Notas de Clase Sobre Regresión Lineal
Regresión Lineal Simple (RLS): Parte I

Nelfi González Alvarez

Profesora Asociada Escuela de Estadística

e-mail: ngonzale@unal.edu.co

Isabel Cristina Ramírez Guevara

Profesora Asociada Escuela de Estadística

e-mail: iscramirezgu@unal.edu.co

Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín

UNIVERSIDAD NACIONAL DE COLOMBIA

Escuela de Estadística

2021

Contenido I

Contenido

Introducción

- Se desea modelar una relación de tendencia entre un par de variables (X, Y);
- *Pero muchas veces sólo se cuenta con datos observacionales sobre estas variables, obtenidos bajo situaciones no controladas estrictamente.*
- Posiblemente, no hubo replicación de las condiciones observadas
- *Bajo estas circunstancias lo más conveniente es el método de mínimos cuadrados (MCO).*

Introducción

- Se desea modelar una relación de tendencia entre un par de variables (X, Y);
- *Pero muchas veces sólo se cuenta con datos observacionales sobre estas variables, obtenidos bajo situaciones no controladas estrictamente.*
- Posiblemente, no hubo replicación de las condiciones observadas
- *Bajo estas circunstancias lo más conveniente es el método de mínimos cuadrados (MCO).*

Introducción

- Se desea modelar una relación de tendencia entre un par de variables (X, Y);
- *Pero muchas veces sólo se cuenta con datos observacionales sobre estas variables, obtenidos bajo situaciones no controladas estrictamente.*
- Posiblemente, no hubo replicación de las condiciones observadas
- *Bajo estas circunstancias lo más conveniente es el método de mínimos cuadrados (MCO).*

Introducción

- Se desea modelar una relación de tendencia entre un par de variables (X, Y);
- *Pero muchas veces sólo se cuenta con datos observacionales sobre estas variables, obtenidos bajo situaciones no controladas estrictamente.*
- Posiblemente, no hubo replicación de las condiciones observadas
- *Bajo estas circunstancias lo más conveniente es el método de mínimos cuadrados (MCO).*

Contenido

Nomenclatura

En el modelo teórico:

- Y: Variable respuesta o dependiente.
- X: Variable predictora, independiente o regresora.
- E: Error aleatorio
- β_0, β_1: Parámetros de la regresión. β_0 es el intercepto y β_1 la pendiente de la línea recta.
- $\mu_{Y|X}$: Valor esperado condicional de Y dado X, es decir, $\mathrm{E}[Y|X]$

En el modelo estimado:

- $\widehat{\beta}_0$: Estimador del parámetro β_0
- $\widehat{\beta}_1$: Estimador del parámetro β_1.
- $\widehat{E}$: Residual, es una estimación del error aleatorio.
- $\widehat{Y}$: Es la estimación de $\mu_{Y|X}$.

Nomenclatura

En el modelo teórico:

- Y: Variable respuesta o dependiente.
- X: Variable predictora, independiente o regresora.
- E: Error aleatorio
- β_0, β_1: Parámetros de la regresión. β_0 es el intercepto y β_1 la pendiente de la línea recta.
- $\mu_{Y|X}$: Valor esperado condicional de Y dado X, es decir, $\mathrm{E}[Y|X]$

En el modelo estimado:

- $\widehat{\beta}_0$: Estimador del parámetro β_0
- $\widehat{\beta}_1$: Estimador del parámetro β_1.
- $\widehat{E}$: Residual, es una estimación del error aleatorio.
- $\widehat{Y}$: Es la estimación de $\mu_{Y|X}$.

Nomenclatura

En el modelo teórico:

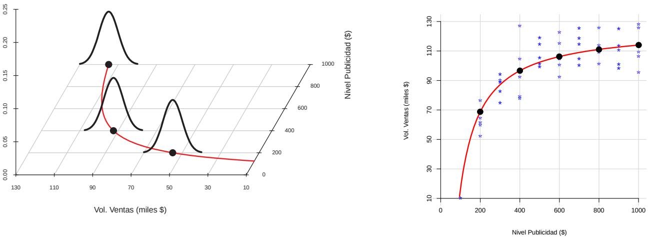
- Y: Variable respuesta o dependiente.
- X: Variable predictora, independiente o regresora.
- E: Error aleatorio
- β_0, β_1: Parámetros de la regresión. β_0 es el intercepto y β_1 la pendiente de la línea recta.
- $\mu_{Y|X}$: Valor esperado condicional de Y dado X, es decir, $\mathrm{E}[Y|X]$

En el modelo estimado:

- $\widehat{\beta}_0$: Estimador del parámetro β_0
- $\widehat{\beta}_1$: Estimador del parámetro β_1.
- $\widehat{E}$: Residual, es una estimación del error aleatorio.
- $\widehat{Y}$: Es la estimación de $\mu_{Y|X}$.

Contenido

Función de medias en términos generales



(a) (b)

Figura 1: Función de regresión: (a) Curva que pasa por las medias de Y dado cada valor de X, es decir, $\mu_{Y|x} = f(x)$ (b) Curva con respecto a la cual es mínima la distancia vertical de las observaciones (x_i, y_i) $i = 1, 2, \ldots, n$.

