Universidad Nacional Autónoma de México

Regresión lineal múltiple.

2022

Karla Michell Trigueros Torres

1. Algunas propiedades

1.1. Supuestos de regresión lineal múltiple por MCO

Tenemos que el modelo planteado por la regresión múltiple es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2+i} + ... + \beta_k x_{ki} + u_i$$

$i \in \{1, .., n\}$, n el tamaño de la muesta.
Y el modelo estimado:

$$\hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2+i} + ... + \hat{\beta}_k x_{ki} + e_i$$

Donde los errores estimados e_i estan dados por:
$e_i = y_i - \hat{y}_i$
El objetivo de la regresión lineal múltiple es minimizar estos errores.

Bajo los siguientes supuestos:

- $E[U|X] = 0$
- U y X_j son independientes para todos las j en la muestra (suponemos es de tamaño n).
- $U \sim N_n(0, \sigma_{u^2} I_n)$

1.2. Estimador de β

1. $\hat{\beta} = E[(X'X)]^{-1}E[X'Y]$

Demostración

Para encontrar el mínimo, derivamos:

$$\begin{aligned}\frac{\partial S}{\partial \hat{\beta}} &= \frac{\partial \sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2+i} + ... + \hat{\beta}_k x_{ki}))^2}{\partial \hat{\beta}} \\ &= -X'Y - X'Y + 2(X'X)\hat{\beta} \Longrightarrow\end{aligned} \tag{1}$$

$$X'Y = (X'X)\hat{\beta} \Longrightarrow \tag{2}$$

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{3}$$

□

2. 1$\Longrightarrow \beta = b + (X'X)^{-1}X'U$

Demostración

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + U) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'U \\ &= \beta + (X'X)^{-1}X'U\end{aligned}$$

(4)

□

3. $\hat{\beta}$ **es insesgado.**

Demostración

Sabemos que:

$$\begin{aligned} E[\hat{\beta}|X] &= E[\beta + (X'X)^{-1}X'U|X] = \beta + (X'X)X'E[U|X] \\ &= \beta + (X'X)X'(0) = b \\ &\Longrightarrow E[E[\hat{\beta}|X]] = E[b] = b \end{aligned} \tag{5}$$

□

4. $Var(\hat{\beta}) = \sigma_u^2 E[(X'Y)^{-1}]$

Demostración

$$\begin{aligned} Var[\hat{\beta}|X] &= Var[\beta + (X'X)^{-1}X'U|X] = Var[(X'X)^{-1}X'U|X] \\ &= (X'X)^{-1}X'Var[U|X] = (X'X)^{-1}X'\sigma_u^2 I_n X(X'X)^{-1} \\ &= \sigma_u^2(X'X)^{-1}(X'X)(X'X)^{-1} = \sigma_u^2(X'X)^{-1} \\ \Longrightarrow Var[\hat{\beta}|X] &= Var(E(\hat{\beta}|X)) + E(Var[\hat{\beta}|X]) \\ &= \sigma_u^2 E[(X'X)^{-1}] \end{aligned} \tag{6}$$

□

5. $\hat{\beta} \sim N_n(b, \sigma_u^2(X'X)^{-1})$

Demostración

Tenemos que: $\hat{\beta} = b + (X'X)^{-1}XU$, si tomamos $A = (X'X)^{-1}X$, sabiendo que $U \sim N_k(0, \sigma_{u^2} I_k)$.

Además, se sabe que $X \sim N(\mu, \sigma) + b \sim N(a\mu + b, \sigma^2 a^2)$.

$$\Longrightarrow \hat{\beta} \sim N_k(b, \sigma_u^2(X'X)^{-1})$$

□

1.3. Estimador de la varianza

Tenemos que:

$$Var[U|X] = \sigma_u^2 I_n \Longrightarrow \sigma_u^2 = E[(U_j - E(U_j))^2] = \frac{U'U}{n-k}$$

Sin embargo U no es observable, por esto, se usa la siguiente estimación:

Sea $\hat{U} = Y - X\hat{\beta}$, $\hat{\sigma_u^2} = \frac{\hat{U}'\hat{U}}{n-k}$

n tamaño de la muestra y k el número de coeficientes estimados.

