(X_i-\bar{X})}}]^2 = (\frac{s_xy}{S_x s_y})^2$$

$$\therefore R^2 = r_{XY}^2$$

2.3. $var(y_i) = \beta^2 var(x_i) + var(u_i)$

Solución:
Considerando el modelo de regresión lineal sin intercepto

$$y_i = \beta x_i + u_i$$

Por lo que, aplicando la varianza en ambos lados de la igualdad, obtenemos

$$var(y_i) = var(\beta x_i) + var(u_i)$$

pero como β es constante y dado que todos los valores de la variable se multiplican por un número, la varianza queda multiplicada por el cuadrado de dicho número, por esta definición nos queda

$$var(y_i) = \beta^2 var(x_i) + var(u_i)$$

donde u_i es el termino de error

```
load("C:/Users/81799/Downloads/TeachingRatings.RData")
```

3. El archivo TeachingRatings.RData contiene datos sobre las evaluaciones de la asignatura, características de la asignatura y del profesor para 463 cursos de la Universidad de Texas en Austin. Se incluye un índice de las características de belleza del profesor, acorde a la clasificación de un jurado de seis jueces, se utiliza para comprobar si hay relación entre la evaluación del curso y la belleza del profesor

$$course\_\ eval_i = \beta_0 + \beta_1 beauty_i + u_i$$

3.1. Construir un diagrama de dispersión para las evaluaciones medias del curso courseeval, sobre la belleza del profesor beauty. ¿Existe relación entre las variables?

El Gráfico de dispersión nos queda de la siguiente manera:

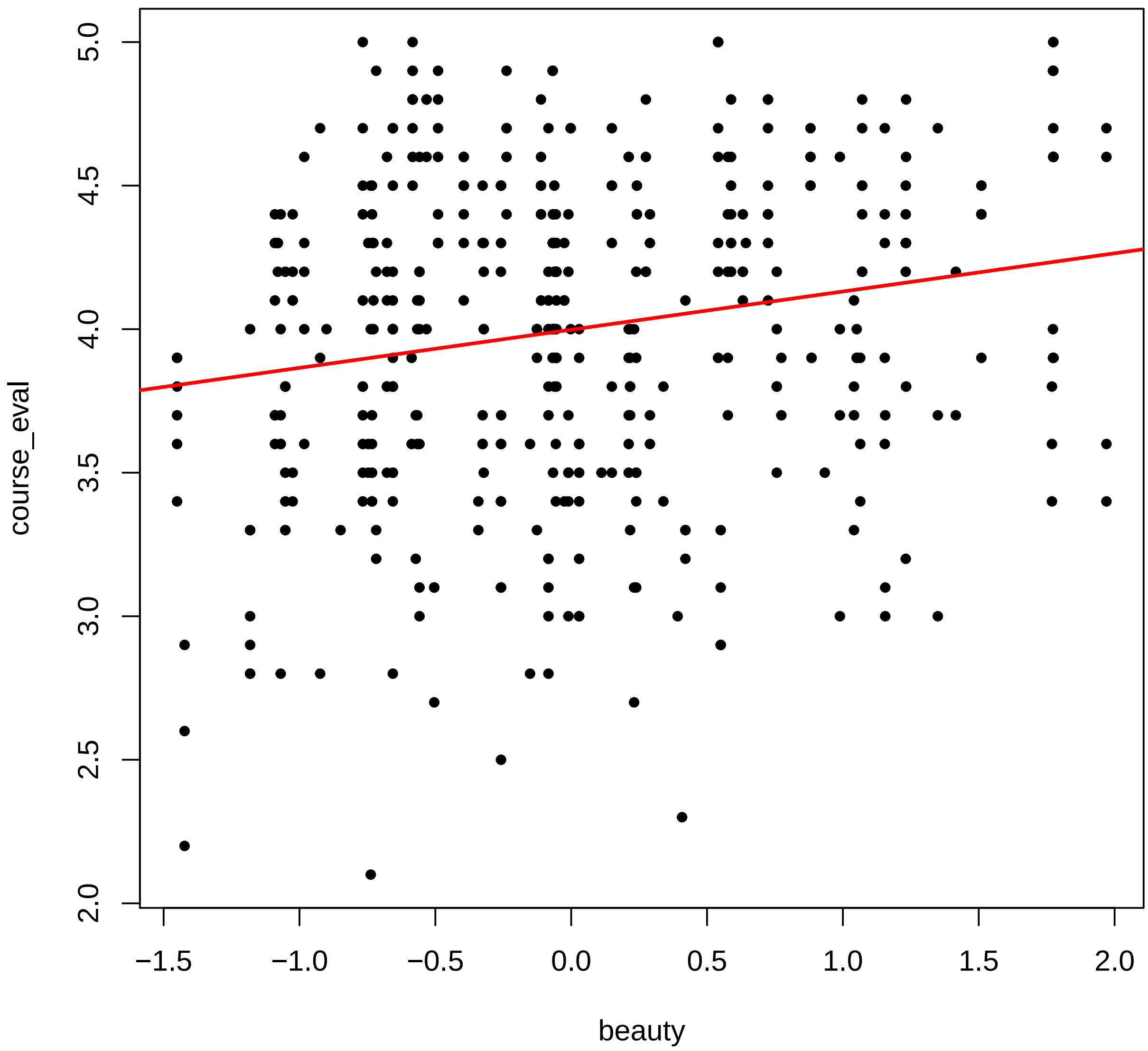
```
b <- sum((TeachingRatings$beauty-mean(TeachingRatings$beauty))*(TeachingRatings$course_eval-mean
```

```
## [1] 0.1330014
```

```
a <- mean(TeachingRatings$course_eval)-b*mean(TeachingRatings$beauty); a
```

```
## [1] 3.998272
```

```
plot(TeachingRatings$beauty, TeachingRatings$course_eval , pch=20, xlim = c(min(TeachingRatings$
abline(a,b, col = "red", lwd = 2 )
```