Función de medias con forma lineal (MRLS)

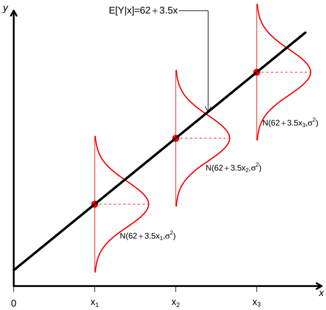


Figura 2: Recta en el MRLS como la función de la media condicional de $Y|x$. En este ejemplo se ha asumido que $Y|x \sim N\left(62 + 3.5x, \sigma^2\right)$, luego, la recta de regresión corresponde a $\mathrm{E}[Y|x] = 62 + 3.5x$ y en cada nivel de x se tiene la misma varianza para Y alrededor de la respectiva media condicional.

Contenido

Modelo de Regresión Lineal Simple (MRLS)

Con n pares (X_i, Y_i), $i = 1, 2, \ldots, n$,

$$Y|X_i = \beta_0 + \beta_1 X_i + E_i, i = 1, 2, \ldots, n \tag{1}$$

$$\mathrm{E}\left[Y|X_i\right] = \beta_0 + \beta_1 X_i \tag{2}$$

Bajo el supuesto $E_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$, iid: Independientes e idénticamente distribuidos.

Nota 4.1

Por simplicidad escribimos,

$$Y_i = \beta_0 + \beta_1 X_i + E_i,\ E_i \overset{iid}{\sim} N\left(0, \sigma^2\right) i = 1, 2, \ldots, n \tag{3}$$

- *Los errores $E_i = Y_i - (\beta_0 + \beta_1 X_i)$,*
- *Los errores E_i no son observables, desde que los parámetros β_j, $j = 0, 1$, son desconocidos y deben estimarse.*

Modelo de Regresión Lineal Simple (MRLS)

Con n pares (X_i, Y_i), $i = 1, 2, \ldots, n$,

$$Y|X_i = \beta_0 + \beta_1 X_i + E_i, i = 1, 2, \ldots, n \tag{1}$$

$$\mathrm{E}\left[Y|X_i\right] = \beta_0 + \beta_1 X_i \tag{2}$$

Bajo el supuesto $E_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$, iid: Independientes e idénticamente distribuidos.

Nota 4.1

Por simplicidad escribimos,

$$Y_i = \beta_0 + \beta_1 X_i + E_i,\ E_i \overset{iid}{\sim} N\left(0, \sigma^2\right)\ i = 1, 2, \ldots, n \tag{3}$$

- *Los errores $E_i = Y_i - (\beta_0 + \beta_1 X_i)$,*
- *Los errores E_i no son observables, desde que los parámetros β_j, $j = 0, 1$, son desconocidos y deben estimarse.*

Modelo de Regresión Lineal Simple (MRLS)

Con n pares (X_i, Y_i), $i = 1, 2, \ldots, n$,

$$Y|X_i = \beta_0 + \beta_1 X_i + E_i, i = 1, 2, \ldots, n \tag{1}$$

$$\mathrm{E}\left[Y|X_i\right] = \beta_0 + \beta_1 X_i \tag{2}$$

Bajo el supuesto $E_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$, iid: Independientes e idénticamente distribuidos.

Nota 4.1

Por simplicidad escribimos,

$$Y_i = \beta_0 + \beta_1 X_i + E_i,\ E_i \overset{iid}{\sim} N\left(0, \sigma^2\right) i = 1, 2, \ldots, n \tag{3}$$

- *Los errores $E_i = Y_i - (\beta_0 + \beta_1 X_i)$,*
- *Los errores E_i no son observables, desde que los parámetros β_j, $j = 0, 1$, son desconocidos y deben estimarse.*

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{4}$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{4}$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \quad (4)$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \quad (4)$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con* E_i*.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\, i$.
- *Los* E_i *son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los* Y_i *(o sea* $Y|X_i$*) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a* σ^2*:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{4}$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{4}$$

Consideraciones

- *Y es una v.a observada mediante la selección de los valores de X en un intervalo de interés.*
- X no se considera v.a, sus valores son definidos con anticipación y se miden sin error.
- *Si lo anterior no se cumple, el método MCO sigue siendo siendo válido si los errores en los valores de X son pequeños comparados con E_i.*
- Los datos constituyen una muestra representativa de un medio acerca del cual se desea generalizar. Si no es así, no es apropiado realizar extrapolaciones.
- *El modelo de regresión es lineal en los parámetros.*
- La varianza de los E_i es igual a σ^2, $\forall\ i$.
- *Los E_i son variables aleatorias normales, independientes e idénticamente distribuidos con media cero. Los Y_i (o sea $Y|X_i$) son variables aleatorias normales e independientes, pero no son idénticamente distribuidos, pues no tienen misma media aunque sí tienen misma varianza igual a σ^2:*

$$Y|X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{4}$$

Interpretación de los parámetros

Nota 4.2

La interpretación de los coeficientes de regresión:

- *β_1: Representa el cambio en la media de Y dado un cambio unitario en X.*
- *β_0: Si el rango en que se observa X incluye a $x = 0$, entonces β_0 corresponde a la media de la distribución de Y cuando $x = 0$. Sin embargo, si $x = 0$ no ha sido observado en los datos, entonces β_0 no tiene interpretación práctica en el modelo de regresión.*