6.$\hat{\sigma}^2 = \frac{U'MU}{n-k}$ **con** $M = I_n - X(X'X)^{-1}X'$

$$\begin{aligned} \hat{U}'\hat{U} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (X\beta + U - X\hat{\beta})'(X\beta + U - X\hat{\beta}) \\ &= (U - X(\beta - \beta))'(U - X(\beta - \beta)) = (U' - (\hat{\beta} - \beta)'X')(U - X(\hat{\beta}' - \beta)) \\ &= U'U - U'X(\hat{\beta} - \beta) - (\hat{\beta} - \beta)'X'U + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \end{aligned} \tag{7}$$

Donde $-U'X(\hat{\beta}-\beta)-(\hat{\beta}-\beta)'X'U$ son escalares del mismo valor.

$$\hat{U}'\hat{U} = U'U - U'X(\hat{\beta}-\beta) - 2U'X(\hat{\beta}-\beta) + (\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta$$
$$= U'U - 2U'X[(X'X)^{-1}X'U] + [(X'X)^{-1}X'U]'X'X(X'X)^{-1}X'U$$
$$= U'U - U'X(X'X)^{-1}X'U = U'(I_n - X(X'X)^{-1}X')U = U \tag{8}$$

□

Esta matriz M:

- es simétrica e idemppotente (MM=M)
- $P(rango(M)) = n-k$

7. $\frac{n-k}{\sigma_u^2}\hat{\sigma_u^2} \sim \chi^2_{n-k}$

Demostración

Proposición. $Z \sim N_m(0, I_m)$, M matriz simétrica a idempotente, y además $rango(M) = l \Longrightarrow Z'MZ \sim \chi^2_l$.

$$\frac{n-k}{\sigma_u^2}\hat{\sigma_u^2} = \frac{n-k}{\sigma_u^2}\frac{U'MU}{n-k} = \frac{U'MU}{\sigma_u^2} = \frac{U'}{\sigma}M\frac{U}{\sigma}$$

Como M es simetrica a idempotente y $Rango(M) = n-k$ y $\frac{U}{\sigma_u^2} \sim N_n(0, I_n) \Longrightarrow$

$$\frac{n-k}{\sigma_u^2}\hat{\sigma_u^2} = \frac{U'}{\sigma}M\frac{U}{\sigma} \sim \chi^2_{n-k}$$

□

1.3.1. Estadístico t

8. $\frac{\hat{\beta}-\beta}{\hat{\sigma_u}\sqrt{(X'X)^{-1}}} \sim t_{n-k}$

Demostración

Sabemos que:

$$\hat{\beta} \sim N_k(\beta, \sigma_u^2(X'X)^{-1} \Longrightarrow \hat{\beta}_i \sim N_k(\beta_i, \sigma_u^2(X'X)^{-1}_{ii} \Longrightarrow \frac{\hat{\beta}_i - \beta_i}{\sigma_u\sqrt{(X'X)^{-1}_{ii}}} =\sim N(0,1)$$

Por otro lado,

$$\frac{n-k}{\sigma_u^2}\hat{\sigma_u^2} \sim \chi^2_{n-k}$$

$$\Longrightarrow \frac{\frac{\hat{\beta}_i-\beta_i}{\sigma_u\sqrt{(X'X)^{-1}_{ii}}}}{\sqrt{\frac{n-k}{\sigma_u^2}\frac{\hat{\sigma_u}^2}{n-k}}} \sim t_{n-k} \Longrightarrow \frac{\hat{\beta}-\beta}{\hat{\sigma}_u\sqrt{(X'X)^{-1}_{ii}}} \sim t_{n-k}$$

□

Esta propiedad se usa para rechazar o aceptar $H_0 = (\beta = 0$ es decir, no hay efecto de la variable independiente en la dependiente.).

9. El intervalo de confianza de $\hat{\beta}_i$ es $(\hat{\beta}_{1i} - t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}}, \hat{\beta}_i + t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}})$

Demostración

$$
\begin{aligned}
1-\alpha &= P[-t_{1-\frac{\alpha}{2}} < \frac{\hat{\beta}-\beta}{\hat{\sigma}_u\sqrt{(X'X)^{-1}_{ii}}} < t_{1-\frac{\alpha}{2}}] \\
&= P[-t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}} - \hat{\beta} < -\beta < t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}} - \hat{\beta}] \\
&= P[t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}} + \hat{\beta} > \beta > -t_{1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{(X'X)^{-1}} + \hat{\beta}]
\end{aligned} \tag{9}
$$

□

1.3.2. Intervalos de confianza de Y

Los intervalos de confianza se calculan de la siguiente forma:

Sean los valores de $X = X_0$ y $t = t_{(n-k,\alpha/2)}$.

