

Análisis para el dataset TeachingRatings

Joel Alejandro Zavala Prieto

Contents

Información de contacto	2
Modelando el dataset TeachingRatings	3
Descripción	3
Propuesta de modelo	4
Visualización	4
Ajuste por forma matricial	5
Modelo ajustado	6
Medidas extras	7
Observaciones ajustadas	7
Residuales	7
Error estándar de la regresión	7
Suma de los residuos al cuadrado	8
Suma explicada de cuadrados	8
Suma total de cuadrados	8
Coeficiente de determinación	8
Coeficiente de determinación ajustado	8
Inferencias a los parámetros obtenidos	9
Test de Shapiro-Wilk	9
Test de Breusch-Pagan	9
Matrix de varianza-covarianza homocedástica	10
¿Es significativo el coeficiente de la variable intro?	11
Error estándar para la variable intro	11
Prueba de hipótesis de significancia	11
Intervalo de confianza al nivel 95%	11
Predicción con intervalos de confianza al 90%	12

Información de contacto

Mail: alejandro.zavala1001@gmail.com

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

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

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

Modelando el dataset TeachingRatings

Descripción

En esta parte se hará un análisis del conjunto de datos “TeachingRatings”. Cuya descripción citare

“Contiene datos sobre evaluaciones del curso, características del curso y profesor características de 463 cursos para los años académicos 2000-2002 en la Universidad de Texas en Austin. Estos datos fueron proporcionados por el profesor Daniel Hamermesh de la Universidad de Texas en Austin y se utilizaron en su artículo con Amy Parker, “Beauty in the Classroom: Instructors’ Pulcritud y productividad pedagógica putativa”. Revisión de la economía de la educación, agosto 2005, vol. 24, núm. 4, págs. 369-376. ”

Viendo variables que contiene este dataset

```
## [1] "minority"    "age"         "female"      "onecredit"   "beauty"
## [6] "course_eval" "intro"       "nnenglish"
```

Ademas la descripción de las variables que integran el dataset son:

Variable	Descripción
course_eval	Puntaje de evaluación docente “general del curso”, en una escala de 1 (muy insatisfactorio) a 5 (excelente)
beauty	Calificación de la apariencia física del instructor por un panel de seis estudiantes, promediado entre los seis panelistas, se desplazaron para tener una media de cero.
female	1 Si el instructor es mujer, 0 si el instructor es hombre
minority	1 Si el instructor es no-Blanco, 0 si el instructor es blanco
nnenglish	1 Si el instructor no es hablante nativo de inglés, 0 si el instructor es hablante nativo de inglés
intro	1 si el curso es introductorio (principalmente cursos grandes para estudiantes de primer y segundo año), 0 Si el curso no es introductorio
onecredit	1 Si el curso es optativo de un solo crédito (yoga, aeróbicos, danza, etc.) , 0 en otro caso
age	Edad del profesor

Fuente: https://wps.pearsoned.com/aw_stock_ie_3/178/45691/11696965.cw/index.html

Propuesta de modelo

Se propone el modelo

$$courseeval_i = \beta_0 + \beta_1 beauty_i + \beta_2 intro_i + \beta_3 onecredit_i + \beta_4 female_i + \beta_5 minority_i + \beta_6 nnenglish_i + \epsilon_i$$

Cuya función ajustada es:

$$\hat{courseeval}_i = \hat{\beta}_0 + \hat{\beta}_1 beauty_i + \hat{\beta}_2 intro_i + \hat{\beta}_3 onecredit_i + \hat{\beta}_4 female_i + \hat{\beta}_5 minority_i + \hat{\beta}_6 nnenglish_i$$

Visualización

Se muestra a continuación las variables a tomar para el análisis (mostrando primeras 25 observaciones)

course_eval	beauty	intro	onecredit	female	minority	nnenglish
4.3	0.2899157	0	0	1	1	0
4.5	-0.7377322	0	0	0	0	0
3.7	-0.5719836	0	0	0	0	0
4.3	-0.6779634	0	0	1	0	0
4.4	1.5097940	0	0	1	0	0
4.2	0.5885687	0	0	0	0	0
4.0	-0.1260010	0	0	1	0	0
3.4	-0.2581899	0	0	1	0	0
4.5	0.1496926	0	0	1	0	0
3.9	0.5409170	0	0	0	0	0
3.1	0.2316134	0	0	0	1	1
4.0	-0.0666737	0	0	0	0	0
3.8	0.2168924	0	0	0	0	0
3.4	-0.2586962	0	0	0	0	1
2.9	0.5502878	0	0	1	0	0
4.5	-0.0620358	0	0	0	0	0
4.0	-0.9823852	0	0	1	0	0
3.8	-0.0543440	1	0	0	0	0
4.3	-0.0680143	0	0	1	0	0
3.4	0.0294137	0	0	1	0	0
3.4	0.2394859	0	0	1	0	0
3.3	-0.8487269	0	0	1	1	1
4.3	-0.7330914	0	0	0	0	0
4.4	-0.1111221	0	0	0	0	0
4.6	1.7755170	0	0	1	0	0