Interpretación de los parámetros

Nota 4.2

La interpretación de los coeficientes de regresión:

- *β_1: Representa el cambio en la media de Y dado un cambio unitario en X.*
- *β_0: Si el rango en que se observa X incluye a $x = 0$, entonces β_0 corresponde a la media de la distribución de Y cuando $x = 0$. Sin embargo, si $x = 0$ no ha sido observado en los datos, entonces β_0 no tiene interpretación práctica en el modelo de regresión.*

Contenido

Estimación por Mínimos Cuadrados (MCO)

Con n pares (x_i, y_i) de observaciones: hallar β_0 y β_1 que minimicen a

Suma de Cuad. del error:
$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_i)]^2 \quad (5)$$

Ecuaciones normales

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0, \quad \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

Conducen a

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \quad (6)$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2, \quad (7)$$

Estimadores resultantes

Intercepto: $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}, \quad (8)$

Pendiente: $\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad (9)$

o bien, $\widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \quad (10)$

Estimación por Mínimos Cuadrados (MCO)

Con n pares (x_i, y_i) de observaciones: hallar β_0 y β_1 que minimicen a

Suma de Cuad. del error: $$S(\beta_0, \beta_1) = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_i)]^2 \qquad (5)$$

Ecuaciones normales

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0, \quad \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

Conducen a

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \qquad (6)$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2, \qquad (7)$$

Estimadores resultantes

Intercepto: $$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}, \qquad (8)$$

Pendiente: $$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad (9)$$

o bien, $$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \qquad (10)$$

Estimación por Mínimos Cuadrados (MCO)

Con n pares (x_i, y_i) de observaciones: hallar β_0 y β_1 que minimicen a

Suma de Cuad. del error: $$S(\beta_0,\beta_1)=\sum_{i=1}^{n}E_i^2=\sum_{i=1}^{n}[Y_i-(\beta_0+\beta_1X_i)]^2 \qquad (5)$$

Ecuaciones normales

$$\frac{\partial S(\beta_0,\beta_1)}{\partial\beta_0}=0, \quad \frac{\partial S(\beta_0,\beta_1)}{\partial\beta_1}=0$$

Conducen a

$$\sum_{i=1}^{n}y_i=n\beta_0+\beta_1\sum_{i=1}^{n}x_i \qquad (6)$$

$$\sum_{i=1}^{n}x_iy_i=\beta_0\sum_{i=1}^{n}x_i+\beta_1\sum_{i=1}^{n}x_i^2, \qquad (7)$$

Estimadores resultantes

Intercepto: $$\widehat{\beta}_0=\bar{y}-\widehat{\beta}_1\bar{x}, \qquad (8)$$

Pendiente: $$\widehat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \qquad (9)$$

o bien, $$\widehat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}. \qquad (10)$$

Ejemplo con datos simulados

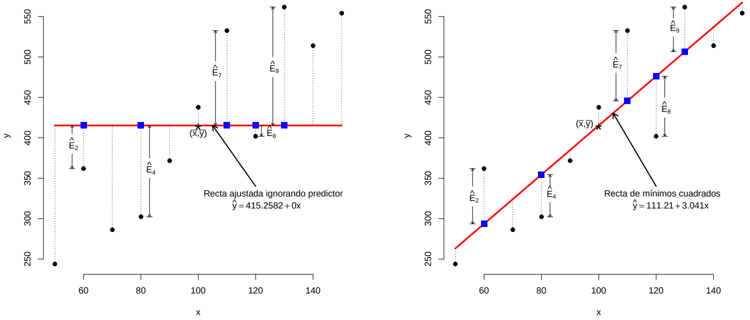
Tabla 1: Datos simulados bajo modelo $Y_i = 62 + 3.5x_i + E_i$, $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$, con $\sigma = 50$, y cálculos para ajuste del modelo de regresión lineal simple por mínimos cuadrados ordinarios.

i	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$
1	50	243.93	-50	2500	-171.33	29353.35	8566.41
2	60	361.95	-40	1600	-53.31	2841.76	2132.33
3	70	286.37	-30	900	-128.89	16612.16	3866.65
4	80	302.25	-20	400	-113.01	12770.85	2260.16
5	90	371.58	-10	100	-43.68	1907.78	436.78
6	100	437.67	0	0	22.41	502.29	0
7	110	532.59	10	100	117.33	13766.76	1173.32
8	120	401.91	20	400	-13.35	178.17	-266.96
9	130	561.66	30	900	146.40	21433.49	4392.05
10	140	513.88	40	1600	98.62	9726.26	3944.87
11	150	554.05	50	2500	138.79	19263.17	6939.59
n	$\bar{x}$	$\bar{y}$	$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)$	$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$	$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)$	$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$	$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$
11	100	415.26	0	11000	0	128356.05	33445.2

$\widehat{\beta}_1 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) / \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 3.041;$ $\quad \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} = 111.211;$