Valor medio.

$$\mu_y = Y' \pm t\hat{s}_u^2\sqrt{X_0'(X'X)^{-1}X_0}$$

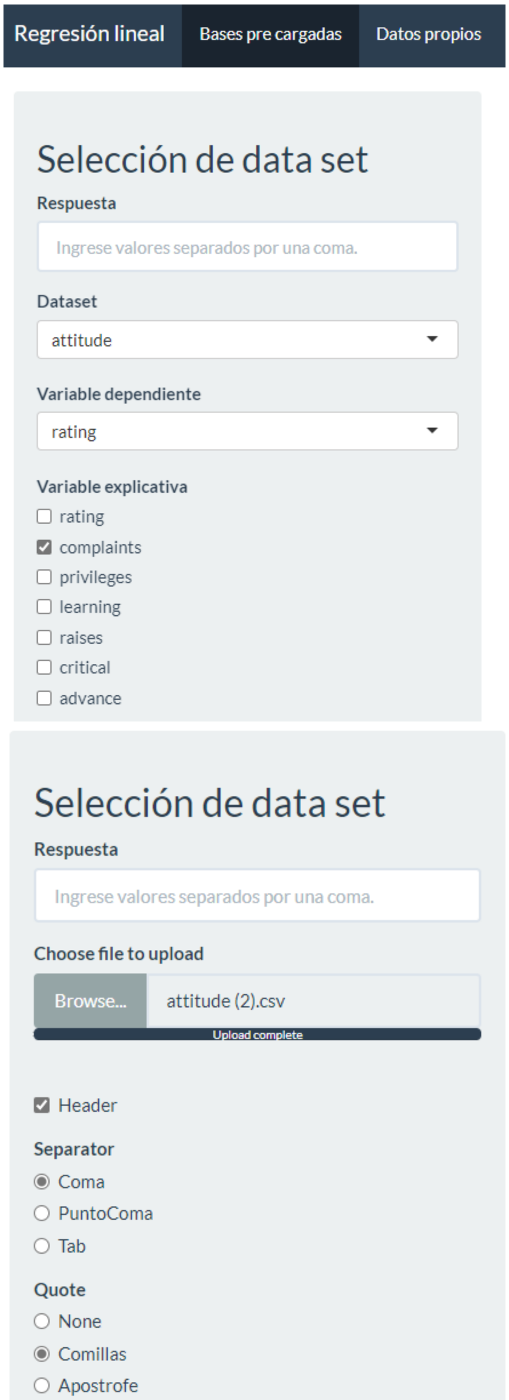
Valor puntual.

$$\mu_y = Y' \pm t\hat{s}_u^2\sqrt{1 + X_0'(X'X)^{-1}X_0}$$

2. App shiny

2.1. Elección de datos

La aplicación permite elegir data sets pre cargados en la aplicación, o subir los propios a partir de un csv, así mismo, las opciones de variables explicativas/dependientes cambian dependiendo del data set, así como el número.



2.2. Exploración de datos

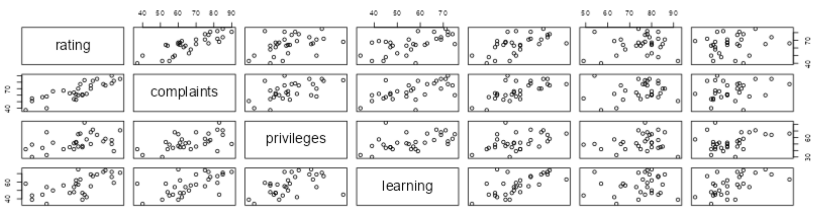
Muestra las gráficas y matriz de covarianza, así como la correlación entre la variable seleccionada como dependiente con cada una de las explicativas.

Exploración de data sets | Respuestas | Residuales

Correlación de las variables selecionadas:

y	Correlacion.con.x
complaints	0.83
learning	0.62

rating	complaints	privileges	learning	raises	critical	advance
43.00	51.00	30.00	39.00	61.00	92.00	45.00
63.00	64.00	51.00	54.00	63.00	73.00	47.00
71.00	70.00	68.00	69.00	76.00	86.00	48.00
61.00	63.00	45.00	47.00	54.00	84.00	35.00
81.00	78.00	56.00	66.00	71.00	83.00	47.00
43.00	55.00	49.00	44.00	54.00	49.00	34.00



2.3. Modelo ajustado

Se muestra la gráfica del modelo ajustado por método de mínimos cuadrados ordinarios con su intervalo de confianza calculado en función del valor ingresado en el slider disponible al usuario, así también se muestra el estadístico t correspondiente a cada coeficiente de regresión y su valor p.