3.2. Es estadísticamente significativo el coeficiente de la pendiente de la regresión. Es deecir, se puede rechazar la hipótesis nula $H_0 : \beta = 0$ frente a una alternativa biateral al nivel de significancia del 10, 5 y 1 por ciento.¿ Cuál es el p-value asociado al estadístico t del coeficiente?

```
ols <- lm(course_eval~beauty, data = TeachingRatings)
```

```
e.2 <- residuals(ols)^2
n <-length(e.2)
desvio.x <-(TeachingRatings$beauty-mean(TeachingRatings$beauty))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2));var.b

## [1] 0.001044513

sd.b <- sqrt(var.b);sd.b

## [1] 0.03231893
```

Varianza del estimador:

```
s2.e <- sum(e.2)/(n-2) #sigma
b.var.homo <- s2.e/(sum(desvio.x))
b.sd.homo <- sqrt(b.var.homo)
```

Estadístico t:

```
t.homo <-ols$coef[2]/b.sd.homo;t.homo

##   beauty
## 4.133368
```

p-value

```
p.value.homo <- 2*pnorm(-abs (t.homo));p.value.homo

##       beauty
## 3.574859e-05
```

3.2.1 Realizando la prueba t, para determinar si es significativo el coeficiente a una bilateral al nivel de significancia del 10 %
Solución: en este caso $\alpha = 0.1 \longrightarrow \frac{\alpha}{2} = .05$ Buscando por tablas de la distribución normal estándar $Z(\frac{\alpha}{2} = .05) = 1.645$ con esto el **Intervalo de confianza nos da:**

```
cbind(inf=ols$coef[2]-1.645*b.sd.homo,sup=ols$coef[2]+1.645*b.sd.homo)

##                inf       sup
## beauty 0.08006946 0.1859334
```

En este caso, **No es estadísticamente** significativo, ya que el $p - evalue = 0.02408721047$) y este, NO pertenece al intervalo de confianza

3.2.2 Realizando la prueba t, para determinar si es significativo el coeficiente a una bilateral al nivel de significancia del 5 %
Solución: en este caso $\alpha = .05 \longrightarrow \frac{\alpha}{2} = .025$ Buscando por tablas de la distribución normal estándar $Z(\frac{\alpha}{2} = .025) = 1.96$ con esto el **Intervalo de confianza nos da:**

```
cbind(inf=ols$coef[2]-1.96*b.sd.homo,sup=ols$coef[2]+1.96*b.sd.homo)

##                inf       sup
## beauty 0.06993355 0.1960694
```

En este caso, **No es estadísticamente** significativo, ya que el $p - evalue = 0.02408721047)$ y este, NO pertenece al intervalo de confianza

3.2.3 Realizando la prueba t, para determinar si es significativo el coeficiente a una bilateral al nivel de significancia del 1 %
Solución: en este caso $\alpha = .01 \longrightarrow \frac{\alpha}{2} = .005$ Buscando por tablas de la distribución normal estándar $Z(\frac{\alpha}{2} = .005) = 2.575$ con esto el **Intervalo de confianza nos da:**

```
cbind(inf=ols$coef[2]-2.575*b.sd.homo,sup=ols$coef[2]+2.575*b.sd.homo)

##               inf       sup
## beauty 0.05014438 0.2158585
```

En este caso, **No es estadísticamente** significativo, ya que el $p - evalue = 0.02408721047)$ y este, NO pertenece al intervalo de confianza

3.3. Realizar una regresión de las evaluaciones medias del curso courseeval sobre la belleza del profesor(beauty). ¿Cuál es el término independiente estimaado?, ¿Cuál es la pendiente estimada?. Explicar porqué el término independiente estimado es igual a la media muestral de la variable courseeval(piesta:¿Cuál es la media muestral de la variable beauty)

El término independiente es:

```
a

## [1] 3.998272
```

La pendiente estimada es:

```
b

## [1] 0.1330014
```

Verificando por linea de comando, tenemos que

```
ols

##
## Call:
## lm(formula = course_eval ~ beauty, data = TeachingRatings)
##
## Coefficients:
## (Intercept)       beauty
##       3.998        0.133
```

La media muestral es:

```
mean(TeachingRatings$course_eval)

## [1] 3.998272
```

$$\therefore \mu_x = \beta_0$$

.

