INTERPRETACIÓN DE LA REGRESIÓN

Este gráfico muestra el salario por hora de 570 individuos.

Interpretación de la regresión

```
. regresión Salario-Estudios

Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =   65.64
Model |  3977.38016     1  3977.38016               Prob > F      =  0.0000
Residual |  34419.6569   568  60.5979875               R-squared     =  0.1036
---------+------------------------------               Adj R-squared =  0.1020
Total |  38397.0371   569  67.4816117               Root MSE      =  7.7845

------------------------------------------------------------------------------
SALARIO  |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
S       |   1.073055   .1324501      8.102   0.000       .8129028    1.333206
_cons   |  -1.391004   1.820305     -0.764   0.445      -4.966354    2.184347
------------------------------------------------------------------------------
```

Esta es una salida de ordenador de un programa econométrico típico.
¿Qué significan los coeficientes?

$\hat{Salario} = -1.39 + 1.073S$

¿Qué mide la pendiente?
¿Qué significa el término constante?
Ajuste cuadrático

**BONDAD DE AJUSTE**

Tres resultados relevantes:

\[ \bar{e} = 0 \quad \bar{Y} = \bar{Y} \quad \text{Cov}(\hat{Y}, e) = 0 \]
Tres resultados relevantes:

\[ \bar{e} = 0 \quad \bar{Y} = \bar{Y} \quad \text{Cov}(\hat{Y}, e) = 0 \]

\[ e_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_i \]

\[ \sum e_i = \sum Y_i - nb_1 - b_2 \sum X_i \]

\[ \frac{1}{n} \sum e_i = \frac{1}{n} \sum Y_i - b_1 - b_2 \frac{1}{n} \sum X_i \]

\[ \bar{e} = \bar{Y} - b_1 - b_2 \bar{X} \quad b_1 = \bar{Y} - b_2 \bar{X} \]

\[ = \bar{Y} - (\bar{Y} - b_2 \bar{X}) - b_2 \bar{X} = 0 \]

---

Tres resultados relevantes:

\[ \bar{e} = 0 \quad \bar{Y} = \bar{Y} \quad \text{Cov}(\hat{Y}, e) = 0 \]

\[ e_i = Y_i - \hat{Y}_i \]

\[ \sum e_i = \sum Y_i - \sum \hat{Y}_i \]

\[ \frac{1}{n} \sum e_i = \frac{1}{n} \sum Y_i - \frac{1}{n} \sum \hat{Y}_i \]

\[ \bar{e} = \bar{Y} - \bar{\hat{Y}} \quad \bar{\hat{Y}} = \bar{Y} \]
**Bondad de ajuste**

Tres resultados relevantes:

\[
\bar{e} = 0 \quad \bar{Y} = \bar{Y} \quad \text{Cov}(\hat{Y}, e) = 0
\]

\[
\text{Cov}(\hat{Y}, e) = \text{Cov}([b_1 + b_2 X], e) = \text{Cov}(b_1, e) + \text{Cov}(b_2 X, e)
\]

\[
= 0 + b_2 \text{Cov}(X, e) = b_2 \text{Cov}(X, [Y - b_1 - b_2 X])
\]

**Demostrad que es igual a 0**

**Bondad de ajuste**

\[
e_i = Y_i - \hat{Y}_i \quad \Rightarrow \quad Y_i = \hat{Y}_i + e_i
\]

Para analizar la bondad del ajuste, descomponemos el valor observado en el valor ajustado y el residuo.
Bondad de ajuste

\[ e_i = Y_i - \hat{Y}_i \quad \Rightarrow \quad Y_i = \hat{Y}_i + e_i \]

\[ \text{Var}(Y) = \text{Var}(\hat{Y} + e) = \text{Var}(\hat{Y}) + \text{Var}(e) + 2\text{Cov}(\hat{Y}, e) = \text{Var}(\hat{Y}) + \text{Var}(e) \]

\[ \frac{1}{n} \sum (Y - \bar{Y})^2 = \frac{1}{n} \sum (\hat{Y} - \bar{Y})^2 + \frac{1}{n} \sum (e - \bar{e})^2 \]

\[ \sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum e^2 \]

\[ S\text{CT} = S\text{CE} + S\text{CR} \]

\[ R^2 = \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)} = \frac{S\text{CE}}{S\text{CT}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2} \]

Un criterio de bondad de ajuste de ajuste es el coeficiente de determinación. Además, \( 0 \leq R^2 \leq 1 \).