Exploración de data sets | Respuestas | Residuales

Intervalo de confianza

0.5 0.9 0.99

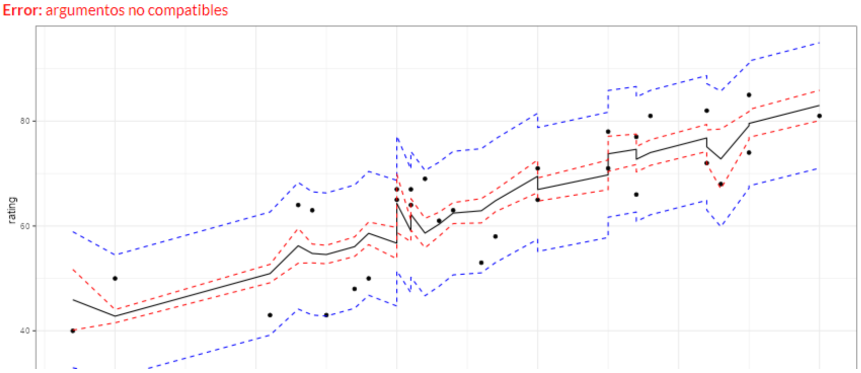
Mostrar gráfica con respecto a:

complaints

Modelo calculado

parametro	valor	error.est	t.est	p.value	lim.inf	lim.sup
alfa	9.87	7.06	1.40	0.09	-9.75	23.87
beta. complaints	0.64	0.12	5.43	0.00	-0.98	1.21
beta. learning	0.21	0.13	1.57	0.06	-0.23	0.49

Error: argumentos no compatibles



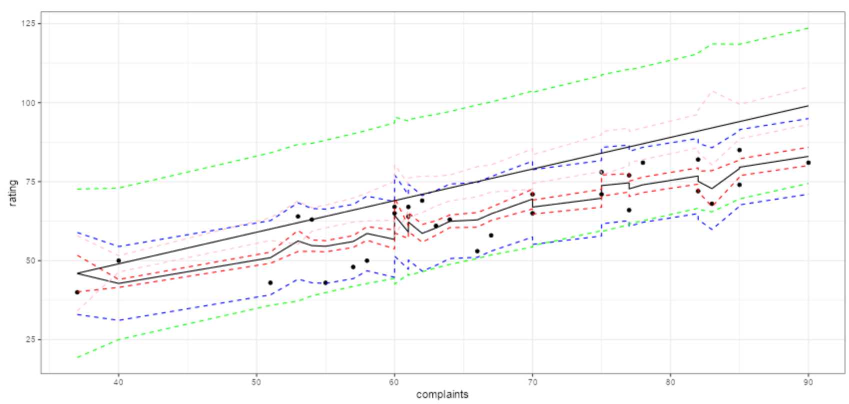
El usuario puede ingresar sus respuestas y comparar la gráfica ajustada de sus propias respuestas y la suma de errores cuadrados, contra los obtenidos de manera automática usando MCO.

Selección de data set

Respuesta

9,1,0

Modelo	Suma.de.errores	Suma.de.errores.2
Usuario	-329.00	5287.00
MCO	-0.00	1254.65



Se muestran los valores calculados por el modelo con valores dados por el usuario, así como sus intervalos de confianza en valor y media.

complaints

Predecir con el modelo ajustado

1,2/5,7

Show 10 entries

Search:

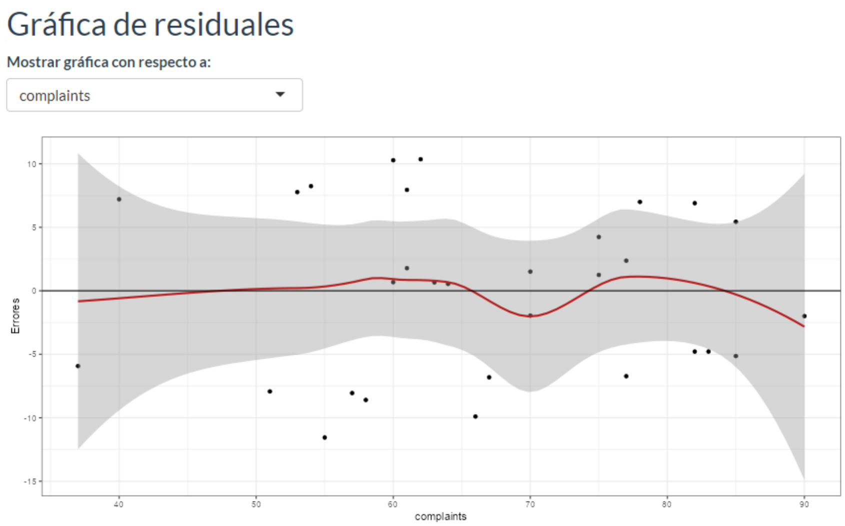
	y.est	lim.punt.inf	lim.punt.sup	lim.med.inf	lim.med.sup
X1.2	10.9367817056384	-0.676761966429025	22.5503253777058	10.6909465414893	11.1826168697875
X5.7	14.5668112963954	2.9397592824642	26.1938633103265	13.9549478806103	15.1786747121805

Showing 1 to 2 of 2 entries

Previous 1 Next

2.4. Pruebas de hipótesis

Se muestra la gráfica de residuales con respecto a la variable elegida.



Se acepta o rechaza la hipótesis de normalidad de los residuales dependiendo del valor de rechazo elegido por el usuario.