3.4. El profesor tiene valor de beauty menos una desviación de la media, la profesora una desviación más respecto a la media, predecir la evaluación del curso de cada profesor

3.4.1. Evaluación para el profesor

Para el **Profesor**, nos queda que:

```
course_eval <- TeachingRatings[TeachingRatings$female == 0, 6]
beauty <- TeachingRatings[TeachingRatings$female == 0, 5]
profesor <- data.frame(course_eval,beauty)

b <- sum((profesor$beauty-mean(profesor$beauty))*(profesor$course_eval-mean(profesor$course_eval

## [1] 0.2002743

a <- mean(profesor$course_eval)-b*mean(profesor$beauty); a

## [1] 4.085949
```

Verificando por linea de comando, tenemos que:

```
regprofe <- lm(course_eval ~ beauty , data = profesor); regprofe

##
## Call:
## lm(formula = course_eval ~ beauty, data = profesor)
##
## Coefficients:
## (Intercept)       beauty
##      4.0859       0.2003
```

```
e.2 <- residuals(regprofe)^2
n <-length(e.2)
desvio.x <-(profesor$beauty-mean(profesor$beauty))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2));var.b

## [1] 0.002249654

sd.b <- sqrt(var.b);sd.b

## [1] 0.04743052
```

Como el profesor tiene valor de beauty menos desviación de la media, tenemos que:

```
course_eval_profe <- b-sd.b ;course_eval_profe

## [1] 0.1528438
```

3.4.2. Evaluación para la profesora

Para la **profesora**, nos queda que:

```
course_eval <- TeachingRatings[TeachingRatings$female == 1, 6]
beauty <- TeachingRatings[TeachingRatings$female == 1, 5]
profesora <- data.frame(course_eval,beauty)
b <- sum((profesora$beauty-mean(profesora$beauty))*(profesora$course_eval-mean(profesora$course_
```

```
## [1] 0.08761645
```

```
a <- mean(profesora$course_eval)-b*mean(profesora$beauty); a
```

```
## [1] 3.890853
```

Verificando por linea de comando, tenemos que:

```
regprofa <- lm(course_eval ~ beauty , data = profesora ); regprofa
```

```
##
## Call:
## lm(formula = course_eval ~ beauty, data = profesora)
##
## Coefficients:
## (Intercept)       beauty
##     3.89085      0.08762
```

```
e.2 <- residuals(regprofa)^2
n <-length(e.2)
desvio.x <-(profesora$beauty-mean(profesora$beauty))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2))
sd.b <- sqrt(var.b);sd.b
```

```
## [1] 0.04049437
```

Como la profesora tiene valor de beauty más desviación de la media, tenemos que:

```
course_eval_profa <- b+sd.b ;course_eval_profa
```

```
## [1] 0.1281108
```

```
load("C:/Users/81799/Downloads/cps08.RData")
```

4. El archivo *cps08.Data*, contiene la distribución de ingresos salariales en Estados Unidos en el 2008, contiene dotos relativos a los ingresos por hora de graduados universitarios de 25 a 34 años de teimpo completo, trabajadores de tiempo completo de 25 a 34 años, en el año 2008,titulados en la escuela secundaria y licenciados (ingenieros) como grado mas alto de la educación alcanzado. El objetivo es ingrsar la relación entre la edad de trabajador y sus ingresos salariales.

4.1. Realizar una regresión de los ingresos medios por hora, sobre la edad, ¿Cuánto aumentarán los ingresos al aumentar la edad de los trabajadores en un año?

De forma manual, tenemos que:

```r
b <- sum((cps08$age-mean(cps08$age))*(cps08$ahe-mean(cps08$ahe)))/sum((cps08$age-mean(cps08$age)

## [1] 0.6049863
```

```r
a <- mean(cps08$ahe)-b*mean(cps08$age); a

## [1] 1.082275
```

Comprando por linea de comando, nos da:

```r
reg<- lm(ahe ~ age, data = cps08); reg

##
## Call:
## lm(formula = ahe ~ age, data = cps08)
##
## Coefficients:
## (Intercept)          age
##       1.082        0.605
```

Calculamos la regresión, si los trabajadores aumenta la edad en un año

De forma maual, tenemos que:

