REGRESIÓN LINEAL

Técnicas de modelamiento y simulación

REGRESIÓN

Una regresión es un análisis estadístico realizado sobre un conjunto de datos con el fin de obtener una relación funcional entre las variable involucradas. Además, este análisis permite realizar predicciones de valores que no se midieron, ya sea dentro del rango de los datos conocidos (interpolación) o fuera de ´este (extrapolación)

REGRESIÓN LINEAL

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

Donde ˆβ0 es el intercepto y ˆβ1 es la pendiente de la recta obtenida de la regresión por mínimos cuadrados.

REGRESIÓN LINEAL

$$\hat{\beta}_0 = \frac{\overline{x^2}\,\bar{y} - \overline{xy}\,\bar{x}}{\overline{x^2} - (\bar{x})^2} \qquad \hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$$\bar{y} = \frac{\sum y_i}{N} \qquad \bar{x} = \frac{\sum x_i}{N} \qquad \overline{xy} = \frac{\sum x_i y_i}{N} \qquad \overline{x^2} = \frac{\sum x_i^2}{N}$$

N es el número de datos.

REGRESIÓN LINEAL

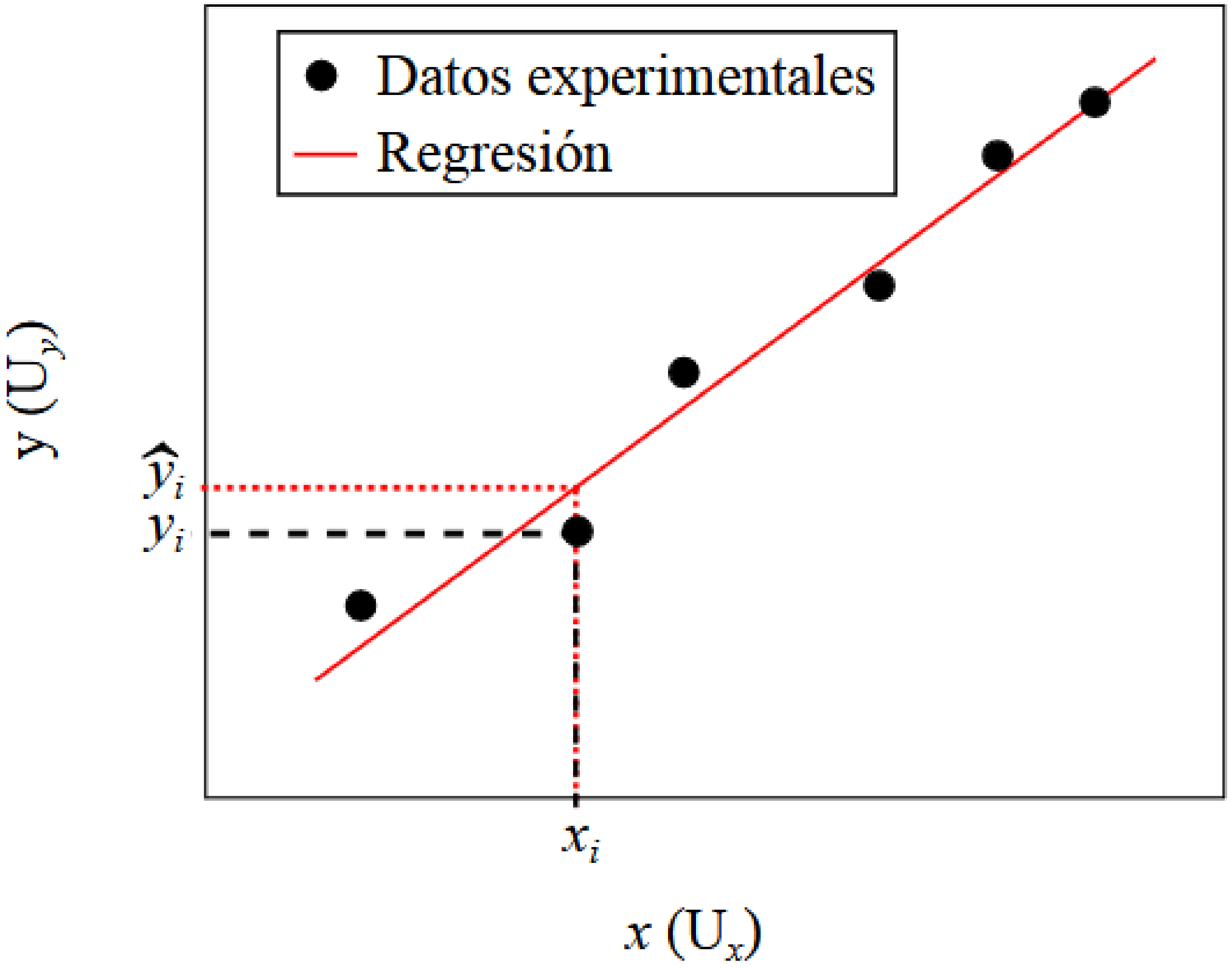
Medición del error de los coeficientes

$$\Delta\hat{\beta}_0 = \Delta\hat{\beta}_1\sqrt{\overline{x^2}}$$

$$\Delta\hat{\beta}_1 = \hat{\beta}_1\sqrt{\frac{1}{N-2}\left(\frac{1}{R^2}-1\right)}$$