Ajuste por forma matricial

Se sabe que de forma matricial se tiene

$$Y = X\beta$$

El tamaño de la muestra es:

```
## [1] 463
```

La matrix $X^t X$

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 4.63e+02    0.000029 157.00000 27.000000 195.00000 64.000000 28.00000
## [2,] 2.90e-05 287.347896    5.62515 -7.240029 22.64127  4.161917  0.89494
## [3,] 1.57e+02    5.625150 157.00000 26.000000 60.00000 32.000000  2.00000
## [4,] 2.70e+01   -7.240029 26.00000 27.000000  9.00000 13.000000  1.00000
## [5,] 1.95e+02   22.641269 60.00000  9.000000 195.00000 36.000000 12.00000
## [6,] 6.40e+01    4.161917 32.00000 13.000000 36.00000 64.000000 15.00000
## [7,] 2.80e+01    0.894940  2.00000  1.000000 12.00000 15.000000 28.00000
```

La matrix $(X^t X)^{-1}$

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 0.005345  0.000403 -0.003694  0.000034 -0.003960 -0.000580 -0.003088
## [2,] 0.000403  0.003581 -0.000431  0.001342 -0.000691 -0.000285 -0.000086
## [3,] -0.003694 -0.000431  0.011255 -0.006756  0.000694 -0.001819  0.003823
## [4,] 0.000034  0.001342 -0.006756  0.047012  0.000937 -0.007350  0.002262
## [5,] -0.003960 -0.000691  0.000694  0.000937  0.009209 -0.001946  0.000994
## [6,] -0.000580 -0.000285 -0.001819 -0.007350 -0.001946  0.022065 -0.010005
## [7,] -0.003088 -0.000086  0.003823  0.002262  0.000994 -0.010005  0.043385
```

Para obtener finalmente la matrix $(X^tX)^{-1}X^tY$

```
##           [,1]
## [1,]  4.0682885
## [2,]  0.16560995
## [3,]  0.01132502
## [4,]  0.63452703
## [5,] -0.17347744
## [6,] -0.16661542
## [7,] -0.24416127
```

Que por linea de comando

```
##
## Call:
## lm(formula = course_eval ~ beauty + intro + onecredit + female +
##      minority + nnenglish, data = TeachingRatings)
##
## Coefficients:
## (Intercept)      beauty      intro  onecredit      female  minority
##    4.06829    0.16561    0.01133    0.63453   -0.17348   -0.16662
##  nnenglish
##   -0.24416
```

Modelo ajustado

De tal forma el modelo ajustado es:

$$\begin{aligned} \hat{courseval}_i = & 4.06829 + 0.16561beauty_i + 0.01133intro_i + 0.63453onecredit_i - 0.17348female_i \\ & - 0.16662minority_i - 0.24416nnenglish_i \end{aligned}$$

Medidas extras

Observaciones ajustadas

Las observaciones ajustadas de forma matricial son (mostrando primeros 5 observaciones):

```
## [1] 3.776209 3.946113 3.973563 3.782534 4.144848
```

Las observaciones ajustadas por linea de comando (mostrando primeras 5 observaciones):

```
##          1          2          3          4          5
## 3.776209 3.946113 3.973563 3.782534 4.144848
```

Residuales

Los residuales de forma matricial son (mostrando primeros 5 observaciones):

```
## [1] 0.5237911 0.5538869 -0.2735627 0.5174661 0.2551517
```

Los residuales por linea de comando son (mostrando primeros 5 observaciones):

```
##          1          2          3          4          5
## 0.5237911 0.5538869 -0.2735627 0.5174661 0.2551517
```

Error estándar de la regresión

El error estándar de la regresión de forma matricial es:

```
##          [,1]
## [1,] 0.513505
```

El error estándar de la regresión por linea de comando es:

```
## [1] 0.513505
```

Suma de los residuos al cuadrado

SRC

```
##           [,1]  
## [1,] 120.2415
```

Suma explicada de cuadrados

SEC

```
##           [,1]  
## [1,] 21.99716
```

Suma total de cuadrados

STC

```
##           [,1]  
## [1,] 142.2386
```

Coefficiente de determinación

De forma manual

```
##           [,1]  
## [1,] 0.1546497
```

Por linea de comando

```
## [1] 0.1546497
```

Coefficiente de determinación ajustado

De forma manual

```
##           [,1]  
## [1,] 0.1435267
```

Por linea de comando

```
## [1] 0.1435267
```


Inferencias a los parámetros obtenidos

Test de Shapiro-Wilk

```
##  
## Shapiro-Wilk normality test  
##  
## data: mco_teachingRatings$residuals  
## W = 0.97881, p-value = 2.826e-06
```