Bondad de ajuste

\[ r_{Y,\hat{Y}} = \frac{\text{Cov}(Y, \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} = \frac{\text{Cov}([\hat{Y} + e], \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} \]

\[ = \frac{\text{Cov}(\hat{Y}, \hat{Y}) + \text{Cov}(e, \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} = \frac{\text{Var}(\hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} \]

\[ = \frac{\sqrt{\text{Var}(\hat{Y}) \cdot \text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} = \sqrt{R^2} \]

Otro criterio de bondad de ajuste es la correlación entre el valor observado y ajustado de la variable \( Y \).
Los estimadores de los coeficientes de regresión son un tipo particular de variable aleatoria: recordar la definición de un estimador. Para analizarlo, visualizaremos cómo se obtienen los estimadores en el caso de una regresión simple.

Estimadores de los Coeficientes de Regresión: variables aleatorias

\[ Y = \beta_1 + \beta_2 X + u \]
\[ \hat{Y} = b_1 + b_2 X \]

\[
b_2 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\text{Cov}(X,[\beta_1 + \beta_2 X + u])}{\text{Var}(X)}
= \frac{\text{Cov}(X,\beta_1) + \text{Cov}(X,\beta_2 X) + \text{Cov}(X,u)}{\text{Var}(X)}
= \frac{0 + \beta_2 \text{Cov}(X,X) + \text{Cov}(X,u)}{\text{Var}(X)}
= \beta_2 + \frac{\text{Cov}(X,u)}{\text{Var}(X)}
\]

Por tanto, hemos descompuesto \( b_2 \) en dos componentes: el verdadero valor del parámetro, \( \beta_2 \), y el término de error.
Un experimento Monte Carlo es un ejercicio de laboratorio, basado en la utilización de ordenadores, cuyo objetivo es evaluar las propiedades de un estimador en situaciones controladas.

El experimento empieza eligiendo los valores de $X$, los parámetros y la distribución de la perturbación. A partir de aquí se genera el valor de $Y$. Una vez que tengo $(X,Y)$, aplica el método de MCO y se obtiene una estimación. Este proceso se repite varias veces.
Elegir los datos, \( X \)

Elegir parámetros

Elegir una distribución para \( u \)

Generar valores de \( Y \)

Estimador

Estimación

Y = \( \beta_1 + \beta_2 X + u \)

\( X = 1, 2, \ldots, 20 \)

\( \beta_1 = 2.0 \)

\( \beta_2 = 0.5 \)

\( u \) es i.i.d. \( N(0,1) \)

\[ Y = 2.0 + 0.5X + u \]

Generar valores de \( Y \)

\[ b_2 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}; \quad b_1 = \bar{Y} - b_2 \bar{X} \]

Estimar el valor de los parámetros

Se regresará \( Y \) sobre \( X \) usando MCO y, a partir de los estimadores \( b_1 \) y \( b_2 \), se obtendrán estimaciones para los verdaderos valores de \( \beta_1 \) y \( \beta_2 \).

Observad lo que es el componente no estocástico.
Estimadores de los Coeficientes de Regresión: variables aleatorias

\[ Y = 2.0 + 0.5X + u \]

<table>
<thead>
<tr>
<th>( X )</th>
<th>2.0+0.5X</th>
<th>( u )</th>
<th>( Y )</th>
<th>( X )</th>
<th>2.0+0.5X</th>
<th>( u )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5</td>
<td>-0.59</td>
<td>1.91</td>
<td>11</td>
<td>7.5</td>
<td>1.59</td>
<td>9.09</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>-0.24</td>
<td>2.76</td>
<td>12</td>
<td>8.0</td>
<td>-0.92</td>
<td>7.08</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>-0.83</td>
<td>2.67</td>
<td>13</td>
<td>8.5</td>
<td>-0.71</td>
<td>7.79</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>0.03</td>
<td>4.03</td>
<td>14</td>
<td>9.0</td>
<td>-0.25</td>
<td>8.75</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>-0.38</td>
<td>4.12</td>
<td>15</td>
<td>9.5</td>
<td>1.69</td>
<td>11.19</td>
</tr>
<tr>
<td>6</td>
<td>5.0</td>
<td>-2.19</td>
<td>2.81</td>
<td>16</td>
<td>10.0</td>
<td>0.15</td>
<td>10.15</td>
</tr>
<tr>
<td>7</td>
<td>5.5</td>
<td>1.03</td>
<td>6.53</td>
<td>17</td>
<td>10.5</td>
<td>0.02</td>
<td>10.52</td>
</tr>
<tr>
<td>8</td>
<td>6.0</td>
<td>0.24</td>
<td>6.24</td>
<td>18</td>
<td>11.0</td>
<td>-0.11</td>
<td>10.89</td>
</tr>
<tr>
<td>9</td>
<td>6.5</td>
<td>2.53</td>
<td>9.03</td>
<td>19</td>
<td>11.5</td>
<td>-0.91</td>
<td>10.59</td>
</tr>
<tr>
<td>10</td>
<td>7.0</td>
<td>-0.13</td>
<td>6.87</td>
<td>20</td>
<td>12.0</td>
<td>1.42</td>
<td>13.42</td>
</tr>
</tbody>
</table>

A partir de aquí obtenemos el valor de la variable dependiente.

