Regresión y Correlación

Manuel Ruiz Marín

Universidad Politécnica de Cartagena

Índice del Tema

¿Qué Necesitamos Saber?

¿Qué Necesitamos Saber?

¿Qué Necesitamos Saber?

1. Calcular la media aritmética, $\overline{x}$.
2. Calcular la varianza s^2 y desviación típica s.
3. Distribuciones bidimensionales.
4. Calcular la covarianza S_{xy}.
5. Cobocer las ecuaciones de la: recta, hiperbola, parábola, función exponencial, función potencial y de un hiperplano.
6. Representación grafica de las funciones anteriores.
7. Propiedades de los logaritmos.
8. Derivar y determinar extremos relativos en funciones reales de variable real.
9. Cálculo matricial.

¿Qué Necesitamos Saber?

¿Qué Necesitamos Saber?

1. Calcular la media aritmética, $\overline{x}$.
2. Calcular la varianza s^2 y desviación típica s.
3. Distribuciones bidimensionales.
4. Calcular la covarianza S_{xy}.
5. Cobocer las ecuaciones de la: recta, hiperbola, parábola, función exponencial, función potencial y de un hiperplano.
6. Representación grafica de las funciones anteriores.
7. Propiedades de los logaritmos.
8. Derivar y determinar extremos relativos en funciones reales de variable real.
9. Cálculo matricial.

Objeto

Objeto

Objeto

Analizar la existencia Establecer relaciones funcionales donde una serie de magnitudes (variables o atributos) $X_1, X_2, \ldots, X_p$ se supone que están relacionadas con una variable Y mediante la expresión

$$Y = f(X_1, X_2, \ldots, X_p).$$

Asimismo pretendemos medir el grado de bondad de la relación establecida.

Objeto

Objeto

Analizar la existencia Establecer relaciones funcionales donde una serie de magnitudes (variables o atributos) $X_1, X_2, \ldots, X_p$ se supone que están relacionadas con una variable Y mediante la expresión

$$Y = f(X_1, X_2, \ldots, X_p).$$

Asimismo pretendemos medir el grado de bondad de la relación establecida.

Regresión con Dos Variables

Regresión con Dos Variables

¿Que relación existe entre X e Y?

	y_1	y_2	$\dots$	y_s
x_1	n_{11}	n_{12}	$\dots$	n_{1s}
x_2	n_{21}	n_{22}	$\dots$	n_{2s}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{ks}

Regresión con Dos Variables

¿Que relación existe entre X e Y?

	y_1	y_2	$\ldots$	y_s
x_1	n_{11}	n_{12}	$\ldots$	n_{1s}
x_2	n_{21}	n_{22}	$\ldots$	n_{2s}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\ldots$	n_{ks}

¿Que relación existe entre X e Y?

Existe algún tipo de asociación, dependencia o covariación entre las variables X e Y.

Regresión con Dos Variables

¿Que relación existe entre X e Y?

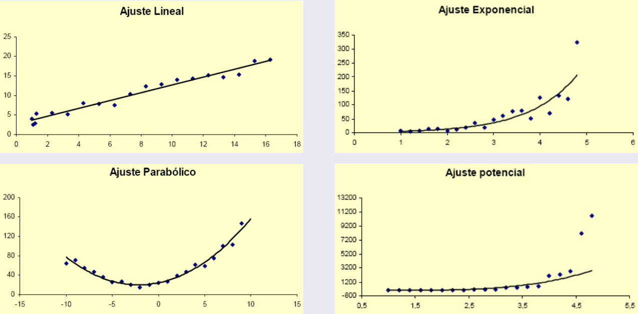
	y_1	y_2	$\ldots$	y_s
x_1	n_{11}	n_{12}	$\ldots$	n_{1s}
x_2	n_{21}	n_{22}	$\ldots$	n_{2s}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\ldots$	n_{ks}

¿Que relación existe entre X e Y?

Existe algún tipo de asociación, dependencia o covariación entre las variables X e Y.

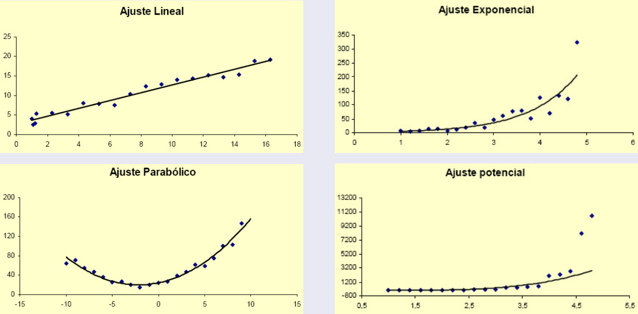
Regresión con Dos Variables

¿Como seleccionar el modelo?. Diagrama de dispersión



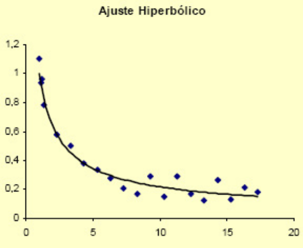
Regresión con Dos Variables

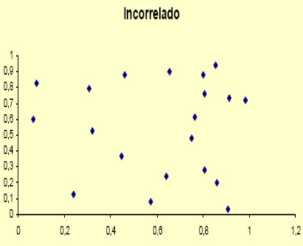
¿Como seleccionar el modelo?. Diagrama de dispersión