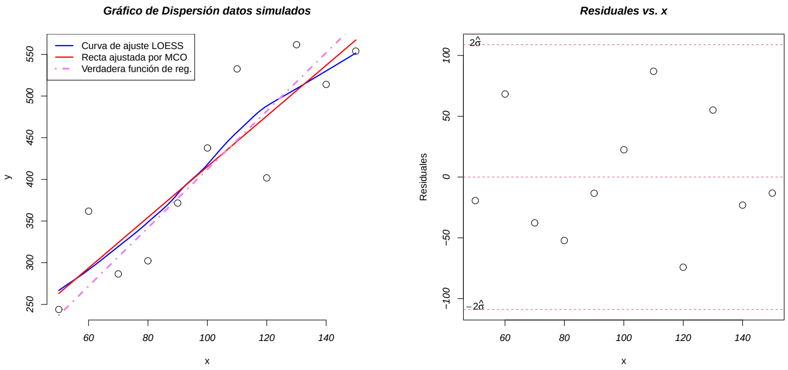
Tabla 2: Datos simulados, ajustes MCO y residuos (continuación Tabla 1).

i	x_i	y_i	$\widehat{y}_i$	$\widehat{E}_i$
1	50	243.93	263.24	-19.31
2	60	361.95	293.64	68.31
3	70	286.37	324.04	-37.67
4	80	302.25	354.45	-52.20
5	90	371.58	384.85	-13.27
6	100	437.67	415.26	22.41
7	110	532.59	445.66	86.93
8	120	401.91	476.07	-74.16
9	130	561.66	506.47	55.19
10	140	513.88	536.88	-22.00
11	150	554.05	567.28	-13.23
n	$\bar{x}$	$\bar{y}$	Media de los $\widehat{y}_i$	Media de los $\widehat{E}_i$
11	100	415.26	415.26	0
Ecuación ajustada: $\widehat{y}_i = 111.211 + 3.041x_i$, SSE $= \sum_{i=1}^{n} \widehat{E}_i^2 = 26666.83$				



(a) (b)

Figura 3: Ilustración del criterio de mínimos cuadrados. (a) Recta ajustada asumiendo como modelo a $Y_i = \beta_0 + E_i$, $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$, (b) Recta ajustada mediante mínimos cuadrados ordinarios con el modelo $Y_i = \beta_0 + \beta_1 x_i + E_i$, $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$. La suma de cuadrados de residuos, $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2 = 26666.83$, del ajuste en (b) es menor que en (a); para este último se tiene que $\text{SSE} = \text{SST} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = 128356$, en cambio en (b), $\text{SSE} < \text{SST}$.



(a) (b)

Figura 4: Resultados de ajuste en datos simulados. (a) Recta ajustada asumiendo como modelo a $Y_i = \beta_0 + \beta_1 x_i + E_i$, $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$, curva LOESS y verdadera función de regresión $\mu_{Y|x} = 62 + 3.5x_i$, (b) Residuos $\widehat{E}_i$ vs. x_i.

Funciones R para regresión lineal

Nombre	Descripción corta
`lm( )`	Ajusta de modelos de regresión lineal.
`summary( )`	Obtiene la tabla de parámetros estimados cuando se aplica a un objeto `lm`.
`anova( )`	Para Tabla ANOVA.
`residuals( )`	Obtiene los valores de los residuos de ajuste.
`fitted( )`	Obteniene los valores ajustados en la variable respuesta.
`predict( )`	Calcula predicciones puntuales y por intervalos de predicción. También permite calcular valores ajustados y los intervalos de confianza para $\mu_{Y\|x}$.
`qqnorm( ),qqline( )`	Para gráfico de probabilidad normal.
`shapiro.test( )`	Para test de normalidad Shapiro-Wilk.
`coef( )`	Extrae en un vector los valores estimados de los parámetros de la regresión.
`confint( )`	Retorna una matriz en la cual en cada fila se dan los límites de los intervalos de confianza para los parámetros del modelo.
`vcov( )`	Produce la matriz de varianzas y covarianzas estimadas, para el vector de parámetros estimado.

Tipos de sumas en MCO

Tabla 3: Principales sumas en el ajuste por mínimos cuadrados

Tipo de sumas	Expresión
Suma de cuadrados corregidos en x:	$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}(x_i - \bar{x})x_i$
Suma de cuadrados corregidos en y. También es conocida como suma de cuadrados totales o SST:	$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n}(y_i - \bar{y})y_i$
Suma de productos cruzados:	$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$
Suma de los cuadrados de los residuos o SSE. Es la estimación de $S(\beta_o, \beta_1)$. Sea $\widehat{E}_i = y_i - \hat{y}_i$ el i-ésimo residuo, entonces:	$\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2 = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = S_{yy} - \widehat{\beta}_1 S_{xy}$
Suma de cuadrados de regresión SSR:	$\text{SSR} = \sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 = \widehat{\beta}_1^2 S_{xx} = \widehat{\beta}_1 S_{xy}$

$$\widehat{\beta}_1 \text{ puede ser expresado así: } \widehat{\beta}_1 = S_{xy}/S_{xx} \tag{11}$$

Contenido

Estimación por máxima verosimilitud (MLE)

Definición 6.1 (Función de verosimilitud)

Dada una muestra observada $\mathcal{M}_n$ de tamaño n y una distribución de probabilidad $F(\cdot,\theta)$, siendo θ un parámetro o un vector de parámetros, la función de verosimilitud $L(\theta|\mathcal{M}_n)$ puede concebirse como una cuantificación de la probabilidad de que la muestra proviene de la distribución F. La verosimilitud indica qué tan probable es que la muestra observada $\mathcal{M}_n$ sea una función de los posibles valores de θ.