Estimadores de los Coeficientes de Regresión: variables aleatorias

![Diagrama de dispersión](image)

Observad que gráficamente las observaciones ya no están en la recta determinista.
Estimadores de los Coeficientes de Regresión: variables aleatorias

\[ \hat{Y} = 1.63 + 0.54X \]

De donde obtenemos la recta de regresión.

Estimadores de los Coeficientes de Regresión: variables aleatorias

\[ \hat{Y} = 1.63 + 0.54X \]

Para comparar, podemos graficar la parte no estocástica verdadera, que es la que surge de la definición del modelo. \( \beta_2 \) (tiene como verdadero valor 0.50) y ha sido sobreestimado mientras que \( \beta_1 \) (con un valor verdadero de 2.00) ha sido subestimada.
El experimento Monte Carlo consiste en repetir el anterior proceso un número elevado de veces, estimar los parámetros y ver cómo se comportan. Vamos a repetirlo. La verdadera recta es la que se observa en el gráfico, que está dada por la definición del experimento.

A los valores no estocásticos les sumamos un término de perturbación y obtenemos los datos que observamos en la realidad.
Nuevamente, ajustamos por MCO y obtenemos la recta de regresión, que nunca coincidirá con la verdadera.

<table>
<thead>
<tr>
<th>replicación</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.63</td>
<td>0.54</td>
</tr>
<tr>
<td>2</td>
<td>2.52</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>2.13</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>2.14</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>1.71</td>
<td>0.56</td>
</tr>
<tr>
<td>6</td>
<td>1.81</td>
<td>0.51</td>
</tr>
<tr>
<td>7</td>
<td>1.72</td>
<td>0.56</td>
</tr>
<tr>
<td>8</td>
<td>3.18</td>
<td>0.41</td>
</tr>
<tr>
<td>9</td>
<td>1.26</td>
<td>0.58</td>
</tr>
<tr>
<td>10</td>
<td>1.94</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Esta tabla resume los valores de las estimaciones si repetimos 10 veces el mismo experimento.
Estimadores de los Coeficientes de Regresión: variables aleatorias

Observad el histograma para $b_2$

<table>
<thead>
<tr>
<th>1-10</th>
<th>11-20</th>
<th>21-30</th>
<th>31-40</th>
<th>41-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54</td>
<td>0.49</td>
<td>0.54</td>
<td>0.52</td>
<td>0.49</td>
</tr>
<tr>
<td>0.48</td>
<td>0.54</td>
<td>0.46</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td>0.45</td>
<td>0.49</td>
<td>0.45</td>
<td>0.54</td>
<td>0.48</td>
</tr>
<tr>
<td>0.50</td>
<td>0.54</td>
<td>0.50</td>
<td>0.53</td>
<td>0.44</td>
</tr>
<tr>
<td>0.56</td>
<td>0.54</td>
<td>0.41</td>
<td>0.51</td>
<td>0.53</td>
</tr>
<tr>
<td>0.51</td>
<td>0.52</td>
<td>0.53</td>
<td>0.51</td>
<td>0.48</td>
</tr>
<tr>
<td>0.56</td>
<td>0.49</td>
<td>0.53</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>0.41</td>
<td>0.53</td>
<td>0.47</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>0.58</td>
<td>0.60</td>
<td>0.51</td>
<td>0.51</td>
<td>0.53</td>
</tr>
<tr>
<td>0.52</td>
<td>0.48</td>
<td>0.47</td>
<td>0.58</td>
<td>0.51</td>
</tr>
</tbody>
</table>

En el caso de 50 replicaciones, estas serían las estimaciones de $\beta_2$.
Estimadores de los Coeficientes de Regresión: variables aleatorias

El histograma ahora sería.

Estimadores de los Coeficientes de Regresión: variables aleatorias

La línea roja muestra la distribución límite, la que se obtendría si hiciéramos muchísimas replicaciones del experimento. Observad que la distribución es simétrica en el verdadero valor del parámetro, confirmando que el estimador es insesgado.
Regresión simple: $Y = \beta_1 + \beta_2 X + u$

Condiciones de Gauss-Markov

1. $E(u_i) = 0$

   Suponga $E(u_i) = \mu_u \neq 0$.