Regresión con Dos Variables

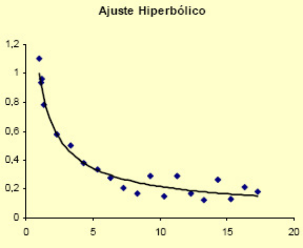
¿Como seleccionar el modelo?. Diagrama de dispersión

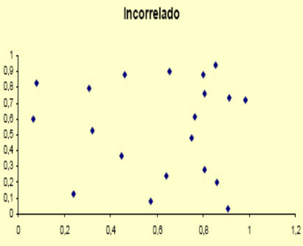




Regresión con Dos Variables

¿Como seleccionar el modelo?. Diagrama de dispersión

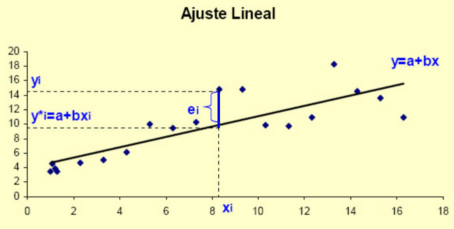




Ajuste por Mínimos Cuadrados a Una Recta

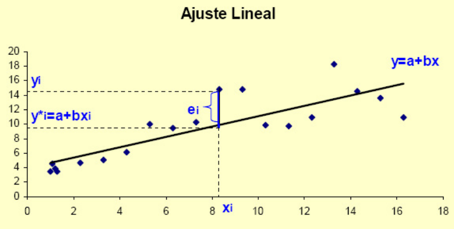
Ajuste por Mínimos Cuadrados a una Recta

Regresión Lineal por Mínimos Cuadrados



Ajuste por Mínimos Cuadrados a una Recta

Regresión Lineal por Mínimos Cuadrados



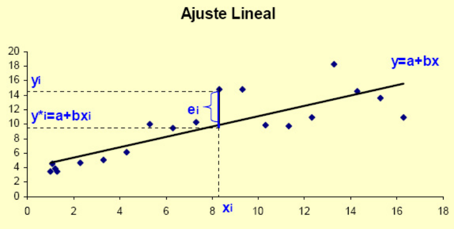
Objetivo

Hacer mínima la suma de los errores al cuadrado

$$f(a,b) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - y_i^*)^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2$$

Ajuste por Mínimos Cuadrados a una Recta

Regresión Lineal por Mínimos Cuadrados



Objetivo

Hacer mínima la suma de los errores al cuadrado

$$f(a,b) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - y_i^*)^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2$$

Ajuste por Mínimos Cuadrados a una Recta

Primera Condición Para Mínimo. Ecuaciones Normales

$$\left.\begin{array}{l}\dfrac{\partial f}{\partial a}(a,b) = -2\left(\displaystyle\sum_{i=1}^{N} y_i - Na - b\sum_{i=1}^{N} x_i\right) = 0 \\ \dfrac{\partial f}{\partial b}(a,b) = -2\left(\displaystyle\sum_{i=1}^{N} y_i x_i - a\sum_{i=1}^{N} x_i - b\sum_{i=1}^{N} x_i^2\right) = 0\end{array}\right\}$$

Ajuste por Mínimos Cuadrados a una Recta

Primera Condición Para Mínimo. Ecuaciones Normales

$$\left.\begin{array}{l} \frac{\partial f}{\partial a}(a,b) = -2\left(\sum_{i=1}^{N} y_i - Na - b\sum_{i=1}^{N} x_i\right) = 0 \\ \frac{\partial f}{\partial b}(a,b) = -2\left(\sum_{i=1}^{N} y_i x_i - a\sum_{i=1}^{N} x_i - b\sum_{i=1}^{N} x_i^2\right) = 0 \end{array}\right\}$$

Segunda Condición Para Mínimo. El Hessiano

$$Hf(a,b) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 a}(a,b) = 2N & \frac{\partial^2 f}{\partial a \partial b}(a,b) = 2N\overline{x} \\ \frac{\partial^2 f}{\partial b \partial a}(a,b) = 2N\overline{x} & \frac{\partial^2 f}{\partial^2 b}(a,b) = 2\sum_{i=1}^{N} x_i^2 \end{pmatrix}$$

$\Delta_1 = 2N > 0$ y $\Delta_2 = 4N^2 s_x^2 > 0$.

Luego los a y b solución de las ecuaciones normales son mínimo para $f(a,b)$.