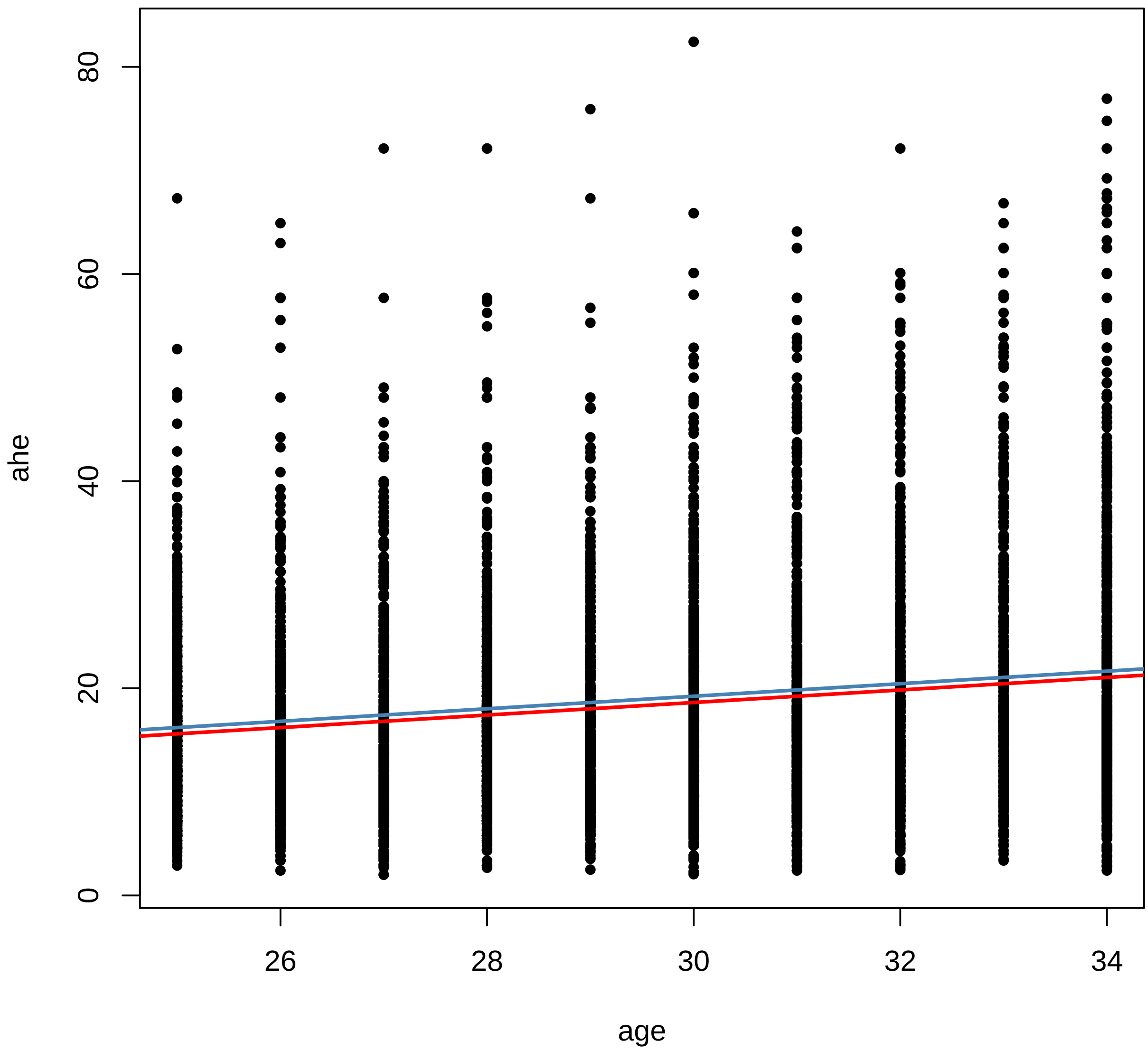
```r
be <- sum(((cps08$age+1)-mean(cps08$age+1))*(cps08$ahe-mean(cps08$ahe)))/sum((cps08$age+1-mean(c

## [1] 0.6049863
```

```r
al <- mean(cps08$ahe)-be*mean(cps08$age+1); al

## [1] 0.4772889
```

Comprobando por linea de comando, nos da:

```r
ahe <- cps08$ahe
age <- cps08$age+1
cps081 <- data.frame(age,ahe)
reg1 <- lm(ahe ~ age, data = cps081 );reg1

##
## Call:
## lm(formula = ahe ~ age, data = cps081)
##
## Coefficients:
## (Intercept)          age
##      0.4773       0.6050
```

```
plot( cps08$age,cps08$ahe, pch=20, ylim = c(min(cps08$ahe), max(cps08$ahe)), xlim = c(min(cps08$
      ylab = "ahe")
abline(a,b, col = "steelblue", lwd = 2 )
abline(al,be, col = "red", lwd = 2 )
```



$$\hat{y_i} = 1.082775 + 0.6049863\hat{x_i} \rightarrow (azul)\text{Regresión sin aumentar edad}$$

$$\hat{y_i} = 0.4772889 + 0.6049863\hat{x_i} \rightarrow (rojo)\text{Regresión con aumentar edad}$$

Nos damos cuenta que la pendiente Regresión sin aumentar edad es igual que la pendiente Regresión sin aumentar edad, como las pendientes son iguales podemos tomar cualquier edad y y sustituilar en las regresiones, para saber cuanto es lo que difiere el salario.

```
salario <- abs((a+b*27)-(al+be*27)); salario

## [1] 0.6049863
```

con esto, vemos que hay una **disminución** en el salario de 0.604963 para la regresión con aumentar respecto a la regresión sin aumentar

4.2. Construir un intervalo de confianza al 95 por ciento para el coeficiente de la pendiente

```
confint(reg1, level = 0.95)

##                  2.5 %    97.5 %
## (Intercept) -1.9219485 2.8765263
## age          0.5268613 0.6831113
```

4.3. Repetir la regresión utilizando solo los datos de los graduados de escuela secundaria

```
age <- cps08[cps08$bachelor == 0, 5 ]+1
ahe <- cps08[cps08$bachelor == 0, 1 ]
graduado <- data.frame(age,ahe)
```

De forma manual, tenemos que:

```
bee <- sum(((graduado$age)-mean(graduado$age))*(graduado$ahe-mean(graduado$ahe)))/sum((graduado$

## [1] 0.2978627
```

```
all <- mean(graduado$ahe)-bee*mean(graduado$age); all

## [1] 6.224078
```

Comprobando por linea de comando, nos da:

```
reg12 <- lm(ahe ~ age, data = graduado );reg12

##
## Call:
## lm(formula = ahe ~ age, data = graduado)
##
## Coefficients:
## (Intercept)          age
##      6.2241       0.2979
```

`La regresión nos queda:`

$$\hat{y_i} = 6.224078 + 0.2978627\hat{x_i}$$

4.4. Repetir la regresión utilizando solo los datos de los graduados universitarios

```
age <- cps08[cps08$bachelor == 1, 5 ]+1
ahe <- cps08[cps08$bachelor == 1, 1 ]
graduado2 <- data.frame(age,ahe)
```