$$\Delta\hat{\beta}_1 = \hat{\beta}_1\sqrt{\frac{1}{N-2}\left[\frac{\left(\overline{x^2}-\bar{x}^2\right)\left(\overline{y^2}-\bar{y}^2\right)(\overline{xy}-\bar{x}\bar{y})^2}{(\overline{xy}-\bar{x}\bar{y})^2}\right]}$$

REGRESIÓN LINEAL



REGRESIÓN LINEAL

Coeficiente de determinación (R^2). Este coeficiente es una medida de la "fuerza" o asertividad que posee el modelo planteado para predecir los resultados. En el caso particular de una línea recta, R^2 coincide con el cuadrado del coeficiente de correlación de Pearson, el cual es una medida de la relación lineal de dos variables y se representa con la letra R.

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N\sqrt{\overline{x^2} - (\bar{x})^2}\sqrt{\overline{y^2} - (\bar{y})^2}}$$

REGRESIÓN LINEAL

Error de la medida i-ésima.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\varepsilon_i = y_i - \hat{y}_i$$

REGRESIÓN LINEAL

Error total.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\sum_{i=1}^{N} \varepsilon_i = \sum_{i=1}^{N} y_i - \hat{y}_i$$

REGRESIÓN LINEAL

Cada componente del error total puede ser positivo o negativo dependiendo de la posición del punto respecto la recta. Como consecuencia de lo anterior, el error total puede se nulo o muy pequeño aun cuando los datos tienen una desviación considerable de la recta.

$$\sum_{i=1}^{N}(\varepsilon_i)^2 = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

REGRESIÓN LINEAL

Como se desea minimizar el error total, se recurre a los métodos de optimización del cálculo.

$$\sum_{i=1}^{N}(\varepsilon_i)^2 = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

yi y *xi* son medidas experimentales, como consecuencia, no se puede optimizar el error respecto de estos valores, por lo tanto, las únicas opciones son la pendiente (ˆβ1) y el intercepto (ˆβ0), los cuales son los parámetros que se desean determinar. Con estos puntos claros, las condiciones de optimización quedan:

$$\frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_0} = 0 \qquad \frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_1} = 0$$

REGRESIÓN LINEAL

$$\sum_{i=1}^{N}(\varepsilon_i)^2 = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$\frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_0} = 0 \qquad \frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_1} = 0$$

$$\frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_0} = \frac{\partial \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{\partial \hat{\beta}_0} \qquad \frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_1} = \frac{\partial \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{\partial \hat{\beta}_1}$$

REGRESIÓN LINEAL

$$\frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_0} = \frac{\partial \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{\partial \hat{\beta}_0}$$

$$-2\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\sum_{i=1}^{N} y_i = N\,\hat{\beta}_0 + \hat{\beta}_1 x_i$$

REGRESIÓN LINEAL

$$\frac{\partial \sum \varepsilon_i^2}{\partial \hat{\beta}_1} = \frac{\partial \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{\partial \hat{\beta}_1}$$

$$-2\sum_{i=1}^{N} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\sum_{i=1}^{N} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{N} x_i + \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

REGRESIÓN LINEAL

Al dividirlos sobre el número de datos N

$$\sum_{i=1}^{N} y_i = N\,\hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \sum_{i=1}^{N} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{N} x_i + \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \qquad \overline{xy} = \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{x^2}$$

REGRESIÓN LINEAL

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ $\qquad$ $\overline{xy} = \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{x^2}$

$$\hat{\beta}_1 = \frac{\bar{y} - \hat{\beta}_0}{\bar{x}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\overline{xy} = \left(\bar{y} - \hat{\beta}_1 \bar{x}\right)\bar{x} + \hat{\beta}_1 \overline{x^2}$$

$$\overline{xy} = \bar{x}\bar{y} - \hat{\beta}_1 \bar{x}^2 + \hat{\beta}_1 \overline{x^2}$$

$$\overline{xy} - \bar{x}\bar{y} = \hat{\beta}_1 \left(\overline{x^2} - \bar{x}^2\right)$$

$$\boxed{\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\left(\overline{x^2} - \bar{x}^2\right)}}$$