Ajuste por Mínimos Cuadrados a una Recta

Primera Condición Para Mínimo. Ecuaciones Normales

$$\left.\begin{array}{l} \frac{\partial f}{\partial a}(a,b) = -2\left(\sum_{i=1}^{N} y_i - Na - b\sum_{i=1}^{N} x_i\right) = 0 \\ \frac{\partial f}{\partial b}(a,b) = -2\left(\sum_{i=1}^{N} y_i x_i - a\sum_{i=1}^{N} x_i - b\sum_{i=1}^{N} x_i^2\right) = 0 \end{array}\right\}$$

Segunda Condición Para Mínimo. El Hessiano

$$Hf(a,b) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 a}(a,b) = 2N & \frac{\partial^2 f}{\partial a \partial b}(a,b) = 2N\overline{x} \\ \frac{\partial^2 f}{\partial b \partial a}(a,b) = 2N\overline{x} & \frac{\partial^2 f}{\partial^2 b}(a,b) = 2\sum_{i=1}^{N} x_i^2 \end{pmatrix}$$

$\Delta_1 = 2N > 0$ y $\Delta_2 = 4N^2 s_x^2 > 0$.
Luego los a y b solución de las ecuaciones normales son mínimo para $f(a,b)$.

Ajuste por Mínimos Cuadrados a una Recta

Sistema de Ecuaciones Normales

$$\left.\begin{array}{l} \displaystyle\sum_{i=1}^{N} y_i = Na + b\sum_{i=1}^{N} x_i \\ \displaystyle\sum_{i=1}^{N} y_i x_i = a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 \end{array}\right\}$$

Ajuste por Mínimos Cuadrados a una Recta

Sistema de Ecuaciones Normales

$$\left.\begin{aligned} \sum_{i=1}^{N} y_i &= Na + b\sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} y_i x_i &= a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 \end{aligned}\right\}$$

Dividiendo las ecuaciones por N tenemos

Ecuaciones Normales

$$\left.\begin{aligned} &\overline{y} = a + b\overline{x} \\ &\frac{1}{N}\sum_{i=1}^{N} y_i x_i = a\overline{x} + b\frac{1}{N}\sum_{i=1}^{N} x_i^2 \end{aligned}\right\}$$

Ajuste por Mínimos Cuadrados a una Recta

Sistema de Ecuaciones Normales

$$\left.\begin{array}{l} \displaystyle\sum_{i=1}^{N} y_i = Na + b\sum_{i=1}^{N} x_i \\ \displaystyle\sum_{i=1}^{N} y_i x_i = a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 \end{array}\right\}$$

Dividiendo las ecuaciones por N tenemos

Ecuaciones Normales

$$\left.\begin{array}{l} \overline{y} = a + b\overline{x} \\ \displaystyle\frac{1}{N}\sum_{i=1}^{N} y_i x_i = a\overline{x} + b\frac{1}{N}\sum_{i=1}^{N} x_i^2 \end{array}\right\}$$

Ajuste por Mínimos Cuadrados a una Recta

Coeficientes de la Recta de Regresión

$$\left.\begin{array}{l} b = \dfrac{S_{xy}}{s_x^2} \\ a = \overline{y} - \dfrac{S_{xy}}{s_x^2}\overline{x} \end{array}\right\}$$

Ajuste por Mínimos Cuadrados a una Recta

Coeficientes de la Recta de Regresión

$$\left.\begin{aligned} b &= \frac{S_{xy}}{s_x^2} \\ a &= \overline{y} - \frac{S_{xy}}{s_x^2}\overline{x} \end{aligned}\right\}$$

El Binomio Cálculos-Gráficos

Ajuste por Mínimos Cuadrados a una Recta

El Binomio Cálculos-Gráficos. Cambio de Escala

El aspecto de un diagrama de dispersión puede alterarse drásticamente con un cambio de escala.

Un aumento en la escala puede hacernos creer en la posible existencia de un ajuste lineal para dos variables incorreladas.

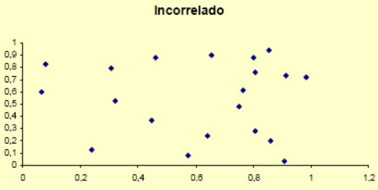
Ajuste por Mínimos Cuadrados a una Recta

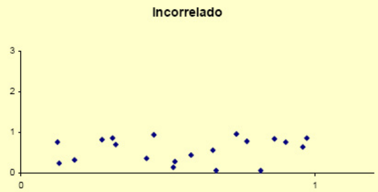
El Binomio Cálculos-Gráficos. Cambio de Escala

El aspecto de un diagrama de dispersión puede alterarse drásticamente con un cambio de escala.

Un aumento en la escala puede hacernos creer en la posible existencia de un ajuste lineal para dos variables incorreladas.