De forma manual, tenemos que:

```r
beee <- sum(((graduado2$age)-mean(graduado2$age))*(graduado2$ahe-mean(graduado2$ahe)))/sum((grad

## [1] 0.924596
```

```r
alll <- mean(graduado2$ahe)-beee*mean(graduado2$age); alll

## [1] -5.363759
```

Comprobando por linea de comando, nos da:

```r
reg12 <- lm(ahe ~ age, data = graduado2 );reg12

##
## Call:
## lm(formula = ahe ~ age, data = graduado2)
##
## Coefficients:
## (Intercept)          age
##     -5.3638       0.9246
```

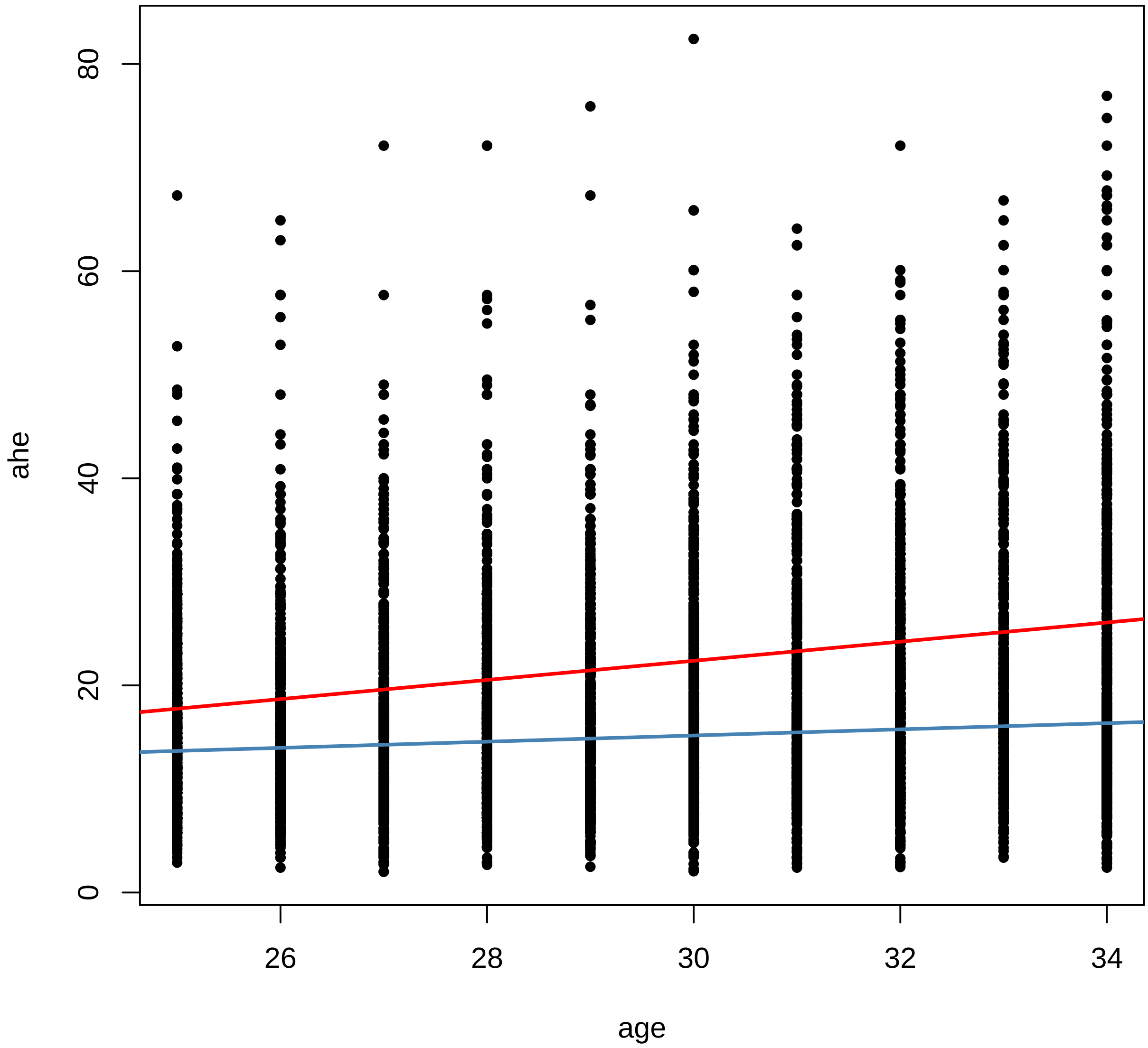
La regresión nos queda:

$$\hat{y}_i = -5.3638 + 0.9246\hat{x}_i$$

4.5. Es distinto el efecto de la variable *age* sobre los ingresos salariales para los graduados de secuandaria que para los graduados universitarios.Explicar.