Definición 6.2 (Estimador de máxima verosimilitud ó MLE)

Dada una muestra observada $\mathcal{M}_n$, el estimador de máxima verosimilitud de θ, parámetro o vector de parámetros de una distribución de probabilidad $F(\cdot,\theta)$, es el valor de este parámetro (o vector de parámetros) que hace máxima la probabilidad de que la muestra $\mathcal{M}_n$ haya sido generada de la distribución $F(\cdot,\theta)$.

Denotaremos por $\widetilde{\theta}$ al MLE de θ.

Estimación por máxima verosimilitud (MLE)

Definición 6.1 (Función de verosimilitud)

Dada una muestra observada $\mathcal{M}_n$ de tamaño n y una distribución de probabilidad $F(\cdot, \theta)$, siendo θ un parámetro o un vector de parámetros, la función de verosimilitud $L(\theta|\mathcal{M}_n)$ puede concebirse como una cuantificación de la probabilidad de que la muestra proviene de la distribución F. La verosimilitud indica qué tan probable es que la muestra observada $\mathcal{M}_n$ sea una función de los posibles valores de θ.

Definición 6.2 (Estimador de máxima verosimilitud ó MLE)

Dada una muestra observada $\mathcal{M}_n$, el estimador de máxima verosimilitud de θ, parámetro o vector de parámetros de una distribución de probabilidad $F(\cdot, \theta)$, es el valor de este parámetro (o vector de parámetros) que hace máxima la probabilidad de que la muestra $\mathcal{M}_n$ haya sido generada de la distribución $F(\cdot, \theta)$.

Denotaremos por $\widetilde{\theta}$ al MLE de θ.

Estimación por máxima verosimilitud (MLE)

Definición 6.1 (Función de verosimilitud)

Dada una muestra observada $\mathcal{M}_n$ de tamaño n y una distribución de probabilidad $F(\cdot,\theta)$, siendo θ un parámetro o un vector de parámetros, la función de verosimilitud $L(\theta|\mathcal{M}_n)$ puede concebirse como una cuantificación de la probabilidad de que la muestra proviene de la distribución F. La verosimilitud indica qué tan probable es que la muestra observada $\mathcal{M}_n$ sea una función de los posibles valores de θ.

Definición 6.2 (Estimador de máxima verosimilitud ó MLE)

Dada una muestra observada $\mathcal{M}_n$, el estimador de máxima verosimilitud de θ, parámetro o vector de parámetros de una distribución de probabilidad $F(\cdot,\theta)$, es el valor de este parámetro (o vector de parámetros) que hace máxima la probabilidad de que la muestra $\mathcal{M}_n$ haya sido generada de la distribución $F(\cdot,\theta)$.

Denotaremos por $\widetilde{\theta}$ al MLE de θ.

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico: $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, y $Y|x_1, \ldots, Y|x_n$ son independientes, y la función de densidad conjunta es $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.*
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud: $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,*

Entonces:

$$
\begin{aligned}
L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\}. \qquad (12)
\end{aligned}
$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico: $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, y $Y|x_1, \ldots, Y|x_n$ son independientes, y la función de densidad conjunta es $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.*
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud: $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,*

Entonces:

$$L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\}. \qquad (12)$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico: $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, y $Y|x_1, \ldots, Y|x_n$ son independientes, y la función de densidad conjunta es $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.*
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud: $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,*

Entonces:

$$L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\}. \qquad (12)$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico:* $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, *y* $Y|x_1, \ldots, Y|x_n$ *son independientes, y la función de densidad conjunta es* $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud:* $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,

Entonces:

$$
\begin{aligned}
L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\}. \qquad (12)
\end{aligned}
$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico:* $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, *y* $Y|x_1, \ldots, Y|x_n$ *son independientes, y la función de densidad conjunta es* $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud:* $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,

Entonces:

$$
\begin{aligned}
L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\}. \qquad (12)
\end{aligned}
$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico: $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, y $Y|x_1, \ldots, Y|x_n$ son independientes, y la función de densidad conjunta es $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.*
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud: $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,*

Entonces:

$$\begin{aligned} L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\} \\ &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\}. \end{aligned} \tag{12}$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

Verosimilitud en el MRLS

- $\mathcal{M}_n = (\mathbf{x}, \mathbf{y})$: los pares de puntos observados $(x_1, y_1), \ldots, (x_n, y_n)$.
- *Modelo probabilístico:* $Y|x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, *y* $Y|x_1, \ldots, Y|x_n$ *son independientes, y la función de densidad conjunta es* $f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$.
- Vector de parámetros desconocidos: $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$.
- *La función de verosimilitud:* $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) = f\left(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2\right)$,

Entonces:

$$\begin{aligned} L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\} \\ &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\}. \qquad (12) \end{aligned}$$

Objetivo: Hallar β_0, β_1 y σ^2 que maximicen a $L\left(\beta_0, \beta_1, \sigma^2 | \mathbf{x}, \mathbf{y}\right)$, o equivalentemente a $\log L$ su logaritmo natural (log-versosimilitud).

