Tema 3: Regresión

introducción

Departamento Matemática Aplicada

Universidad de Málaga

Curso 2017-2018

Regresión y correlación

Definición

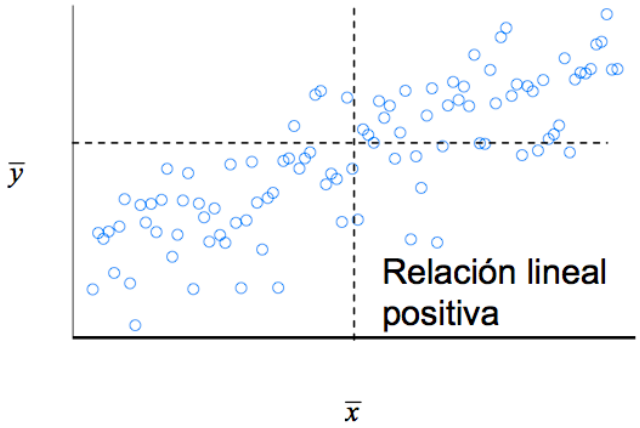
Correlación *es una medida del grado de dependencia entre las variables. La* **regresión** *pretende encontrar un modelo aproximado de la dependencia entre las variables.*

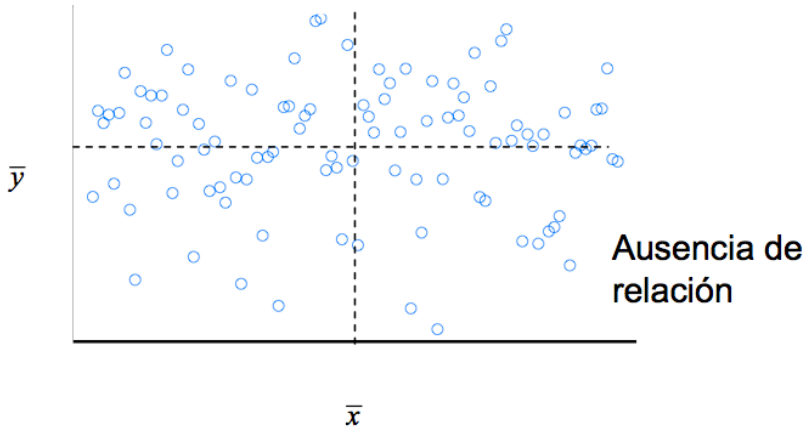
Representando los datos de la muestra de la variable bidimensional obtenemos una nube de puntos. Se llama *línea o curva de regresión* a la función que mejor se ajusta a esa nube de puntos.

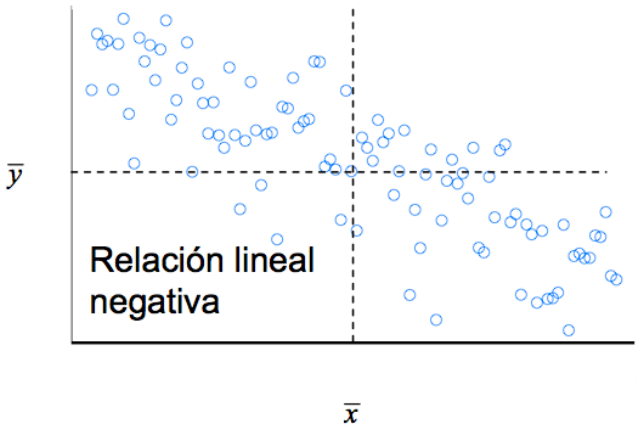
Si todos los valores de la variable satisfacen la ecuación calculada, se dice que las variables están perfectamente correladas. La ecuación de la curva de regresión nos permite predecir valores desconocidos.

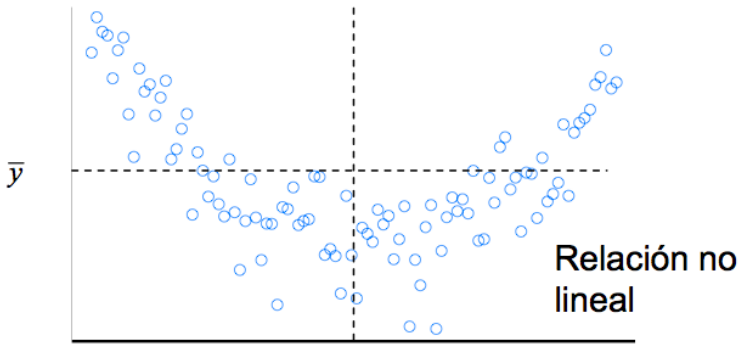
Regresión y correlación

El diagrama de dispersión muestra el tipo de relación existente:



