

CPS08

Joel Alejandro Zavala Prieto

Contents

Informacion de contacto	2
Descripción	3
Modelo	3
Visualización de los datos	4
Modelo ajustado	4
Predicción	6
Resumen general	7
Distribucion de los residuales	7
Medidas estadisticas extras	8

Informacion de contacto

Mail: alejandro.zavala1001@gmail.com

Facebook: <https://www.facebook.com/AlejandroZavala1001>

Git: <https://github.com/AlejandroZavala98>

Descripción

La base de datos cps08.RData contiene la distribución de ingresos salariales en Estados Unidos en 2008, contiene datos relativos a los ingresos por hora de los graduados universitarios de 25 a 34 años. El objetivo es investigar la relación entre la edad de un trabajador y sus ingresos salariales

Modelo

Se propone el modelo

$$\begin{aligned} ahe_i &= \beta_0 + \beta_1 age_i + u_i \\ i &= 1, 2, \dots, n \end{aligned}$$

Una pequeña descripción de las variables de la base de datos se da a continuación

Variables	Definition
ahe	Average earning hours
female	1 if female 0 if male
yrseduc	years of education
BACHELOR	1 if worker has a bachelor's degree 0 if worker has a high school degree

El nombre de columnas de la base de datos se muestra a continuación

```
## [1] "ahe"      "year"     "bachelor" "female"   "age"
```

Mostrando las primeras observaciones de la tabla para las variables requeridas

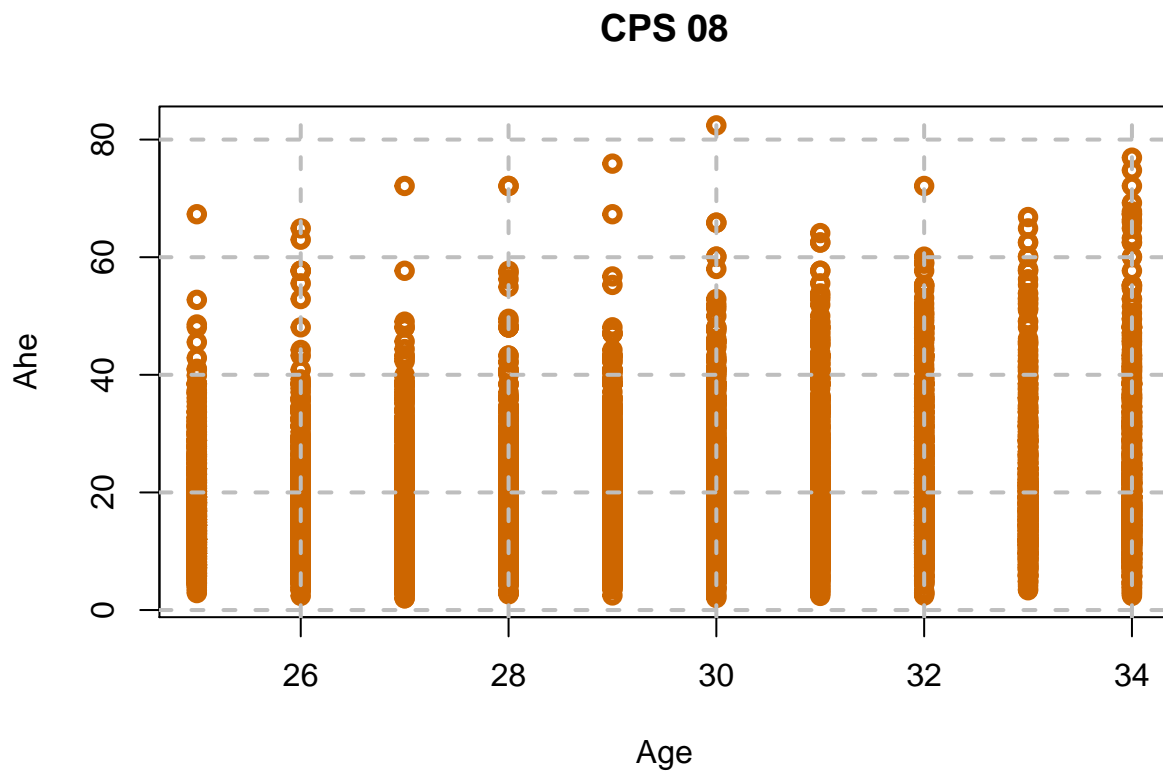
ahe	year	bachelor	female	age
38.461540	2008	1	0	33
12.500000	2008	1	0	31
9.857142	2008	0	0	30
8.241758	2008	0	0	30
17.788462	2008	0	0	31
10.096154	2008	0	1	29
17.788462	2008	1	0	26
30.384615	2008	1	0	28
23.668638	2008	1	0	30
12.019231	2008	1	1	25