Graficando el diagrama de dispersión con la recta de regresión de lo graduados universitarios y los graduados de la escuela secundaria:

```r
plot( cps08$age,cps08$ahe, pch=20, ylim = c(min(cps08$ahe), max(cps08$ahe)), xlim = c(min(cps08$
     ylab = "ahe")
abline(all,bee, col = "steelblue", lwd = 2 )
abline(alll,beee, col = "red", lwd = 2 )
```



Tenemos que nuestras regresiones son:

$$\hat{y}_i = -5.3638 + 0.9246\hat{x}_i \rightarrow (Rojo)\text{Regresión de graduados universitarios}$$

$$\hat{y}_i = 6.224078 + 0.2978627\hat{x}_i \rightarrow (azul)\text{Regresión de graduados de escuela secundaria}$$

Con esto llegamos a la conclusión de que, **si es distinto el efecto de la variable *age* sobre los ingresos salarias**, ya que la regresión de los graduados universitarios es mayor que la de la regresión de los graduados de la escuela secundaria(esto se nota ya que la pendiente de los graduados universitarios es mayor que los de secundria), con esto los universitarios tienen un crecimiento salarial más grande que los de secundaria)

5. El archivo Growth.RData contiene datos sobre las tasas medias de crecimiento desde 1960 a 1995 para 65 países, junto con variables que potencialmente están relacionadas con el crecimiento. En este ejercicio se investigará la relación entre crecimiento y el comercio

5.1. Utilizar todas las observaciones

5.1.1. Construir un diagrama de dispersión de la cuota media de participación del comercio (**tradeshare**) sobre la tasa media de crecimiento anual (**growth**). Existe relación entre las variables

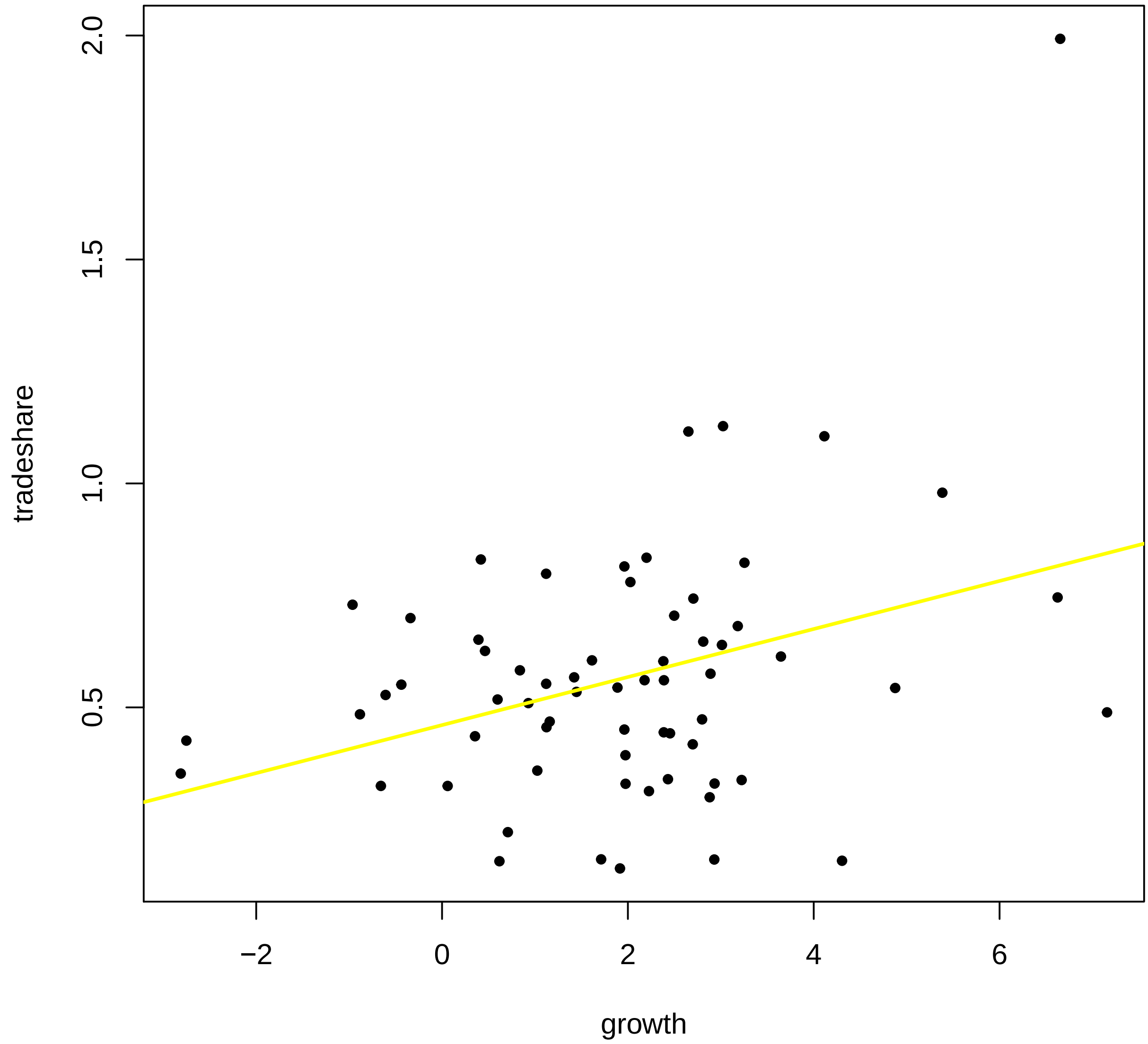
Utilizando la base de datos Growth.Rdata

```
load("C:/Users/81799/Downloads/Growth.RData")
```

El Gráfico de dispersión nos queda de la siguiente manera:

```
b <- sum((Growth$growth-mean(Growth$growth))*(Growth$tradeshare-mean(Growth$tradeshare)))/sum(((

## [1] 0.053624

a <- mean(Growth$tradeshare)-b*mean(Growth$growth ); a

## [1] 0.4605269
```

```
plot( Growth$growth ,Growth$tradeshare , pch=20, xlim = c(min(Growth$growth), max(Growth$growth)
abline(a,b, col = "yellow", lwd = 2 )
```



5.1.2. Realizar una regresión de **tradeshare** sobre **growth**

```
ols <- lm(tradeshare~growth, data = Growth);ols

##
## Call:
## lm(formula = tradeshare ~ growth, data = Growth)
##
## Coefficients:
## (Intercept)       growth
##     0.46053      0.05362
```

5.1.3. ¿Cuál es la pendiente estimada y que significa?