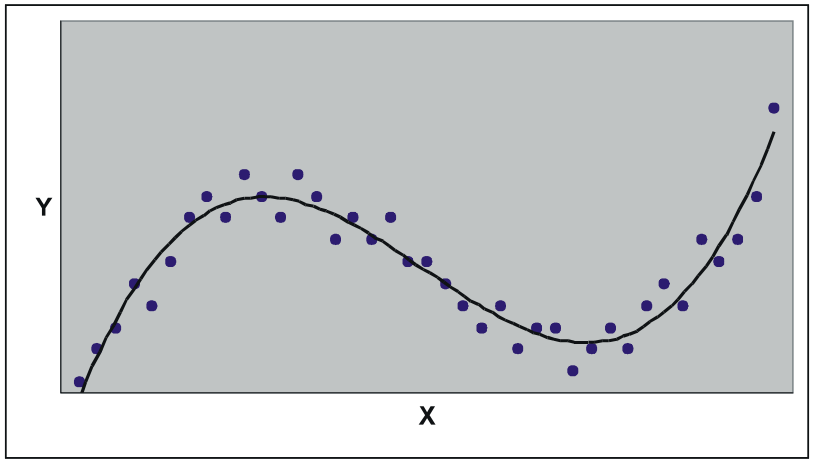






Curva de regresión

A la vista de la nube de puntos, podemos elegir el tipo de modelo a elegir: lineal, cuadrático, exponencial, etc.



Al representar la curva de regresión y la nube de puntos conjuntamente, se puede observar la bondad del ajuste.

Regresión general

La regresión puede realizarse para todo tipo de variables, incluso cualitativas.

	C_1	C_2	Turista	Tripulación	
Español	3	10	25	2	40
Alemán	7	14	8	1	30
Francés	3	0	12	15	30

Regresión de Tipo respecto a Nacionalidad: Si es 'Español' lo más probable es que viaje como 'Turista'. Si es 'Alemán' en C_2 y si es 'Francés' sea de la 'Tripulación'.

Regresión de Nacionalidad respecto a Tipo: Si viaja en 'Clase-1´ o en 'Clase 2' lo más probable es que sea 'Alemán', si en 'Turista' que sea 'Español' y si es de la 'Tripulación' que sea 'Francés'.

Regresión general-2

Regresión de cualitativa con cuantitativa.

L.\ Hr.	$[0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1]$	
Málaga	0	0	3	26	2	31
Granada	1	4	8	12	6	31
Cádiz	0	0	7	11	13	31

Regresión de 'Humedad relativa' respecto a 'Localidad': Para Málaga y Granada lo más probable es (0.6,0.8]. Para 'Cádiz' del (0.8,1].
Regresión de 'Localidad' respecto a 'Humedad relativa': Si es inferior a 0.6 lo más probable es que se haya medido en 'Granada', si es de (0.6,0.8] que sea de 'Málaga' y que sea de 'Cádiz' si es superior a 0.8.

Ajuste por el método de mínimos cuadrados

Sean los datos $\{x_i, y_i\}$, para dos variables estadísticas X e Y cuantitativas. El objetivo es encontrar la función $y = f(x)$ de un subconjunto de las funciones reales (rectas, parábolas, hipérbolas, ...) que más se aproxime a los datos. Se trata pues de minimizar la **función objetivo mínimo-cuadrática:**

$$\mathbf{F} = \sum_{\mathbf{i}} \left(\mathbf{y_i} - \mathbf{y_i^{est}}\right)^2 = \sum_{\mathbf{i}} \left(\mathbf{y_i} - \mathbf{f(x_i)}\right)^2$$

$y_i^{est} = f(x_i)$ es el valor de y estimado por la regresión para x_i.
$e_i = y_i - y_i^{est}$ es el error cometido por el ajuste para el i-ésimo dato.
Minimizar la función objetivo significa minimizar el Error Cuadrático Medio $\left(ECM = \frac{\sum_i e_i^2}{N}\right)$ y la media cuadrática de los errores $\left(MC = \sqrt{\frac{\sum_i e_i^2}{N}}\right)$.

Tipos de ajuste

El tipo de ajuste mínimos cuadrados está determinado por el tipo de función $y = f(x)$ elegido. Los más usados son:

- **Ajuste lineal:** $\mathbf{y = f(x) = a + bx}$ (parámetros a y b).
- **Ajuste parabólico:** $\mathbf{y = a + bx + cx^2}$ (parámetros a, b y c).
- **Ajuste hiperbólico:** $y = \frac{\mathbf{1}}{\mathbf{a+bx}}$ (parámetros a y b).
- **Exponencial:** $\mathbf{y = ae^{bx}}$ (parámetros a y b).