El Binomio Cálculos-Gráficos. Cambio de Escala



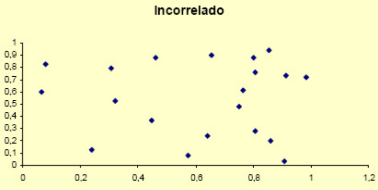


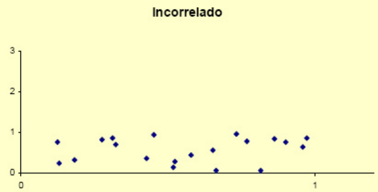
Ajuste por Mínimos Cuadrados a una Recta

El Binomio Cálculos-Gráficos. Cambio de Escala

El aspecto de un diagrama de dispersión puede alterarse drásticamente con un cambio de escala.
Un aumento en la escala puede hacernos creer en la posible existencia de un ajuste lineal para dos variables incorreladas.

El Binomio Cálculos-Gráficos. Cambio de Escala

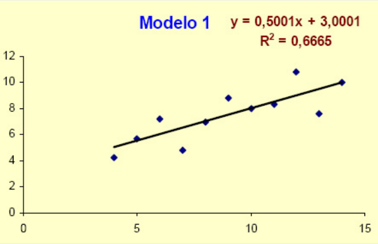


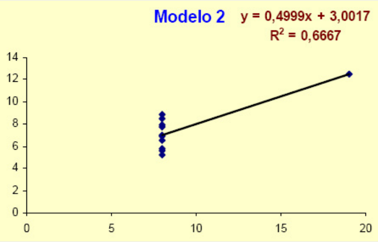


Ajuste por Mínimos Cuadrados a una Recta

El Binomio Cálculos-Gráficos. Hacer Cálculos Sin Representar Datos

Presentamos dos modelos de regresión lineales con los mismos coeficientes y mismo grado de bondad.
Pero el ajuste lineal del Modelo 2 no es el más adecuado para los datos proporcionados.
De ahí la importancia de la representación gráfica.

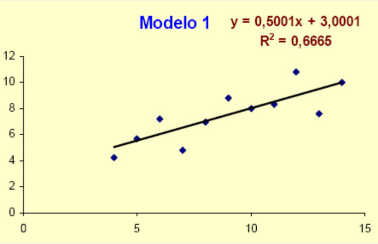


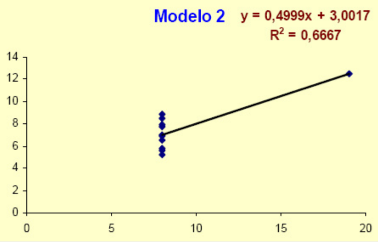


Ajuste por Mínimos Cuadrados a una Recta

El Binomio Cálculos-Gráficos. Hacer Cálculos Sin Representar Datos

Presentamos dos modelos de regresión lineales con los mismos coeficientes y mismo grado de bondad.
Pero el ajuste lineal del Modelo 2 no es el más adecuado para los datos proporcionados.
De ahí la importancia de la representación gráfica.





Propiedades de la Recta de Regresión

Ajuste por Mínimos Cuadrados a una Recta

Propiedades de la recta de regresión

- $\sum_{i=1}^{N} e_i = 0$ (por la primera ecuación normal)
- Dadas las rectas de regresión

$$\left.\begin{array}{r} Y = a + bX \\ X = c + dY \end{array}\right\}$$

  se cortan en el punto $(\overline{x}, \overline{y})$.
- Las variables Y^* y E son incorreladas, es decir $S_{y^*e} = 0$ (por la primera propiedad y la segunda ecuación normal).

Ajuste por Mínimos Cuadrados a una Recta

Propiedades de la recta de regresión

- $\sum_{i=1}^{N} e_i = 0$ (por la primera ecuación normal)
- Dadas las rectas de regresión

$$\left.\begin{array}{l} Y = a + bX \\ X = c + dY \end{array}\right\}$$

  se cortan en el punto $(\overline{x}, \overline{y})$.
- Las variables Y^* y E son incorreladas, es decir $S_{y^*e} = 0$ (por la primera propiedad y la segunda ecuación normal).

Coeficientes de Determinación y Correlación

Coeficientes de Determinación y Correlación

Objetivo

Una vez determinada la recta de regresión queremos determinar si el ajuste es bueno.

Objetivo

Encontrar un coeficiente que indique el grado de bondad del ajuste.

Coeficientes de Determinación y Correlación

Objetivo

Una vez determinada la recta de regresión queremos determinar si el ajuste es bueno.

Objetivo

Encontrar un coeficiente que indique el grado de bondad del ajuste.

Coeficientes de Determinación y Correlación

La Varianza Residual

Como los coeficientes a y b calculados por mínimos cuadrados hacen mínima

$$f(a,b) = \sum_{i=1}^{N} e_i^2,$$

parece lógico que una primera medida del grado de bondad del ajuste sería calcular el valor de ese mínimo.