La pendiente estimada ($\hat{\beta}$) es:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (1)$$

```
b <- b <- sum((Growth$growth-mean(Growth$growth))*(Growth$tradeshare-mean(Growth$tradeshare)))/s

## [1] 0.053624
```

Recordemos que la **pendiente** de la recta relaciona X con Y es elefecto de la variación en una unidad de X sobre Y, con esto, la variación de tradeshare sobre growth nos queda

```
b

## [1] 0.053624
```

5.1.4. Calcular la R^2 y explicar

$$R^2 = 1 - \frac{SRC}{STC} \tag{2}$$

```
growth.hat <- a+b*Growth$growth
e <- Growth$growth - growth.hat
STC <-  sum((Growth$growth - mean(Growth$growth))^2)
SEC <-  sum((growth.hat - mean(Growth$growth))^2)
SRC <- sum(e^2)
R.2 <- 1-(SRC/STC); R.2

## [1] -0.431486
```

5.1.5. Obtener el error estándar de la regresión

```
ols <- lm(tradeshare~growth, data = Growth)
```

Calcular el error estándar del estimador:

$$\hat{\sigma}^2_{\hat{\beta}} = \frac{1}{n} \frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u_i}^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} \tag{3}$$

```
e.2 <- residuals(ols)^2
n <- length(e.2)
desvio.x <- (Growth$growth-mean(Growth$growth))^2
var.b <- (n/(n-2))*sum(desvio.x*e.2)/sum(desvio.x)^2
sd.b <- sqrt(var.b);sd.b

## [1] 0.02833622
```

5.1.6. Realizar la prueba t, para determinar si es significativo el coeficiente al 95 de confianza, utilizar forma manual y linea de comando

5.1.6.1 Errores homocedásticos **Calcular el error estándar del estimador:**

```
e.2 <- residuals(ols)^2
n <-length(e.2)
desvio.x <-(Growth$growth-mean(Growth$growth))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2))
sd.b <- sqrt(var.b);sd.b

## [1] 0.02833622
```

Varianza del estimador:

```
s2.e <- sum(e.2)/(n-2) #sigma
b.var.homo <- s2.e/(sum(desvio.x))
b.sd.homo <- sqrt(b.var.homo)
```

Estadístico t:

```
t.homo <-ols$coef[2]/b.sd.homo;t.homo

##   growth
## 2.981873
```

p-value

```
p.value.homo <- 2*pnorm(-abs (t.homo));p.value.homo

##      growth
## 0.002864913
```

Intervalo de confianza

```
cbind(inf=ols$coef[2]-1.96*b.sd.homo,sup=ols$coef[2]+1.96*b.sd.homo)

##               inf        sup
## growth 0.01837667 0.08887133
```

Linea de comando de Homocedásticos

```
coeftest(ols,vcov=vcovHC, type="const")

##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.460527   0.048648  9.4666 1.017e-13 ***
## growth      0.053624   0.017983  2.9819   0.00407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

coefci(ols,vcov=vcovHC, type="const")

##                  2.5 %     97.5 %
## (Intercept) 0.36331236 0.55774134
## growth      0.01768718 0.08956082
```

5.1.6.2 Errores heterocedásticos **Calcular el error estándar del estimador:**

```
e.2 <- residuals(ols)^2
n <-length(e.2)
desvio.x <-(Growth$growth-mean(Growth$growth))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2))
sd.b <- sqrt(var.b);sd.b

## [1] 0.02833622
```

Calcular el estadístico t:

```
t <- ols$coef[2]/sd.b;t

##   growth
## 1.892419
```

Calcular p-value:

```
p.value <- 2*pnorm(-abs (t));p.value

##    growth
## 0.0584352
```

Intervalo de confianza de $\hat{\beta}$.

```
cbind(inf=ols$coef[2]-1.96*sd.b,sup=ols$coef[2]+1.96*sd.b)

##                  inf      sup
## growth -0.001914995 0.109163
```

Linea de comando de Robustos a la heterocedasticidad

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

##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.460527   0.043897 10.4911 1.866e-15 ***
## growth      0.053624   0.028336  1.8924   0.06303 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

coefci(ols,vcov=vcovHC, type="HC1")

##                    2.5 %    97.5 %
## (Intercept)  0.372805968 0.5482477
## growth      -0.003001421 0.1102494
```

```
ols <- lm(growth~tradeshare  , data = Growth)
```

5.1.7. Construir intervalo de confianza del 95 por ciento de la variable tradeshare, utilizar formula y linea de comando

5.1.7.1 Errores homocedásticos **Calcular el error estándar del estimador:**

```
e.2 <- residuals(ols)^2
n <-length(e.2)
desvio.x <-(Growth$tradeshare-mean(Growth$tradeshare))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2))
sd.b <- sqrt(var.b);sd.b