Un ajuste de mínimos cuadrados requiere del cálculo de los valores de los parámetros del modelo que minimizan la función objetivo:
$F(a, b, \ldots) = \sum_i (y_i - f(x_i))^2 = \sum_i e_i^2$.
Existen otros tipos de ajuste. En particular, se define la **curva general de regresión de Y sobre X** como la función que asigna a cada valor x_i de la variable X, la media de la variable $Y/_{x_i}$.

Ajuste de la recta Y/X

Dado un conjunto de puntos $\{(x_i, y_i)\}_{i\in\mathcal{I}}$ queremos calcular una recta de la forma $\mathbf{y = a + bx}$ que mejor se ajuste a esos datos en el sentido 'mínimos cuadrados', es decir, que minimice la función:

$$\mathbf{F} = \sum_{\mathbf{i}\in\mathcal{I}} (\mathbf{y_i} - (\mathbf{a} + \mathbf{bx_i}))^2$$

Los valores de los parámetros **a** y **b** que minimizan esa función se obtienen resolviendo el sistema de ecuaciones:

$$\nabla \mathbf{F} = \begin{bmatrix} \frac{\partial F}{\partial a} \\ \frac{\partial F}{\partial b} \end{bmatrix} = \vec{0} \Rightarrow \left\{ \begin{array}{l} \frac{\partial F}{\partial a} = -2\sum_i (y_i - a - bx_i) = 0 \\ \frac{\partial F}{\partial b} = -2\sum_i (y_i - a - bx_i)x_i = 0 \end{array} \right\} \Rightarrow$$

Ecuaciones normales recta regresión Y/X:

$$\begin{array}{c} \sum_i y_i = Na + b\sum_i x_i \\ \sum_i x_i y_i = a\sum_i x_i + b\sum_i x_i^2 \end{array}$$

Ajuste lineal X/Y

Análogamente, dado un conjunto de puntos $\{(x_i, y_i)\}_{i\in\mathcal{I}}$ queremos ajustar una recta de X sobre Y, de la forma $\mathbf{x} = \mathbf{a}' + \mathbf{b}'\mathbf{y}$ que mejor se ajuste a esos datos en el sentido de 'mínimos cuadrados', la función a minimizar es:

$$\mathbf{G(a', b')} = \sum_{\mathbf{i}\in\mathcal{I}} (\mathbf{x_i} - (\mathbf{a'} + \mathbf{b'y_i}))^2$$

Ahora los parámetros a' y b' deberan satisfacer las ecuaciones:

$$\nabla G = \begin{bmatrix} \frac{\partial G}{\partial a'} \\ \frac{\partial G}{\partial b'} \end{bmatrix} = \vec{0} \Rightarrow \left\{ \begin{array}{l} \frac{\partial G}{\partial a'} = -2\sum_i (x_i - a' - b'y_i) = 0 \\ \frac{\partial G}{\partial b'} = -2\sum_i (x_i - a' - b'y_i) y_i = 0 \end{array} \right\} \Rightarrow$$

Ecuaciones normales recta regresión X/Y:

$$\begin{array}{c} \sum_i x_i = Na' + b' \sum_i y_i \\ \sum_i x_i y_i = a' \sum_i y_i + b' \sum_i y_i^2 \end{array}$$

Ajuste lineal forma matricial

Los sistemas de ecuaciones normales, en forma matricial, para el caso de la regresión lineal son:

Recta de Y sobre X: (y=a+bx)

$$\begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

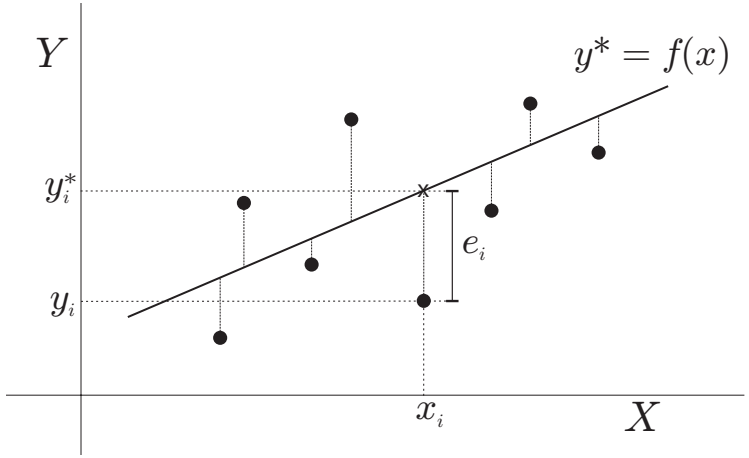
Recta de X sobre Y: (x=a'+b'y)

$$\begin{pmatrix} N & \sum_i y_i \\ \sum_i y_i & \sum_i y_i^2 \end{pmatrix} \begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \sum_i x_i \\ \sum_i x_i y_i \end{pmatrix}$$