Obsérvese que como $\sum_{i=1}^{N} e_i = 0$ tenemos que:

$$s_E^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2,$$

término que se le conoce como varianza residual.

Coeficientes de Determinación y Correlación

La Varianza Residual

Como los coeficientes a y b calculados por mínimos cuadrados hacen mínima

$$f(a, b) = \sum_{i=1}^{N} e_i^2,$$

parece lógico que una primera medida del grado de bondad del ajuste sería calcular el valor de ese mínimo.

Obsérvese que como $\sum_{i=1}^{N} e_i = 0$ tenemos que:

$$s_E^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2,$$

término que se le conoce como varianza residual.

Coeficientes de Determinación y Correlación

Cálculo de la Varianza Residual

$$s_e^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N} e_i(y_i - y_i^*) = \frac{1}{N}\sum_{i=1}^{N} e_i y_i - \frac{1}{N}\sum_{i=1}^{N} e_i y_i^*$$

Como las variables Y^* y E son incorreladas tenemos que $\sum_{i=1}^{N} e_i y_i^* = 0$ y por tanto:

$$s_e^2 = \frac{1}{N}\sum_{i=1}^{N} e_i y_i = \frac{\sum_{i=1}^{N} y_i^2 - a\sum_{i=1}^{N} y_i - b\sum_{i=1}^{N} x_i y_i}{N}$$

Coeficientes de Determinación y Correlación

Cálculo de la Varianza Residual

$$s_e^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N} e_i(y_i - y_i^*) = \frac{1}{N}\sum_{i=1}^{N} e_i y_i - \frac{1}{N}\sum_{i=1}^{N} e_i y_i^*$$

Como las variables Y^* y E son incorreladas tenemos que $\sum_{i=1}^{N} e_i y_i^* = 0$ y por tanto:

$$s_e^2 = \frac{1}{N}\sum_{i=1}^{N} e_i y_i = \frac{\sum_{i=1}^{N} y_i^2 - a\sum_{i=1}^{N} y_i - b\sum_{i=1}^{N} x_i y_i}{N}$$

Coeficientes de Determinación y Correlación

La Varianza Residual

La Varianza residual mide la dispersión existente entre los valores observados (y_i) y los valores ajustados (y_i^*).

- Si s_e^2 es muy grande la bondad del ajuste será baja.
- Si s_e^2 es muy pequeña la bondad del ajuste será alta.
- Si $s_e^2 = 0$ es porque todos los $e_i = 0$ y por tanto todos los puntos estan sobre la recta. Habria una relación funcional perfecta.

Coeficientes de Determinación y Correlación

La Varianza Residual

La Varianza residual mide la dispersión existente entre los valores observados (y_i) y los valores ajustados (y_i^*).

- Si s_e^2 es muy grande la bondad del ajuste será baja.
- Si s_e^2 es muy pequeña la bondad del ajuste será alta.
- Si $s_e^2 = 0$ es porque todos los $e_i = 0$ y por tanto todos los puntos estan sobre la recta. Habria una relación funcional perfecta.

Coeficientes de Determinación y Correlación

Relación Entre la Varianzas $s_y^2, s_{y^*}^2$ y s_e^2

En el caso de ajuste lineal, como $Y = Y^* + E$ y $\overline{e} = 0$ tenemos que $\overline{y} = \overline{y}^*$ y además $\sum_{i=1}^{N} y_i^* e_i = 0$. Por tanto:

$$s_y^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i^* + e_i)^2 - \overline{y}^{*2} =$$
$$= \frac{1}{N}\sum_{i=1}^{N} y_i^{*2} - \overline{y}^{*2} + \frac{1}{N}\sum_{i=1}^{N} e_i^2 = s_{y^*}^2 + s_e^2$$

Coeficientes de Determinación y Correlación

Relación Entre la Varianzas $s_y^2, s_{y^*}^2$ y s_e^2

En el caso de ajuste lineal, como $Y = Y^* + E$ y $\overline{e} = 0$ tenemos que $\overline{y} = \overline{y}^*$ y además $\sum_{i=1}^{N} y_i^* e_i = 0$. Por tanto:

$$s_y^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i^* + e_i)^2 - \overline{y}^{*2} =$$
$$= \frac{1}{N}\sum_{i=1}^{N} y_i^{*2} - \overline{y}^{*2} + \frac{1}{N}\sum_{i=1}^{N} e_i^2 = s_{y^*}^2 + s_e^2$$