## [1] 0.6632868
```

Varianza del estimador:

```
s2.e <- sum(e.2)/(n-2) #sigma
b.var.homo <- s2.e/(sum(desvio.x))
b.sd.homo <- sqrt(b.var.homo)
```

Estadístico t:

```
t.homo <-ols$coef[2]/b.sd.homo;t.homo

## tradeshare
##   2.981873
```

p-value

```
p.value.homo <- 2*pnorm(-abs (t.homo));p.value.homo

##  tradeshare
## 0.002864913
```

Intervalo de confianza

```
cbind(inf=ols$coef[2]-1.96*b.sd.homo,sup=ols$coef[2]+1.96*b.sd.homo)

##                  inf      sup
## tradeshare 0.7904031 3.822464
```

Linea de comando de Homocedásticos

```
coeftest(ols,vcov=vcovHC, type="const")

##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.64027    0.48998  1.3067  0.19606
## tradeshare   2.30643    0.77349  2.9819  0.00407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coefci(ols,vcov=vcovHC, type="const")

##                  2.5 %   97.5 %
## (Intercept) -0.3388749 1.619405
## tradeshare   0.7607473 3.852120
```

5.1.7.2 Errores heterocedásticos **Calcular el error estándar del estimador:**

```r
e.2 <- residuals(ols)^2
n <-length(e.2)
desvio.x <-(Growth$tradeshare-mean(Growth$tradeshare))^2
var.b <-(n/(n-2))*(sum(desvio.x*e.2)/(sum(desvio.x)^2))
sd.b <- sqrt(var.b);sd.b

## [1] 0.6632868
```

Calcular el estadístico t:

```r
t <- ols$coef[2]/sd.b;t

## tradeshare
##    3.47728
```

Calcular p-value:

```r
p.value <- 2*pnorm(-abs (t));p.value

##   tradeshare
## 0.0005065294
```

Intervalo de confianza de $\hat{\beta}$.

```r
cbind(inf=ols$coef[2]-1.96*sd.b,sup=ols$coef[2]+1.96*sd.b)

##                 inf      sup
## tradeshare 1.006392 3.606476
```

Linea de comando de Robustos a la heterocedasticidad

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

##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  0.64027    0.45915  1.3945 0.1680736
## tradeshare   2.30643    0.66329  3.4773 0.0009235 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

coefci(ols,vcov=vcovHC, type="HC1")

##                  2.5 %   97.5 %
## (Intercept) -0.2772641 1.557795
## tradeshare   0.9809608 3.631907
```

5.1.8. Utilizar la regresión para predecir la tasa de crecimiento de un país con una participación del comercio con un .5 y con una participación del 1

La tasa de crecimiento de un país con una participación del comercio igual a .5 es:

```
y <- a+(b*.5);y

## [1] 0.4873389
```

La tasa de crecimiento de un país con una participación del comercio igual a 1 es:

```
y <- a+(b*1);y

## [1] 0.5141509
```

5.2. Estimar de nuevo la regresión excluyendo los datos de Malta:comparar los coeficientes de ambas regresiones, que puedes explicar al recpecto

```
tradeshare <- Growth[Growth$country_name != "Malta" , 5 ]
growth <- Growth[Growth$country_name != "Malta" , 2 ]
sinmalta <- data.frame(tradeshare,growth)
```

```
b <- sum((sinmalta$growth-mean(sinmalta$growth))*(sinmalta$tradeshare-mean(sinmalta$tradeshare))

## [1] 0.02656789

a <- mean(sinmalta$tradeshare)-b*mean(sinmalta$growth); a

## [1] 0.4927333
```

Verificando con la linea de comando, tenemos que:

```
ols <- lm(tradeshare~growth, data = sinmalta);ols

##
## Call:
## lm(formula = tradeshare ~ growth, data = sinmalta)
##
## Coefficients:
## (Intercept)       growth
##     0.49273      0.02657
```

La tasa de crecimiento de un país(excluyendo Malta) con una participación del comercio igual a .5 es:

```
y <- a+(b*.5);y

## [1] 0.5060173
```

La tasa de crecimiento de un país(excluyendo Malta) con una participación del comercio igual a 1 es:

```
y <- a+(b*1);y

## [1] 0.5193012
```

En este caso, nos damos cuenta que cuando exluimos Malta, baja la tasa de crecimiento de un país.

5.3. ¿Debería estar incluida o excluida del análisis?

Tenemos que el error estándar de la regresión con malta, es de :

```
ols <- lm(tradeshare~growth, data = Growth)
e.2 <- residuals(ols)^2
n <- length(e.2)
desvio.x <- (Growth$growth-mean(Growth$growth))^2
var.b <- (n/(n-2))*sum(desvio.x*e.2)/sum(desvio.x)^2
sd.b <- sqrt(var.b);sd.b

## [1] 0.02833622

sd.b

## [1] 0.02833622
```

Y el error estándar de la regresión sin malta, es de:

```
ols <- lm(tradeshare~growth, data = sinmalta)
e.2 <- residuals(ols)^2
n <- length(e.2)
desvio.x <- (sinmalta$growth-mean(sinmalta$growth))^2
var.b <- (n/(n-2))*sum(desvio.x*e.2)/sum(desvio.x)^2
sd.b <- sqrt(var.b);sd.b

## [1] 0.01408595
```

Como el error estándar de la regresión es más pequeño excluyendo a Malta, entonces concluimos que **Malta debería de ser excluida**