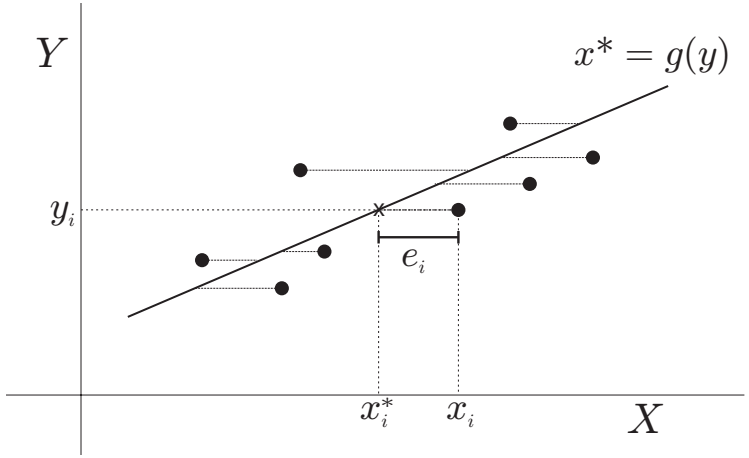
NOTA: Las ecuaciones normales pueden ser fácilmente adaptadas para los casos de disponer de datos con frecuencias.

Significado de los ajustes Y/X y X/Y

Ajuste Y/X: Minimiza $F = \sum_i e_i^2 = \sum_i (y_i - y_i^*)^2$



Ajuste X/Y: Minimiza $G = \sum_i (e_i')^2 = \sum_i (x_i - x_i^*)^2$



Ejemplo ajuste lineal

Ejemplo

La tabla siguiente muestra la evolución de la población española de edad comprendida entre 80 y 89, entre los años 2002 y 2011.

y=Número	893218	926708	963513	1003857	1088204
x=Año	2002	2003	2004	2005	2006
y=Número	1126204	1126704	1166200	1202349	1239183
x=Año	2007	2008	2009	2010	2011

Ajustar las rectas de Y/X y de X/Y

Las ecuaciones normales para Y/X: $\left\{ \begin{array}{c} \sum_i y_i = Na + b\sum_i x_i \\ \sum_i x_i y_i = a\sum_i x_i + b\sum_i x_i^2 \end{array} \right.$

$$\Rightarrow \left\{ \begin{array}{c} 10736140 = 10a + 20065b \\ 21545296484 = 20065a + 40260505b \end{array} \right. \Rightarrow \left\{ \begin{array}{c} a = -77522182.7 \\ b = 39170.5939 \end{array} \right\}$$

La recta ajustada es: $\mathbf{y = -77522182.7 + 39170.5939x}$

Puede usarse para estimar el número previsto para 2012:

Numero=-77522182.7+39170.5939(2012)=1289052.23

Ejemplo ajuste lineal - continuación

Las ecuaciones normales para X/Y: $\left\{ \begin{array}{c} \sum_i x_i = Na' + b' \sum_i y_i \\ \sum_i x_i y_i = a' \sum_i y_i + b' \sum_i y_i^2 \end{array} \right.$

$$\Rightarrow \left\{ \begin{array}{c} 20065 = 10a' + 1073614b' \\ 21545296484 = 1073614a' + 11655937654544b' \end{array} \right\} \Rightarrow$$

$\mathbf{a = 1979.702},\ \mathbf{b = 0.00002496}$

La recta ajustada es: $\mathbf{x = 1979.702 + 0.00002496051y}$

Puede usarse para estimar el año en que se prevén 1300000 individuos de esa edad es:

Año=1979.702+0.00002496051(1300000)= 2012.1507

Así, si los datos están medidos a 1 de enero, se prevé esa cantidad para el 24/2/2012. (0.1507*366= 55.16 días)

Ajuste lineal. Propiedades

Dividiendo por N las ecuaciones normales:

$$\left.\begin{array}{c} \frac{\sum_i y_i}{N} = a + b\frac{\sum_i x_i}{N} \\ \frac{\sum_i x_i y_i}{N} = a\frac{\sum_i x_i}{N} + b\frac{\sum_i x_i^2}{N} \end{array}\right\} \Rightarrow \left\{\begin{array}{c} \bar{y} = a + b\bar{x} \\ m_{11} = a\bar{x} + bm_{20} \end{array}\right\}$$

$$\left.\begin{array}{c} \frac{\sum_i x_i}{N} = a' + b'\frac{\sum_i y_i}{N} \\ \frac{\sum_i x_i y_i}{N} = a'\frac{\sum_i y_i}{N} + b'\frac{\sum_i y_i^2}{N} \end{array}\right\} \Rightarrow \left\{\begin{array}{c} \bar{x} = a' + b'\bar{y} \\ m_{11} = a'\bar{y} + b'm_{02} \end{array}\right\}$$