Relación Entre la Varianzas $s_y^2, s_{y^*}^2$ y s_e^2

$$\mathbf{s_y^2 = s_{y^*}^2 + s_e^2}$$

Coeficientes de Determinación y Correlación

Relación Entre la Varianzas $s_y^2, s_{y^*}^2$ y s_e^2

En el caso de ajuste lineal, como $Y = Y^* + E$ y $\overline{e} = 0$ tenemos que $\overline{y} = \overline{y}^*$ y además $\sum_{i=1}^{N} y_i^* e_i = 0$. Por tanto:

$$s_y^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i^* + e_i)^2 - \overline{y}^{*2} =$$
$$= \frac{1}{N}\sum_{i=1}^{N} y_i^{*2} - \overline{y}^{*2} + \frac{1}{N}\sum_{i=1}^{N} e_i^2 = s_{y^*}^2 + s_e^2$$

Relación Entre la Varianzas $s_y^2, s_{y^*}^2$ y s_e^2

$$\mathbf{s_y^2 = s_{y^*}^2 + s_e^2}$$

Coeficientes de Determinación y Correlación

Problema

- s_e^2 viene dada en las unidades de medida de la variable dependiente al cuadrado.
- ¿A partir de que valores s_e^2 es suficientemente pequeña o grande como para admitir un buen o mal ajuste?.

Coeficientes de Determinación y Correlación

Problema

- s_e^2 viene dada en las unidades de medida de la variable dependiente al cuadrado.
- ¿A partir de que valores s_e^2 es suficientemente pequeña o grande como para admitir un buen o mal ajuste?.

Solución

Determinar un coeficiente que mida el grado de bondad del ajuste lineal de manera que:

- Sea adimensional, es decir, carezca de unidades de medida.
- Nos permita decidir si el juste es aceptable o no.

Coeficientes de Determinación y Correlación

Problema

- s_e^2 viene dada en las unidades de medida de la variable dependiente al cuadrado.
- ¿A partir de que valores s_e^2 es suficientemente pequeña o grande como para admitir un buen o mal ajuste?.

Solución

Determinar un coeficiente que mida el grado de bondad del ajuste lineal de manera que:

- Sea adimensional, es decir, carezca de unidades de medida.
- Nos permita decidir si el juste es aceptable o no.

Coeficientes de Determinación y Correlación

Coeficiente de Determinación

El coeficiente de determinación R^2 determina la proporción de la varianza de la variable Y que queda explicada por la variable ajustada Y^*:

$$R^2 = \frac{s_{y^*}^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

Coeficientes de Determinación y Correlación

Coeficiente de Determinación

El coeficiente de determinación R^2 determina la proporción de la varianza de la variable Y que queda explicada por la variable ajustada Y^*:

$$R^2 = \frac{s_{y^*}^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

Rango de Variación de R^2

Como R^2 es un cociente de varianzas tenemos que $0 \leq R^2$. Por otro lado $1 - R^2 = \frac{s_e^2}{s_y^2} \geq 0$ y por tanto $R^2 \leq 1$. Por tanto:

$$\mathbf{0 \leq R^2 \leq 1 \quad \Rightarrow \quad -1 \leq R \leq 1}$$

Coeficientes de Determinación y Correlación

Coeficiente de Determinación

El coeficiente de determinación R^2 determina la proporción de la varianza de la variable Y que queda explicada por la variable ajustada Y^*:

$$R^2 = \frac{s^2_{y^*}}{s^2_y} = \frac{s^2_y - s^2_e}{s^2_y} = 1 - \frac{s^2_e}{s^2_y}$$

Rango de Variación de R^2

Como R^2 es un cociente de varianzas tenemos que $0 \leq R^2$. Por otro lado $1 - R^2 = \frac{s^2_e}{s^2_y} \geq 0$ y por tanto $R^2 \leq 1$. Por tanto:

$$\mathbf{0 \leq R^2 \leq 1 \quad \Rightarrow \quad -1 \leq R \leq 1}$$

Coeficientes de Determinación y Correlación

Coeficiente de Correlación Lineal

Se define el coeficiente de correlación lineal como:

$$r = \frac{s_{xy}}{s_x s_y}$$

Coeficientes de Determinación y Correlación

Coeficiente de Correlación Lineal

Se define el coeficiente de correlación lineal como:

$$r = \frac{s_{xy}}{s_x s_y}$$

Coeficientes de Determinación y Correlación

Coeficiente de Correlación Lineal

Teorema En el caso de ajuste lineal se verifica que $R^2 = r^2$.

Demostración:

$$R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = \frac{s_y^2 - \dfrac{\sum_{i=1}^{N} y_i^2 - a\sum_{i=1}^{N} y_i - b\sum_{i=1}^{N} x_i y_i}{N}}{s_y^2}$$

Por otro lado como

$$\left.\begin{array}{l} a = \overline{y} - \dfrac{S_{xy}}{s_x}\overline{x} \\ b = \dfrac{S_{xy}}{s_x} \end{array}\right\}$$

tenemos que

$$R^2 = \frac{s_y^2 - (s_y^2 + \overline{y}) + (\overline{y} - \frac{S_{xy}}{s_x}\overline{x})\overline{y} + \frac{S_{xy}}{s_x}\overline{xy} + S_{xy}}{s_y^2} = \frac{\frac{S_{xy}}{s_x^2}}{s_y^2} = \frac{S_{xy}}{s_x^2 s_y^2} = r^2$$

Coeficientes de Determinación y Correlación

Coeficiente de Correlación Lineal

Teorema En el caso de ajuste lineal se verifica que $R^2 = r^2$.