El MLE de $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$

Con n pares (x_i, y_i) de observaciones: hallar β_0, β_1 y σ^2 que maximicen al log-versosimilitud:

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}_{S(\beta_0,\beta_1)=\sum_{i=1}^{n} E_i^2}. \tag{13}$$

Sistema de ecuaciones

$$\frac{\partial \log L}{\partial \beta_0} = 0, \quad \frac{\partial \log L}{\partial \beta_1} = 0, \quad \frac{\partial \log L}{\partial \sigma^2} = 0:$$

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \tag{14}$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \tag{15}$$

$$\sigma^2 = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2 / n \tag{16}$$

Estimadores resultantes

Intercepto: $\widetilde{\beta}_0 = \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x},$ (17)

Pendiente: $\widetilde{\beta}_1 = \widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (18)

Varianza: $\widetilde{\sigma}^2 = \dfrac{1}{n}\underbrace{\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}_{\mathrm{SSE}=\sum_{i=1}^{n}\widehat{E}_i^2}.$ (19)

El MLE de $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$

Con n pares (x_i, y_i) de observaciones: hallar β_0, β_1 y σ^2 que maximicen al log-versosimilitud:

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}_{S(\beta_0,\beta_1)=\sum_{i=1}^{n} E_i^2}. \quad (13)$$

Sistema de ecuaciones

$$\frac{\partial \log L}{\partial \beta_0} = 0, \quad \frac{\partial \log L}{\partial \beta_1} = 0, \quad \frac{\partial \log L}{\partial \sigma^2} = 0:$$

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \quad (14)$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \quad (15)$$

$$\sigma^2 = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2 / n \quad (16)$$

Estimadores resultantes

Intercepto: $\widetilde{\beta}_0 = \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x},$ (17)

Pendiente: $\widetilde{\beta}_1 = \widehat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (18)

Varianza: $\widetilde{\sigma}^2 = \dfrac{1}{n}\underbrace{\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}_{\mathrm{SSE}=\sum_{i=1}^{n}\widehat{E}_i^2}.$ (19)

El MLE de $\theta = \left(\beta_0, \beta_1, \sigma^2\right)$

Con n pares (x_i, y_i) de observaciones: hallar β_0, β_1 y σ^2 que maximicen al log-versosimilitud:

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}_{S(\beta_0,\beta_1)=\sum_{i=1}^{n} E_i^2}. \qquad (13)$$

Sistema de ecuaciones

$$\frac{\partial \log L}{\partial \beta_0} = 0, \quad \frac{\partial \log L}{\partial \beta_1} = 0, \quad \frac{\partial \log L}{\partial \sigma^2} = 0:$$

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \qquad (14)$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \qquad (15)$$

$$\sigma^2 = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2/n \qquad (16)$$

Estimadores resultantes

Intercepto: $$\widetilde{\beta}_0 = \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}, \qquad (17)$$

Pendiente: $$\widetilde{\beta}_1 = \widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (18)$$

Varianza: $$\widetilde{\sigma}^2 = \frac{1}{n}\underbrace{\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}_{\text{SSE}=\sum_{i=1}^{n}\widehat{E}_i^2}. \qquad (19)$$

Estimadores MLE vs. estimadores MCO

- Los estimadores MLE: $\widetilde{\beta}_0$ y $\widetilde{\beta}_1$ son iguales, respectivamente, a los estimadores MCO: $\widehat{\beta}_0$ y $\widehat{\beta}_1$, sólo si se cumple los supuestos sobre los errores de ajuste: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$.
- *La diferencia entre los dos métodos se da en la estimación de la varianza σ^2 (ver más adelante).*
- Cuando alguno de los supuestos sobre los errores no es válido, hay que buscar soluciones alternativas: Transformaciones, modelos lineales generalizados, métodos no paramétricos, entre otros.

Estimadores MLE vs. estimadores MCO

- Los estimadores MLE: $\widetilde{\beta}_0$ y $\widetilde{\beta}_1$ son iguales, respectivamente, a los estimadores MCO: $\widehat{\beta}_0$ y $\widehat{\beta}_1$, sólo si se cumple los supuestos sobre los errores de ajuste: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$.
- *La diferencia entre los dos métodos se da en la estimación de la varianza σ^2 (ver más adelante).*
- Cuando alguno de los supuestos sobre los errores no es válido, hay que buscar soluciones alternativas: Transformaciones, modelos lineales generalizados, métodos no paramétricos, entre otros.

Estimadores MLE vs. estimadores MCO

- Los estimadores MLE: $\widetilde{\beta}_0$ y $\widetilde{\beta}_1$ son iguales, respectivamente, a los estimadores MCO: $\widehat{\beta}_0$ y $\widehat{\beta}_1$, sólo si se cumple los supuestos sobre los errores de ajuste: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$.
- *La diferencia entre los dos métodos se da en la estimación de la varianza σ^2 (ver más adelante).*
- Cuando alguno de los supuestos sobre los errores no es válido, hay que buscar soluciones alternativas: Transformaciones, modelos lineales generalizados, métodos no paramétricos, entre otros.