Test de Breusch-Pagan

```
##  
## studentized Breusch-Pagan test  
##  
## data: mco_teachingRatings  
## BP = 4.3493, df = 6, p-value = 0.6295
```

Matrix de varianza-covarianza homocedástica

De forma manual

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,]  0.001409  0.000106 -0.000974  0.000009 -0.001044 -0.000153 -0.000814
## [2,]  0.000106  0.000944 -0.000114  0.000354 -0.000182 -0.000075 -0.000023
## [3,] -0.000974 -0.000114  0.002968 -0.001782  0.000183 -0.000480  0.001008
## [4,]  0.000009  0.000354 -0.001782  0.012396  0.000247 -0.001938  0.000597
## [5,] -0.001044 -0.000182  0.000183  0.000247  0.002428 -0.000513  0.000262
## [6,] -0.000153 -0.000075 -0.000480 -0.001938 -0.000513  0.005818 -0.002638
## [7,] -0.000814 -0.000023  0.001008  0.000597  0.000262 -0.002638  0.011440
```

Por linea de comando

```
##           (Intercept)      beauty      intro onecredit      female      minority
## (Intercept)    0.001409  0.000106 -0.000974  0.000009 -0.001044 -0.000153
## beauty         0.000106  0.000944 -0.000114  0.000354 -0.000182 -0.000075
## intro          -0.000974 -0.000114  0.002968 -0.001782  0.000183 -0.000480
## onecredit       0.000009  0.000354 -0.001782  0.012396  0.000247 -0.001938
## female         -0.001044 -0.000182  0.000183  0.000247  0.002428 -0.000513
## minority        -0.000153 -0.000075 -0.000480 -0.001938 -0.000513  0.005818
## nnenglish       -0.000814 -0.000023  0.001008  0.000597  0.000262 -0.002638
##               nnenglish
## (Intercept) -0.000814
## beauty       -0.000023
## intro         0.001008
## onecredit     0.000597
## female        0.000262
## minority     -0.002638
## nnenglish     0.011440
```

¿Es significativo el coeficiente de la variable intro?

Error estándar para la variable intro

Se calculo que el estimado del coeficiente de la variable intro es:

```
##           [,1]
## [1,] 0.01132502
```

De forma matricial el error estándar es:

```
## [1] 0.05447785
```

Prueba de hipótesis de significancia

Obteniendo estadístico t para el coeficiente de la variable intro de forma matricial

```
##           [,1]
## [1,] 0.2078831
```

Por linea de comando

```
##
## t test of coefficients:
##
##           Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  4.068289   0.037543 108.3636 < 2.2e-16 ***
## beauty       0.165610   0.030730  5.3893 1.137e-07 ***
## intro        0.011325   0.054478  0.2079 0.8354131
## onecredit    0.634527   0.111339  5.6990 2.165e-08 ***
## female      -0.173477   0.049279 -3.5203 0.0004744 ***
## minority     -0.166615   0.076278 -2.1843 0.0294483 *
## nnenglish    -0.244161   0.106958 -2.2828 0.0229028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Intervalo de confianza al nivel 95%

Obteniendo un intervalo de confianza al 95% para el coeficiente de la variable intro de forma matricial

```
##           [,1]      [,2]
## [1,] -0.09545156 0.1181016
```

Por linea de comando

```
##           2.5 %      97.5 %
## (Intercept)  3.99451020  4.14206751
## beauty       0.10522081  0.22599910
## intro       -0.09573375  0.11838379
## onecredit    0.41572568  0.85332839
## female      -0.27031970 -0.07663518
## minority     -0.31651619 -0.01671464
## nnenglish    -0.45435268 -0.03396985
```

Predicción con intervalos de confianza al 90%

Se quiere hacer una predicción para los siguientes estándares al 90%

beauty	intro	onecredit	female	minority	nnenglish
0.25	0	1	1	1	1
0.40	1	1	1	1	1
0.65	1	0	1	1	1
0.80	1	0	1	1	0

De forma matricial:

```
##           [,1]  
## [1,] 4.159964  
## [2,] 4.196131  
## [3,] 3.603006  
## [4,] 3.872009
```

Que por linea de comando a un nivel del 90%

	fit	lwr	upr
4.159964	3.279513	5.040415	
4.196131	3.317326	5.074936	
3.603006	2.734839	4.471173	
3.872009	3.014737	4.729281	