El modelo ajustado es

$$\begin{aligned} a\hat{h}e_i &= \hat{\beta}_0 + \hat{\beta}_1 age_i \\ i &= 1, 2, \dots, n \end{aligned}$$

Visualización de los datos

Una visualización previa de los datos



La regresión del modelo es

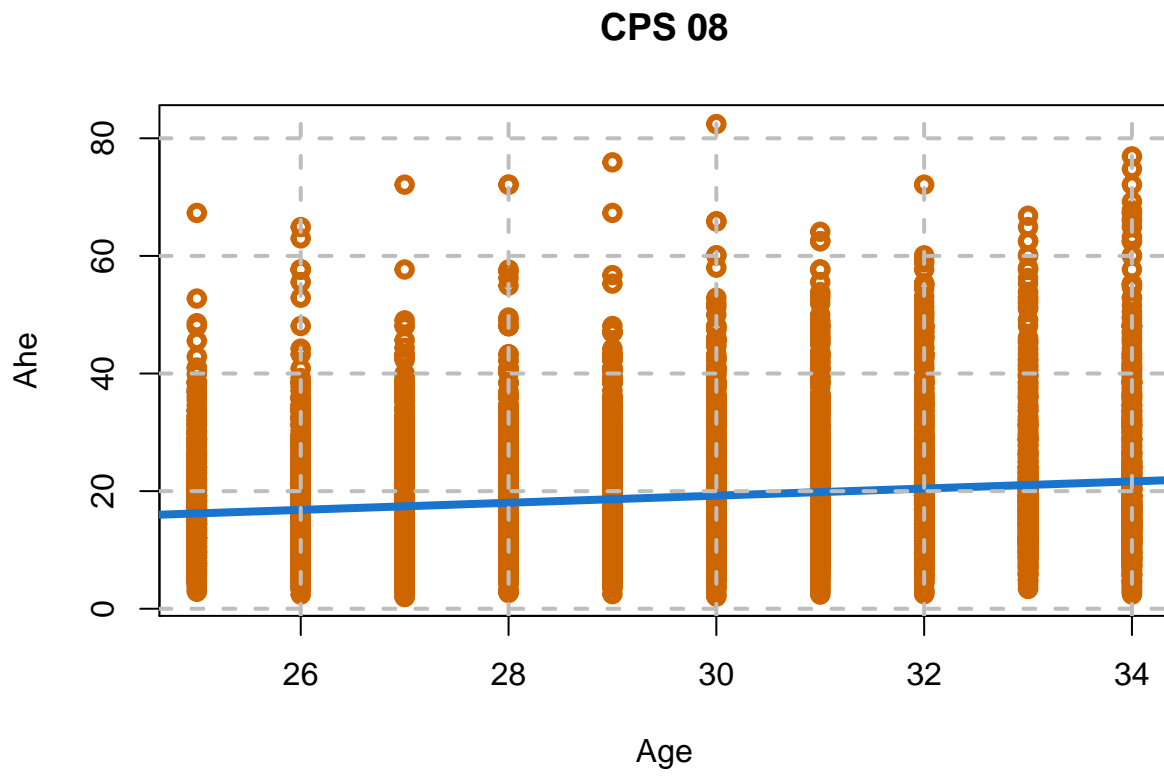
```
##  
## Call:  
## lm(formula = ahe ~ age, data = cps08)  
##  
## Coefficients:  
## (Intercept)      age  
##      1.082      0.605
```

Modelo ajustado

El modelo ajustado es

$$\hat{ahe}_i = 1.082 + 0.605age_i$$
$$i = 1, 2, \dots, n$$

De tal forma



Predicción

supongamos que Alejandro es un trabajador de 22 años. Vamos a predecir los ingresos de Alejandro

```
f_predict <- function(x,a,b)
{
  a+(b*x)
}

beta_0 <- ols_cps$coefficients[1]
beta_1 <- ols_cps$coefficients[2]

f_predict(22,beta_0,beta_1)
```

```
## (Intercept)
##      14.39197
```

Por linea de comando

```
nuevas.edades <- data.frame(age = c(22))
predict(ols_cps, nuevas.edades)
```