   Definimos $v = u - \mu_u$, entonces $u = v + \mu_u$

   Entonces $Y = \beta_1 + \beta_2 X + v + \mu_u$

   $= (\beta_1 + \mu_u) + \beta_2 X + v$

   donde $E(v) = E(u - \mu_u) = E(u) - E(\mu_u) = 0$

Esta condición está relacionada con la perturbación, $u$. Este supuesto sostiene que el valor esperado de la perturbación es cero, por lo que, en media, no afecta al valor de la variable dependiente: no tiene una tendencia sistemática en ninguna dirección (positiva o negativa).

Nótese que el término constante normalmente recoge cualquier tendencia de $Y$ no tomada en cuenta por las variables explicativas.

Las condiciones de Gauss-Markov y la inesgadez de los estimadores

Regresión simple: $Y = \beta_1 + \beta_2 X + u$

Condiciones de Gauss-Markov

2. La varianza poblacional de $u_i$ es la misma para todo $i$

   $\sigma^2_{u_i} = \sigma^2_u$ para todo $i$.

La segunda condición es que los valores del término de perturbación en las diferentes observaciones son extraídos de una distribución con varianza constante: Homocedasticidad.
Las condiciones de Gauss-Markov y la inesgadez de los estimadores

Regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Condiciones de Gauss-Markov

3. La covarianza poblacional entre \( u_i \) y \( u_j \) es igual a 0, \( i \) distinto de \( j \)

\[ \sigma_{u_i u_j} = 0 \text{ para todo } i \neq j. \]

La tercera condición sostiene que el valor del término de perturbación para una observación no podrá co-variar con ninguna de las otras observaciones: las \( u \) se distribuyen independientemente: Incorrelación.

Las condiciones de Gauss-Markov y la inesgadez de los estimadores

Regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Condiciones de Gauss-Markov

4. \( X \) no estocástico

Esta condición final puede verse en dos versiones: la fuerte, que supone que las variables explicativas son no estocásticas. La débil: son aleatorias pero se distribuyen de forma independiente al término de perturbación.

Utilizaremos el supuesto fuerte pues facilita el análisis de la propiedad de los estimadores.
Regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Condiciones de Gauss-Markov

4. \( X \) no estocástica

Este ejemplo, de muestra estratificada, es un caso de variables no estocásticas.

<table>
<thead>
<tr>
<th>( S )</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td>430</td>
</tr>
<tr>
<td>13</td>
<td>100 , etc</td>
</tr>
</tbody>
</table>

Supongamos que del censo nacional se sabe que el 1% de la población tiene \( S = 8 \), el 3% tiene \( S = 9 \), el 5% tiene 10, el 7% \( S = 11 \), el 43% \( S = 12 \), etc.

Las condiciones de Gauss-Markov y la insesgadez de los estimadores

Regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Supuesto de Normalidad

5. \( u \) tiene una distribución normal.

Además de las condiciones de Gauss-Markov, generalmente se supone que la perturbación tiene distribución normal. Este supuesto permite derivar, de forma sencilla, la distribución de los estadísticos.

Justificación: Teorema Central del Límite.
Las condiciones de Gauss-Markov y la insesgadez de los estimadores

Regresión simple: \( Y = \beta_1 + \beta_2X + u \)

Insesgadez

\[
b_2 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\text{Cov}(X,\beta_1 + \beta_2X + u)}{\text{Var}(X)}
\]

\[
= \frac{\text{Cov}(X,\beta_1) + \text{Cov}(X,\beta_2X) + \text{Cov}(X,u)}{\text{Var}(X)}
\]

\[
= \frac{0 + \beta_2\text{Cov}(X,X) + \text{Cov}(X,u)}{\text{Var}(X)}
\]

\[
= \beta_2 + \frac{\text{Cov}(X,u)}{\text{Var}(X)}
\]

Por lo tanto, llegamos a que el estimador se descompone en el verdadero valor del parámetro y un término de error que depende de la covarianza entre las variables explicativas y el término de perturbación.

EJERCICIO: Demostrar que \( E(\text{Cov}(X,u)) = 0 \).

Para ello, utilizar la definición de covarianzas
Las condiciones de Gauss-Markov y la insegadez de los estimadores

Regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Insegadez

\[ b_1 = \bar{Y} - b_2 \bar{X} \]

**EJERCICIO**

Demostrar que el estimador \( b_1 \) es insegado con respecto a \( \beta_1 \).