REGRESIÓN LINEAL

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \qquad \overline{xy} = \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{x^2}$$

$$\overline{xy} - \hat{\beta}_0 \bar{x} = \hat{\beta}_1 \overline{x^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\overline{xy} - \hat{\beta}_0 \bar{x}}{\overline{x^2}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \hat{\beta}_0 \bar{x}}{\overline{x^2}} \bar{x} \qquad \hat{\beta}_0 = \bar{y} + \frac{\hat{\beta}_0 \bar{x} - \overline{xy}}{\overline{x^2}} \bar{x} \qquad \hat{\beta}_0 = \bar{y} + \frac{\hat{\beta}_0 \bar{x}^2 - \bar{x}\,\overline{xy}}{\overline{x^2}}$$

REGRESIÓN LINEAL

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \hat{\beta}_0\bar{x}}{\overline{x^2}}\bar{x} \qquad \hat{\beta}_0 = \bar{y} + \frac{\hat{\beta}_0\bar{x} - \overline{xy}}{\overline{x^2}}\bar{x} \qquad \hat{\beta}_0 = \bar{y} + \frac{\hat{\beta}_0\bar{x}^2 - \bar{x}\,\overline{xy}}{\overline{x^2}}$$

$$\hat{\beta}_0\overline{x^2} = \bar{y}\overline{x^2} + \hat{\beta}_0\bar{x}^2 - \bar{x}\,\overline{xy}$$

$$\hat{\beta}_0\overline{x^2} - \hat{\beta}_0\bar{x}^2 = \bar{y}\overline{x^2} - \bar{x}\,\overline{xy}$$

$$\hat{\beta}_0\left(\overline{x^2} - \bar{x}^2\right) = \bar{y}\overline{x^2} - \bar{x}\,\overline{xy}$$

$$\boxed{\hat{\beta}_0 = \frac{\bar{y}\overline{x^2} - \bar{x}\,\overline{xy}}{\left(\overline{x^2} - \bar{x}^2\right)}}$$

REGRESIÓN LINEAL

Las unidades de los coeficientes son:

$$[\hat{\beta}_0] = U_y$$

$$[\hat{\beta}_1] = \frac{U_y}{U_x}$$

Con todos los parámetros calculados, la ecuación de la recta se expresa como:

$$y = (\hat{\beta}_1 \pm \Delta\hat{\beta}_1)x + (\hat{\beta}_0 \pm \Delta\hat{\beta}_0)$$

$$A = (\hat{\beta}_1 \pm \Delta\hat{\beta}_1)l + (\hat{\beta}_0 \pm \Delta\hat{\beta}_0)$$

Las ecuaciones deben ser descriptivas, por lo tanto, si la variable dependiente se representa con la letra A y la independiente con la letra l, entonces la ecuación debe escribirse en términos de A y l no y y x.

REGRESIÓN

t (s)	s (m)
1.0	10.1
2.0	14.8
3.0	21.0
4.0	25.5
5.0	31.1
6.0	34.2
7.0	41.3
8.0	43.6
9.0	52.1
10.0	55.2

REGRESIÓN

Función potencial

$$y = \alpha_0 x^{\alpha_1}$$

REGRESIÓN

Función exponencial:

$$y = \alpha_0 c^{x\alpha_1}$$

REGRESIÓN

Distancia ±0.01 (m)	Fuerza ±1 (N)
2.00	6
1.90	6
1.80	7
1.70	8
1.60	9
1.50	10
1.40	11
1.30	13
1.20	16
1.10	19
1.00	23
0.90	28
0.80	35
0.70	46

REGRESIÓN

Con el objetivo de estudiar una sepa de cierta bacteria, es necesario cultivarla en un ambiente rico en cierto tipo de nutrientes, hasta que su población alcance la cantidad de 2.4 × 105 bacterias. Desafortunadamente, durante el transporte de las muestras, murió el 98 % de la población de las bacterias que inicialmente se transportaban, sobreviviendo ´únicamente 45 de estas. Con el fin de predecir el tiempo necesario para que la población alcance el mínimo necesario para realizar el estudio, un biólogo mide la población de bacterias y los resultados se muestran en la tabla de al lado.

t (horas)	N (bacterias)
0.0 ± 0.1	45 ± 5
0.5 ± 0.1	54 ± 5
1.0 ± 0.1	63 ± 6
1.5 ± 0.1	76 ± 5
2.0 ± 0.1	93 ± 10
2.5 ± 0.1	110 ± 11
3.0 ± 0.1	130 ± 13
3.5 ± 0.1	151 ± 15
4.0 ± 0.1	178 ± 19

REGRESIÓN

Microprocesador	t (años)	N (transistores)
4004	0	2250
8008	1	2500
8086	3	5000
286	7	29000
Intel386™	11	120000
Intel® Pentium®	14	275000
Intel® Pentium® II	18	1180000
Intel® Pentium® III	22	3100000
Intel® Pentium® 4	26	7500000