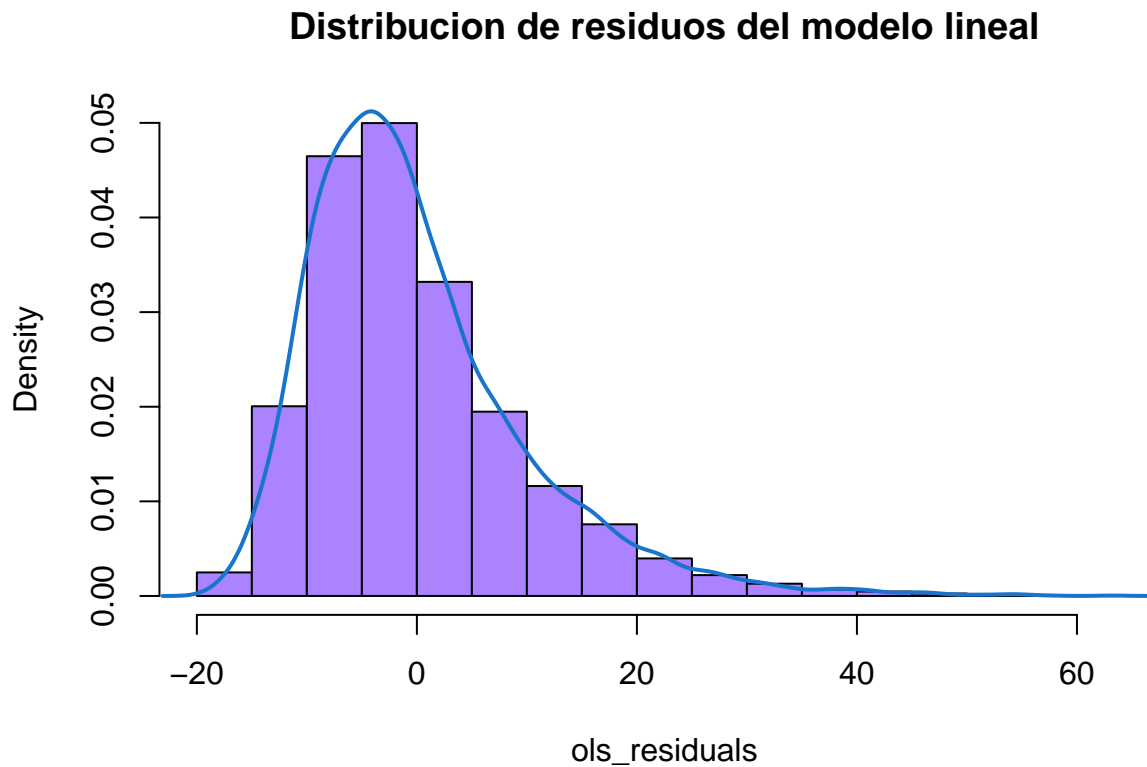
```
##           1
## 14.39197
```

Resumen general

Ahe	Ahe_Ajustados	Residuales
38.461540	21.04682	17.4147173
12.500000	19.83685	-7.3368503
9.857142	19.23186	-9.3747216
8.241758	19.23186	-10.9901057
17.788462	19.83685	-2.0483886
10.096154	18.62688	-8.5307235
17.788462	16.81192	0.9765428
30.384615	18.02189	12.3627235
23.668638	19.23186	4.4367742
12.019231	16.20693	-4.1877017
15.865385	16.81192	-0.9465338
11.846154	16.20693	-4.3607783
28.846153	19.23186	9.6142892
9.440909	19.83685	-10.3959409
20.192308	20.44184	-0.2495282

Distribucion de los residuales

Observando la distribucion de los errores para estos datos



Medidas estadísticas extras

Cantidad de datos totales

```
n <- length (cps08$ahe); n
```

```
## [1] 7711
```

Suma total de cuadrados

```
# alpha <- 1.082; beta <- 0.605  
# ahe.hat <- alpha + beta * cps08$age estos son los ajustados manualmente  
stc <- sum((cps08$ahe - mean(cps08$ahe))^2); stc
```

```
## [1] 792651.5
```

Suma explicada de cuadrados

```
sec <- sum((ols_ajustados - mean(cps08$ahe))^2); sec
```

```
## [1] 23005.74
```

Suma de residuales al cuadrado

```
errores <- cps08$ahe - ols_ajustados  
src <- sum(errores^2); src
```

```
## [1] 769645.7
```

Error estándar de la regresión

```
esr <- sqrt (src/(n-2)) ; esr
```

```
## [1] 9.991862
```

Coefficiente de determinación R^2

```
R_2 <- sec/stc; R_2
```

```
## [1] 0.02902378
```

Lo que nos dice este modelo es que ajusta los datos en un 2.902%

Por linea de comando

```
summary(ols_cps)$sigma
```

```
## [1] 9.991862
```

```
summary(ols_cps)$r.squared
```

```
## [1] 0.02902378
```