Deducimos que **el centro de gravedad $\mathbf{G} = (\bar{\mathbf{x}}, \bar{\mathbf{y}})$ pertenece a ambas rectas**. Las rectas Y/X y X/Y se cortan en G.
Eliminando a en la de Y/X y a' en la de X/Y:

$$m_{11} - \bar{x}\bar{y} = b(m_{20} - \bar{x}^2) \Rightarrow \mathbf{b} = \frac{\mathbf{Cov(x,y)}}{\sigma_{\mathbf{x}}^2} = \frac{\mu_{\mathbf{11}}}{\mathbf{V(x)}}$$

$$m_{11} - \bar{x}\bar{y} = b'(m_{02} - \bar{y}^2) \Rightarrow \mathbf{b'} = \frac{\mathbf{Cov(x,y)}}{\sigma_{\mathbf{y}}^2} = \frac{\mu_{\mathbf{11}}}{\mathbf{V(y)}}$$

Coeficiente de correlación lineal de Pearson:

Definición

El **coeficiente de correlación lineal** *mide el grado de relación lineal (magnitud y dirección) entre las variables:*

$$\rho = r = \frac{Cov}{\sigma_x \sigma_y} \qquad (-1 \leq r \leq 1)$$

Significado: La correlación mide la magnitud y la dirección de la dependencia lineal.

- **r>0** Correlación lineal directa.
- **r<0** Correlación lineal inversa.
- **r=0** Variables incorreladas.
- **r=1 ó r=-1** Correlación lineal perfecta (directa o inversa).

Expresiones del ajuste lineal

1. De $b = \frac{\mu_{11}}{Var(x)} = \frac{\mu_{11}\sigma_y}{\sigma_y\sigma_x^2} = r\frac{\sigma_y}{\sigma_x} \Rightarrow \mathbf{b = r\frac{\sigma_y}{\sigma_x}}$
2. De $b' = \frac{\mu_{11}}{Var(y)} = \frac{\mu_{11}\sigma_x}{\sigma_x\sigma_y^2} = r\frac{\sigma_x}{\sigma_y} \Rightarrow \mathbf{b' = r\frac{\sigma_x}{\sigma_y}}$
3. $\mathbf{r = \sqrt{bb'}}$: (El signo de r será el de b. $(r = \mathrm{sgn(b)}\sqrt{\mathrm{bb'}})$
4. Como ambas rectas pasan por $G = (\bar{x}, \bar{y})$, las rectas quedan:

   Recta Y/X: $y - \bar{y} = b(x - \bar{x} \Rightarrow \mathbf{y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})}$

   Recta X/Y: $y - \bar{y} = \frac{1}{b'}(x - \bar{x}) \Rightarrow \mathbf{y - \bar{y} = \frac{1}{r}\frac{\sigma_y}{\sigma_x}(x - \bar{x})}$
5. Llamando $Y = \frac{y-\bar{y}}{\sigma_y}$, $X = \frac{x-\bar{x}}{\sigma_x}$ quedan:

   $$\text{Recta Y/X:} \qquad Y = rX$$

   $$\text{Recta X/Y:} \qquad Y = \frac{1}{r}X$$
6. Las pendientes de las rectas Y/X y X/Y son $m_{X/Y} = b$ y $m_{X/Y} = \frac{1}{b'} \Rightarrow |m_{X/Y}| \geq |m_{Y/X}|$

Varianza residual. Coeficiente de determinación.

Dada una nube de puntos $\{(x_i, y_i)\}$, llamamos **vector residuo** $\vec{e} = (e_i)$ con $e_i = y_i - y_i^{est}$. Es decir, e_i es el error cometido por el ajuste para la i-ésima observación.

Definición

Llamamos **varianza residual** *a la varianza del vector residuo.*

$$\mathbf{V_r = \sum_i f_i (e_i - \bar{e})^2 = \sum_i f_i e_i^2 - \bar{e}^2}$$

Definición

Llamamos **coeficiente de determinación** *a:*

$$\mathbf{R^2 = 1 - \frac{V_r}{V(y)}}$$

Discusión

El coeficiente de determinación R^2 (caso lineal) verifica $0 \leq R^2 \leq 1$.

Definición

Llamamos **varianza explicada** *por la regresión a* $\mathbf{V_e = R^2 V(y)}$

De $R^2 = 1 - \frac{V_r}{V(y)}$, obtenemos: $V_r = (1 - R^2)V(y)$, luego:

$$\mathbf{V(y)} = R^2 V(y) + (1 - R^2)V(y) = R^2 V(y) + V_r = \mathbf{V_e + V_r}$$

Así, $R^2 = \frac{V_e}{V(y)}$ representa la fracción de la varianza explicada por el ajuste.