Demostración:

$$R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = \frac{s_y^2 - \dfrac{\sum_{i=1}^{N} y_i^2 - a\sum_{i=1}^{N} y_i - b\sum_{i=1}^{N} x_i y_i}{N}}{s_y^2}$$

Por otro lado como

$$\left.\begin{aligned} a &= \overline{y} - \frac{S_{xy}}{s_x}\overline{x} \\ b &= \frac{S_{xy}}{s_x} \end{aligned}\right\}$$

tenemos que

$$R^2 = \frac{s_y^2 - (s_y^2 + \overline{y}) + (\overline{y} - \frac{S_{xy}}{s_x}\overline{x})\overline{y} + \frac{S_{xy}}{s_x}\overline{xy} + S_{xy}}{s_y^2} = \frac{\frac{S_{xy}}{s_x^2}}{s_y^2} = \frac{S_{xy}}{s_x^2 s_y^2} = r^2$$

Coeficientes de Determinación y Correlación

Interpretación de los Valores de R

- $R = 1 \Rightarrow s_e^2 = 0$. Todos los valores teóricos coinciden con los observados. Existe correlaión perfecta positiva.
- $R = -1 \Rightarrow s_e^2 = 0$. Todos los valores teóricos coinciden con los observados. Existe correlaión perfecta negativa.
- $R = 0 \Rightarrow s_{y*}^2 = 0$ y por tanto $s_y^2 = s_e^2$. Luego no se consigue ninguna explicación de la variable Y relacionándola con la X. La correlación es nula.
- $-1 \leq R \leq 0$, la correlación sería negativa, siendo *más intensa cuanto mas cerca esté de* -1.
- $0 \leq R \leq 1$, la correlación sería positiva, indicando una *mayor interrelación cuanto más próximo esté de* 1.

Coeficientes de Determinación y Correlación

Interpretación de los Valores de R

- $R = 1 \Rightarrow s_e^2 = 0$. Todos los valores teóricos coinciden con los observados. Existe correlaión perfecta positiva.
- $R = -1 \Rightarrow s_e^2 = 0$. Todos los valores teóricos coinciden con los observados. Existe correlaión perfecta negativa.
- $R = 0 \Rightarrow s_{y*}^2 = 0$ y por tanto $s_y^2 = s_e^2$. Luego no se consigue ninguna explicación de la variable Y relacionándola con la X. La correlación es nula.
- $-1 \leq R \leq 0$, la correlación sería negativa, siendo *más intensa cuanto mas cerca esté de* -1.
- $0 \leq R \leq 1$, la correlación sería positiva, indicando una *mayor interrelación cuanto más próximo esté de* 1.

Regresión No Lineal

Regresión No Lineal

Ajuste Hiperbólico $Y = a + \frac{b}{X}$

Queremos determinar el ajuste $Y = a + \frac{b}{X}$ por mínimos cuadrados.

Es equivalente a:

Mediante el cambio de variable $Z = \frac{1}{X}$ realizar el ajuste lineal

$$Y = a + bZ.$$

Regresión No Lineal

Ajuste Hiperbólico $Y = a + \frac{b}{X}$

Queremos determinar el ajuste $Y = a + \frac{b}{X}$ por mínimos cuadrados.

Es equivalente a:

Mediante el cambio de variable $Z = \frac{1}{X}$ realizar el ajuste lineal

$$Y = a + bZ.$$

Regresión No Lineal

Ajuste Exponencial $Y = ae^{bX}$

Queremos determinar el ajuste $Y = ae^{bX}$ por mínimos cuadrados.

Es equivalente a:

Tomando logaritmos $W = \ln(Y)$ y $A = \ln(a)$ realizar el ajuste lineal

$$W = A + bX.$$

Luego los parámetros que determinan la función exponencial son

$$a = e^{A} \text{ y } b$$

.

Regresión No Lineal

Ajuste Exponencial $Y = ae^{bX}$

Queremos determinar el ajuste $Y = ae^{bX}$ por mínimos cuadrados.

Es equivalente a:

Tomando logaritmos $W = \ln(Y)$ y $A = \ln(a)$ realizar el ajuste lineal

$$W = A + bX.$$

Luego los parámetros que determinan la función exponencial son

$$a = e^{A} \text{ y } b$$

.

Regresión No Lineal

Ajuste Potencial $Y = aX^b$

Queremos determinar el ajuste $Y = aX^b$ por mínimos cuadrados.