Contenido

Estimación de σ^2

Sean $\widehat{E}_i = y_i - \widehat{y}_i$ los residuos del ajuste del MRLS con base en una muestra de tamaño n. Sea tambien la suma de cuadrados de los residuos de ajuste, SSE $= \sum_{i=1}^{n} \widehat{E}_i^2$. Bajo supuestos: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$,

Tabla 4: Comparación estimadores de la varianza

Estimador	Ecuación	Valor Esperado
Insesgado: MSE	$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-2}$	$\text{E}\left[\widehat{\sigma}^2\right] = \sigma^2$
MLE	$\widetilde{\sigma}^2 = \frac{1}{n}\text{SSE}$	$\text{E}\left[\widetilde{\sigma}^2\right] = \left(\frac{n-2}{n}\right)\sigma^2$

- Note que $\widetilde{\sigma}^2 = \left(\frac{n-2}{n}\right)\widehat{\sigma}^2$.
- $\widehat{\sigma}^2$ es insesgado mientras que $\widetilde{\sigma}^2$ es sesgado con respecto a σ^2, sin embargo, este último es asintóticamente insesgado: $\lim_{n\to\infty} \text{E}\left[\widetilde{\sigma}^2\right] = \sigma^2$.
- El estimador MLE es de mínima varianza y consistente.

Estimación de σ^2

Sean $\widehat{E}_i = y_i - \widehat{y}_i$ los residuos del ajuste del MRLS con base en una muestra de tamaño n. Sea tambien la suma de cuadrados de los residuos de ajuste, SSE $= \sum_{i=1}^{n} \widehat{E}_i^2$. Bajo supuestos: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$,

Tabla 4: Comparación estimadores de la varianza

Estimador	Ecuación	Valor Esperado
Insesgado: MSE	$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-2}$	$\text{E}\left[\widehat{\sigma}^2\right] = \sigma^2$
MLE	$\widetilde{\sigma}^2 = \frac{1}{n}\text{SSE}$	$\text{E}\left[\widetilde{\sigma}^2\right] = \left(\frac{n-2}{n}\right)\sigma^2$

Estimación de σ^2

Sean $\widehat{E}_i = y_i - \widehat{y}_i$ los residuos del ajuste del MRLS con base en una muestra de tamaño n. Sea tambien la suma de cuadrados de los residuos de ajuste, SSE $= \sum_{i=1}^{n} \widehat{E}_i^2$. Bajo supuestos: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$,

Tabla 4: Comparación estimadores de la varianza

Estimador	Ecuación	Valor Esperado
Insesgado: MSE	$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-2}$	$\text{E}\left[\widehat{\sigma}^2\right] = \sigma^2$
MLE	$\widetilde{\sigma}^2 = \frac{1}{n}\text{SSE}$	$\text{E}\left[\widetilde{\sigma}^2\right] = \left(\frac{n-2}{n}\right)\sigma^2$

- Note que $\widetilde{\sigma}^2 = \left(\frac{n-2}{n}\right)\widehat{\sigma}^2$.
- $\widehat{\sigma}^2$ es insesgado mientras que $\widetilde{\sigma}^2$ es sesgado con respecto a σ^2, sin embargo, este último es asintóticamente insesgado: $\lim_{n\to\infty} \text{E}\left[\widetilde{\sigma}^2\right] = \sigma^2$.
- El estimador MLE es de mínima varianza y consistente: $\widetilde{\sigma}^2 \xrightarrow{\mathcal{P}} \sigma^2$.

Estimación de σ^2

Sean $\widehat{E}_i = y_i - \widehat{y}_i$ los residuos del ajuste del MRLS con base en una muestra de tamaño n. Sea tambien la suma de cuadrados de los residuos de ajuste, SSE $= \sum_{i=1}^{n} \widehat{E}_i^2$. Bajo supuestos: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$,

Tabla 4: Comparación estimadores de la varianza

Estimador	Ecuación	Valor Esperado
Insesgado: MSE	$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-2}$	$\text{E}\left[\widehat{\sigma}^2\right] = \sigma^2$
MLE	$\widetilde{\sigma}^2 = \frac{1}{n}\text{SSE}$	$\text{E}\left[\widetilde{\sigma}^2\right] = \left(\frac{n-2}{n}\right)\sigma^2$

- Note que $\widetilde{\sigma}^2 = \left(\frac{n-2}{n}\right)\widehat{\sigma}^2$.
- $\widehat{\sigma}^2$ es insesgado mientras que $\widetilde{\sigma}^2$ es sesgado con respecto a σ^2, sin embargo, este último es asintóticamente insesgado: $\lim_{n\to\infty} \text{E}\left[\widetilde{\sigma}^2\right] = \sigma^2$.
- El estimador MLE es de mínima varianza y consistente: $\widetilde{\sigma}^2 \xrightarrow{\mathcal{P}} \sigma^2$.