- $R^2 = 1 \Rightarrow$ Ajuste perfecto.
- $R^2 = 0 \Rightarrow$ El ajuste no explica nada.

Simplificación varianza residual caso lineal

En el caso lineal $\mathbf{V_e} = V(\vec{y^{est}}) = \sum_i f_i(y_i^{est} - \bar{y})^2 = \sum_i f_i(a + bx_i - (a + b\bar{x}))^2 = b^2 \sum_i f_i(x_i - \bar{x})^2 = \mathbf{b^2 V(x)} \Rightarrow$

$$\mathbf{V_e} = b^2\sigma_x^2 = \left(r\frac{\sigma_y}{\sigma_x}\right)^2 \sigma_x^2 = \mathbf{r^2 V(y)}$$

Luego la varianza residual puede obtenerse desde el coeficiente de regresión lineal r:

- $\mathbf{R^2 = r^2}$
- $\mathbf{V_r = (1 - r^2)V(y)}$

Ajuste exponencial $y = ae^{bx}$

$\mathbf{y = ae^{bx}}$ (introducimos logaritmos) $\Rightarrow$

$$ln(y) = ln(ae^{bx}) = ln(a) + bx$$

Llamando: $Y = ln(y)$, $A = ln(a)$ obtenemos: $Y = A + bx$.
Podemos ajustar una recta a $\{(ln(y_i), x_i)\}$ obteniendo $A = ln(a)$, $(a = e^A)$ y b que sustituiremos en $y = ae^{bx}$

Ejemplo

Ajustar una curva del tipo $y = ae^{bx}$ a los datos de la tabla:

x_i	0	1	2	3	6
y_i	7	5	4	3.5	3

Hallar: Varianza residual y coeficiente de determinación.

Ejemplo ajuste exponencial

x_i	0	1	2	3	6	12
y_i	7	5	4	3.5	3	22.5
$Y_i = ln(y_i)$	1.9459	1.6094	1.3863	1.2528	1.0986	7.293
x_i^2	0	1	4	9	36	50
$x_i Y_i$	0	1.6094	2.7726	3.7583	6.5917	14.732
y_i^{est}	5.8846	5.1635	4.5308	3.9756	2.6859	
$y_i - y_i^{est}$	1.1154	−0.1635	−0.5308	−0.4756	0.3141	0.2597
$(y_i - y_i^{est})^2$	1.2442	0.0267	0.2817	0.2262	0.0987	1.8775
y_i^2	49	25	16	12.25	9	111.25

Las ecuaciones normales son: $\begin{cases} 7.293 = 5A + 12b \\ 14.732 = 12A + 50b \end{cases} \Rightarrow \begin{matrix} A = 1.7723 \\ b = -0.1307 \end{matrix}$

$\Rightarrow a = e^{1.7723} = 5.8846$ Obtenemos: $\mathbf{y = 5.8846e^{-0.1307x}}$

$$\mathbf{V_y} = \frac{111.25}{5} - \left(\frac{22.5}{5}\right)^2 = \mathbf{2} \qquad \mathbf{V_r} = \frac{1.8775}{5} - \left(\frac{0.2597}{5}\right)^2 = \mathbf{0.3728}$$

$$\mathbf{R^2} = 1 - \frac{V_r}{V(y)} = 1 - \frac{0.3728}{2} = \mathbf{0.8136}$$

Ajuste hiperbólico $\mathbf{y = \frac{1}{a+bx}}$

$\mathbf{y = \frac{1}{a+bx}}$ (inviertiendo) $\Rightarrow \frac{1}{y} = a + bx$ Llamando: $Y = \frac{1}{y}$, obtenemos: $Y = a + bx$.

Podemos ajustar una recta a $\{(\frac{1}{y_i}, x_i)\}$ obteniendo a, y b que sustituiremos en $y = \frac{1}{a+bx}$

Ejemplo

Ajustar una curva del tipo $y = \frac{1}{a+bx}$ a los datos del problema anterior:

x_i	0	1	2	3	6
y_i	7	5	4	3.5	3

Hallar: Varianza residual y coeficiente de determinación.
¿Qué ajuste es mejor el exponencial, el hiperbólico o el lineal?