---

**ANÁLISIS DE LA PRECISIÓN DE LA REGRESIÓN**

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Varianza de los estimadores

![distribución de \( b_2 \)](image)

Hemos visto que los estimadores \( b_1 \) y \( b_2 \) son variables aleatorias que generan estimaciones puntuales de \( \beta_1 \) y \( \beta_2 \). Sabemos, además, que dichos estimadores son insegados: la media de la densidad del estimador coincide con el verdadero valor del parámetro, como se observa en el gráfico.
Modelo de regresión simple: $Y = \beta_1 + \beta_2 X + u$

Varianza de los estimadores

La desviación típica de un estimador mide la precisión del mismo: “lo cerca que estoy del verdadero valor del parámetro”, si interpreto la desviación típica como una distancia. Dada esta interpretación, la desviación típica permite realizar contrastes relacionados con posibles valores del verdadero valor del parámetro.

Análisis de la precisión de la regresión

Modelo de regresión simple: $Y = \beta_1 + \beta_2 X + u$

Varianza de los estimadores

Varianza poblacional de $b_1 = \sigma^2_{b_1} = \frac{\sigma^2_u}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\}$

Varianza poblacional de $b_2 = \sigma^2_{b_2} = \frac{\sigma^2_u}{n \text{Var}(X)}$

Inversamente con $n$
Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Varianza de los estimadores

Varianza poblacional

\[
b_1 = \sigma_{b_1}^2 = \frac{\sigma_u^2}{n} \left( 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right)
\]

Varianza poblacional

\[
b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n \text{Var}(X)}
\]

Proporcional a la varianza de la perturbación

Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Varianza de los estimadores

\[
Y = 3.0 + 0.8X
\]

Observar cómo cambia el ajuste cuando cambia la varianza de las observaciones de la variable dependiente. La gráfica punteada es la verdadera recta de regresión, \( Y = 3.0 + 0.8X \).
Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Varianza de los estimadores

Varianza poblacional de \( b_1 = \sigma^2_{b_1} = \frac{\sigma^2_u}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\} \)

Varianza poblacional de \( b_2 = \sigma^2_{b_2} = \frac{\sigma_u^2}{n \text{Var}(X)} \)

Inversamente proporcional a la varianza de \( X \)

\( Y = 3.0 + 0.8X \)
Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Varianza de los estimadores

\[
\sigma^2_{b_1} = \frac{\sigma^2_u}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\} \\
\sigma^2_{b_2} = \frac{\sigma^2_u}{n \text{Var}(X)}
\]

\[
s^2_u = \frac{n}{n-k} \text{Var}(e) = \frac{n}{n-2} \text{Var}(e)
\]

\[
s.e.(b_1) = \sqrt{\frac{s^2_u}{n} \left\{ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right\}} \\
s.e.(b_2) = \sqrt{\frac{s^2_u}{n \text{Var}(X)}}
\]

También podemos estimar las desviaciones típicas de los estimadores (las denotamos por s.e., del inglés standard error)

Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X = u \)

Varianza de los estimadores

```
. reg EARNINGS S

Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =   65.64
Model |  3977.38016     1  3977.38016               Prob > F      =  0.0000
Residual |  34419.6569   568  60.5979875               R-squared     =  0.1036
---------+------------------------------               Adj R-squared =  0.1020
Total |  38397.0371   569  67.4816117               Root MSE      =  7.7845

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   1.073055   .1324501      8.102   0.000       .8129028    1.333206
  _cons |  -1.391004   1.820305     -0.764   0.445      -4.966354    2.184347
------------------------------------------------------------------------------
```

50
Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Eficiencia

El teorema de Gauss-Markov establece que si el modelo está bien especificado, los estimadores de MCO son los de menor varianza dentro de todos los estimadores insesgados.

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Eficiencia

\[ b_2 = \frac{Y_n - Y_1}{X_n - X_1} \]
Análisis de la precisión de la regresión

Modelo de regresión simple: \( Y = \beta_1 + \beta_2 X + u \)

Eficiencia

\[
b_2 = \frac{Y_n - Y_1}{X_n - X_1}
\]

Investigaremos las propiedades de este estimador

**EJERCICIO:** calcular la media y la varianza de este estimador

\[
\sigma^2_{b_2} = \frac{2\sigma^2_u}{(X_n - X_1)^2}
\]

\[
b_2^{OLS} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}
\]

\[
\sigma^2_{b_2^{OLS}} = \frac{\sigma^2_u}{n\text{Var}(X)}
\]

Se demuestra que la varianza del estimador MCO es menor que la del estimador alternativo.

**EJERCICIO:** realizar dicha demostración