Estimación de σ^2

Sean $\widehat{E}_i = y_i - \widehat{y}_i$ los residuos del ajuste del MRLS con base en una muestra de tamaño n. Sea tambien la suma de cuadrados de los residuos de ajuste, $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$. Bajo supuestos: $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$,

Tabla 4: Comparación estimadores de la varianza

Estimador	Ecuación	Valor Esperado
Insesgado: MSE	$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-2}$	$\text{E}\left[\widehat{\sigma}^2\right] = \sigma^2$
MLE	$\widetilde{\sigma}^2 = \frac{1}{n}\text{SSE}$	$\text{E}\left[\widetilde{\sigma}^2\right] = \left(\frac{n-2}{n}\right)\sigma^2$

- Note que $\widetilde{\sigma}^2 = \left(\frac{n-2}{n}\right)\widehat{\sigma}^2$.
- $\widehat{\sigma}^2$ es insesgado mientras que $\widetilde{\sigma}^2$ es sesgado con respecto a σ^2, sin embargo, este último es asintóticamente insesgado: $\lim_{n\to\infty} \text{E}\left[\widetilde{\sigma}^2\right] = \sigma^2$.
- El estimador MLE es de mínima varianza y consistente: $\widetilde{\sigma}^2 \xrightarrow{\mathcal{P}} \sigma^2$.

Contenido

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\text{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionadas.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionadas.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: $\text{SST} = \text{SSR} + \text{SSE}$, la veremos más adelante como Análisis de varianza o ANOVA.

$$\text{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \quad (20)$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionadas.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: $\text{SST} = \text{SSR} + \text{SSE}$, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: $\text{SST} = \text{SSR} + \text{SSE}$, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n}\widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2$

La descomposición: $\text{SST} = \text{SSR} + \text{SSE}$, la veremos más adelante como Análisis de varianza o ANOVA.

$$\text{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionadas.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\mathrm{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$
- Suma de cuadrados de residuos: $\mathrm{SSE} = \sum_{i=1}^{n} \hat{E}_i^2$
- Suma de cuadrados de regresión: $\mathrm{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$

La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionados.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\text{SSE} = \sum_{i=1}^{n} \widehat{E}_i^2$
- Suma de cuadrados de regresión: $\text{SSR} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\text{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionados.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionados.*

Coeficiente de determinación ó R^2 de una regresión

Considere las siguientes sumas de cuadrados:

- Suma de cuadrados totales (vista previamente como S_{yy}): $\mathrm{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$
- Suma de cuadrados de residuos: $\mathrm{SSE} = \sum_{i=1}^{n} \hat{E}_i^2$
- Suma de cuadrados de regresión: $\mathrm{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

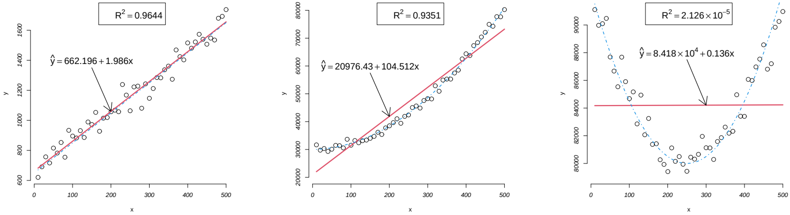
La descomposición: SST = SSR + SSE, la veremos más adelante como Análisis de varianza o ANOVA.

$$\mathrm{R}^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{SST} \tag{20}$$

Interpretación: *La proporción de la variabilidad total observada en la variable respuesta, que es explicada por la relación lineal con la variable predictora considerada.*

Interpretaciones erróneas:

- *Un R^2 alto indica que el modelo puede hacer predicciones útiles.*
- *Un R^2 alto indica que la recta de regresión tiene buen ajuste.*
- *Un R^2 cercano a cero indica que X y Y no están relacionados.*



(a) (b) (c)

Figura 5: Datos simulados.(a) la verdadera relación estadística es lineal: $\mu_{Y|x} = 650 + 2x$ y el ajuste RLS arroja R^2 cercano a 1; (b) La verdadera relación no es lineal: $\mu_{Y|x} = 30000 + 2x + 0.2x^2$ aunque el ajuste por RLS arroja R^2 cercano a 1; (c) La verdadera relación no es lineal: $\mu_{Y|x} = 92500 - 100x + 0.2x^2$ y el ajuste RLS produce un R^2 de casi cero, sin embargo, en este caso, no se puede decir que no existe relación estadística entre X y Y sino que la relación es no lineal.

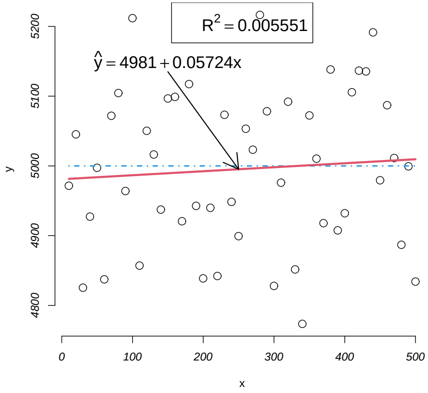


Figura 6: Datos simulados. El verdadero modelo es $Y_i = 5000 + E_i$, $E_i \overset{\text{iid}}{\sim} N(0, 10000)$, es decir, no hay ningún tipo de asociación estadística de Y con X, sin embargo. se ajustó modelo de RLS asumiendo que $\mu_{Y|x} = \beta_0 + \beta_1 x$, y su ajuste da un R^2 pequeño, como era de esperarse bajo estas circuntancias. El modelo $Y_i = \beta_0 + E_i$, $E_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$ da como estimación $\widehat{\beta}_0 = 4995.508$ muy próximo a la media verdadera de Y.