Ejemplo ajuste hiperbólico

x_i	0	1	2	3	6	12
y_i	7	5	4	3.5	3	22.5
$Y_i = \frac{1}{y_i}$	0.1429	0.2	0.25	0.2857	0.3333	1.2119
x_i^2	0	1	4	9	36	50
$x_i Y_i$	0	0.2	0.5	0.8571	2	3.5571
y_i^{est}	5.9186	5.0113	4.3451	3.8353	2.8368	
$e_i = y_i - y_i^{est}$	1.0814	−0.0113	−0.3451	−0.3353	0.1632	0.5530
$(y_i - y_i^{est})^2$	1.1693	0.0001	0.1191	0.1124	0.0266	1.4276
y_i^2	49	25	16	12.25	9	111.25
$x_i y_i$	0	5	8	10.5	18	41.5

Ajustamos: $\begin{cases} 1.2119 = 5a + 12b \\ 3.5571 = 12a + 50b \end{cases} \Rightarrow \begin{array}{c} a = 0.1690 \\ b = -0.0306 \end{array} \Rightarrow \mathbf{y = \frac{1}{0.169 - 0.0306x}}$

$\mathbf{V_y = 2}$, $\mathbf{V_r} = \frac{1.4276}{5} - \left(\frac{0.553}{5}\right)^2 = \mathbf{0.2733}$ $\mathbf{R^2} = 1 - \frac{V_r}{V(y)} = 1 - \frac{0.2733}{2} = \mathbf{0.8634}$

$\mathbf{V(x) = 4.24}$, $\mathbf{cov} = \frac{41.5}{5} - \frac{12}{5}\frac{22.5}{5} = -2.5 \Rightarrow r = \frac{-2.5}{\sqrt{2}\sqrt{4.24}} = -0.8585 \Rightarrow \mathbf{r^2 = 0.7370}$

Luego el mejor ajuste es el hiperbólico que explica el 86.34 % de la varianza de y.

NOTA: También se usa como criterio $SSE = \sum_i e_i^2$ que para el exponencial, hiperbólico y lineal dan respectivamente: $SSE_e = 1.8775$, $SSE_h = 1.4276$ y $SSE_L = 2.63$.

Ajuste parabólico $\mathbf{y = a + bx + cx^2}$

Podemos deducir las ecuaciones normales minimizando la función: $F = \sum_i (y_i - (a + bx_i + cx_i))^2$ mediante $\frac{\partial F}{\partial a} = 0$, $\frac{\partial F}{\partial b} = 0$ y $\frac{\partial F}{\partial c} = 0$, tal como se hizo en el caso lineal, pero lo vamos a hacer de otra forma:

Se trata de obtener la función de la forma $f(x) = a \cdot 1 + b \cdot x + c \cdot x^2$ más próxima a los $\{y_i\}$. Debemos elegir un elemento del subespacio vectorial de las funciones que tiene como base $B = \{1, x, x^2\}$.
Las ecuaciones normales salen de considerar que el vector error debe cumplir:

$$\left.\begin{array}{c} \langle \vec{e}, \vec{1} \rangle = 0 \\ \langle \vec{e}, \vec{x} \rangle = 0 \\ \langle \vec{e}, \vec{x^2} \rangle = 0 \end{array}\right\} \Leftrightarrow \left.\begin{array}{c} \sum_i (y_i - (a + bx_i + cx_i^2)) = 0 \\ \sum_i (y_i - (a + bx_i + cx_i^2))x_i = 0 \\ \sum_i (y_i - (a + bx_i + cx_i^2))x_i^2 = 0 \end{array}\right\} \Leftrightarrow$$

$$\left\{\begin{array}{c} \sum_i y_i = Na + b\sum_i x_i + c\sum_i x_i^2 \\ \sum_i y_i x_i = a\sum_i x_i + b\sum_i x_i^2 + c\sum_i x_i^3 \\ \sum_i y_i x_i^2 = a\sum_i x_i^2 + b\sum_i x_i^3 + c\sum_i x_i^4 \end{array}\right\}$$

Otros ajustes

Dado un conjunto de puntos $\{(x_i, y_i)\}$.
1) Ajustar una función del tipo $\mathbf{y = asen(x) + bcos(x)}$.
Una base de las funciones es: $B = \{sen(x), cos(x)\}$.

Luego se debe cumplir:
$$\left.\begin{array}{l} \langle \vec{e}, \vec{sen}(x) \rangle = 0 \\ \langle \vec{e}, \vec{cos}(x) \rangle = 0 \end{array}\right\} \Leftrightarrow$$

$$\left.\begin{array}{l} \sum_i (y_i - (asen(x_i) + bcos(x_i)))sen(x_i) = 0 \\ \sum_i (y_i - (asen(x_i) + bcos(x_i)))cos(x_i) = 0 \end{array}\right\} \Leftrightarrow$$

$$\left\{\begin{array}{l} \sum_i \boldsymbol{y_i sen(x_i) = a \sum_i sen^2(x_i) + b \sum_i cos(x_i)sen(x_i)} \\ \sum_i \boldsymbol{y_i cos(x_i) = a \sum_i sen(x_i)cos(x_i) + b \sum_i cos^2(x_i)} \end{array}\right\}$$

Ajuste de un plano $\mathbf{z = a + bx + cy}$

Dada una nube de puntos $\{(x_i, y_i, z_i)\}_{i\in\mathcal{I}}$, podemos deducir las ecuaciones normales minimizando la función: $F = \sum_i (z_i - (a + bx_i + cy_i))^2$ mediante $\frac{\partial F}{\partial a} = 0$, $\frac{\partial F}{\partial b} = 0$ y $\frac{\partial F}{\partial c} = 0$, pero también:

Debemos obtener la función de la forma $z = f(x, y) = a \cdot 1 + b \cdot x + c \cdot y$ y obtener las componentes del elemento (vector $\vec{z}$ del subespacio vectorial de las funciones que tiene como base $B = \{\vec{1}, \vec{x}, \vec{y}\}$.
Las ecuaciones normales salen de considerar que para el óptimo, el vector error debe ser ortogonal con cualquiera de la base, es decir:

$$\left.\begin{array}{l} \langle \vec{e}, \vec{1} \rangle = 0 \\ \langle \vec{e}, \vec{x} \rangle = 0 \\ \langle \vec{e}, \vec{y} \rangle = 0 \end{array}\right\} \Leftrightarrow \left.\begin{array}{c} \sum_i (z_i - (a + bx_i + cy_i)) = 0 \\ \sum_i (z_i - (a + bx_i + cy_i))x_i = 0 \\ \sum_i (z_i - (a + bx_i + cy_i))y_i = 0 \end{array}\right\} \Leftrightarrow$$

$$\left\{\begin{array}{c} \sum_i z_i = Na + b\sum_i x_i + c\sum_i y_i \\ \sum_i z_i x_i = a\sum_i x_i + b\sum_i x_i^2 + c\sum_i y_i x_i \\ \sum_i z_i y_i = a\sum_i y_i + b\sum_i x_i y_i + c\sum_i y_i^2 \end{array}\right\}$$

Ejemplo de ajuste de un plano

Ejemplo

Ajustar un plano a los puntos:

x_i	0	1	0	-1	0	1	1	2
y_i	0	1	1	1	3	2	2	0
z_i	2	3	4	5	7	4	5	0

Hallar: Suma de los cuadrados de los errores (SSE), Varianza residual y coeficiente de determinación.

las ecuaciones normales son:

$$\left\{ \begin{array}{c} \sum_i z_i = Na + b\sum_i x_i + c\sum_i y_i \\ \sum_i z_i x_i = a\sum_i x_i + b\sum_i x_i^2 + c\sum_i y_i x_i \\ \sum_i z_i y_i = a\sum_i y_i + b\sum_i x_i y_i + c\sum_i y_i^2 \end{array} \right\}$$

Ejemplo ajuste de un plano

x_i	y_i	z_i	x_i^2	y_i^2	x_iy_i	z_ix_i	z_iy_i	z_i^*	$e_i = z_i - z_i^*$	e_i^2
0	0	2	0	0	0	0	0	2.2045	−0.2045	0.0418
1	1	3	1	1	1	3	3	2.8068	0.1932	0.0373
0	1	4	0	1	0	0	4	3.8636	0.1364	0.0186
−1	1	5	1	1	−1	−5	5	4.9205	0.0795	0.0063
0	3	7	0	9	0	0	21	7.1818	−0.1818	0.0331
1	2	4	1	4	2	4	8	4.4659	−0.4659	0.2171
1	2	5	1	4	2	5	10	4.4659	0.5341	0.2853
2	0	0	4	0	0	0	0	0.0909	−0.0909	0.0083
4	10	30	8	20	4	7	51		0	0.6477

Las ecuaciones normales son: $(z_i^* = a + bx_i + cy_i)$

$$\begin{cases} 30 = 8a + 4b + 10c \\ 7 = 4a + 8b + 4c \\ 51 = 10a + 4b + 20c \end{cases} \Rightarrow \begin{cases} a = 2.2045 \\ b = -1.0568 \Rightarrow \mathbf{z = 2.2045 - 1.0568x + 1.6591y} \\ c = 1.6591 \end{cases}$$

$\mathbf{SSE} = \sum_i \mathbf{e_i^2} = \mathbf{0.6477}$, $\mathbf{V(y)} = \frac{20}{8} - \left(\frac{10}{8}\right)^2 = \mathbf{0.9375}$,

$\mathbf{V_r} = \frac{0.6477}{8} - 0^2 = \mathbf{0.0810}$, $\mathbf{R^2} = 1 - \frac{V_r}{V(y)} = 1 - \frac{0.0810}{0.9375} \approx \mathbf{0.9794}$