Es equivalente a:

Tomando logaritmos $W = \ln(Y)$, $A = \ln(a)$ y $Z = \ln(X)$ realizar el ajuste lineal

$$W = A + bZ.$$

Luego los parámetros que determinan la función exponencial son

$$a = e^A \text{ y } b$$

.

Regresión No Lineal

Ajuste Potencial $Y = aX^b$

Queremos determinar el ajuste $Y = aX^b$ por mínimos cuadrados.

Es equivalente a:

Tomando logaritmos $W = \ln(Y)$, $A = \ln(a)$ y $Z = \ln(X)$ realizar el ajuste lineal

$$W = A + bZ.$$

Luego los parámetros que determinan la función exponencial son

$$a = e^A \text{ y } b$$

.

Regresión No Lineal

Ajuste Parabólico $Y = a + bX + cX^2$

Queremos determinar el ajuste $Y = a + bX + cX^2$ por mínimos cuadrados. Es decir, determinar a,b y c tales que minimizan:

$$f(a, b, c) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - a - bx_i - cx_i^2)^2$$

Para obtener las ecuaciones normales:

$$\begin{aligned}
\frac{\partial f}{\partial a} &= -2\sum_{i=1}^{N}(y_i - a - bx_i - cx_i^2) = 0 \\
\frac{\partial f}{\partial b} &= -2\sum_{i=1}^{N}(y_i - a - bx_i - cx_i^2)x_i = 0 \\
\frac{\partial f}{\partial c} &= -2\sum_{i=1}^{N}(y_i - a - bx_i - cx_i^2)x_i^2 = 0
\end{aligned}$$

Regresión No Lineal

Ajuste Parabólico $Y = a + bX + cX^2$

Queremos determinar el ajuste $Y = a + bX + cX^2$ por mínimos cuadrados. Es decir, determinar a,b y c tales que minimizan:

$$f(a, b, c) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - a - bx_i - cx_i^2)^2$$

Para obtener las ecuaciones normales:

$$\frac{\partial f}{\partial a} = -2 \sum_{i=1}^{N} (y_i - a - bx_i - cx_i^2) = 0$$
$$\frac{\partial f}{\partial b} = -2 \sum_{i=1}^{N} (y_i - a - bx_i - cx_i^2)x_i = 0$$
$$\frac{\partial f}{\partial c} = -2 \sum_{i=1}^{N} (y_i - a - bx_i - cx_i^2)x_i^2 = 0$$

Regresión No Lineal

Ajuste Parabólico $Y = a + bX + cX^2$

Por tanto las ecuaciones normales quedan:

$$\left.\begin{array}{l}
\sum_{i=1}^{N} y_i = Na + b\sum_{i=1}^{N} x_i + c\sum_{i=1}^{N} x_i^2 \\
\sum_{i=1}^{N} y_i x_i = a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 + c\sum_{i=1}^{N} x_i^3 \\
\sum_{i=1}^{N} y_i x_i^2 = a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i^3 + c\sum_{i=1}^{N} x_i^4
\end{array}\right\}$$

Para medir la bondad del ajuste se utilizará el coeficiente de determinación

$$R^2 = 1 - \frac{s_e^2}{s_y^2}$$

Regresión No Lineal

Ajuste Parabólico $Y = a + bX + cX^2$

Por tanto las ecuaciones normales quedan:

$$\left.\begin{array}{l}
\sum_{i=1}^{N} y_i = Na + b\sum_{i=1}^{N} x_i + c\sum_{i=1}^{N} x_i^2 \\
\sum_{i=1}^{N} y_i x_i = a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 + c\sum_{i=1}^{N} x_i^3 \\
\sum_{i=1}^{N} y_i x_i^2 = a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i^3 + c\sum_{i=1}^{N} x_i^4
\end{array}\right\}$$

Para medir la bondad del ajuste se utilizará el coeficiente de determinación

$$R^2 = 1 - \frac{s_e^2}{s_y^2}$$

Regresión Lineal Múltiple

Regresión Lineal Múltiple

Aplicaciones de la Regresión

- Explicativo. Determinar que variables explican una variable dada.
- Predictivo. Predecir valores de una variabe a partir de valores conocidos de otras.

Regresión Lineal Múltiple

Aplicaciones de la Regresión

- Explicativo. Determinar que variables explican una variable dada.
- Predictivo. Predecir valores de una variabe a partir de valores conocidos de otras.

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Dando valores a la ecuación del hiperplano tenemos:

$$\left.\begin{array}{l} y_1^* = b_0 + b_1 x_{11} + b_2 x_{21} + \cdots + b_k x_{k1} \\ y_2^* = b_0 + b_1 x_{12} + b_2 x_{22} \cdots + b_k x_{k2} \\ \ldots \\ y_n^* = b_0 + b_1 x_{1n} + b_2 x_{2n} + \cdots + b_k x_{kn} \end{array}\right\}$$

Matricialmente sería $Y^* = XB$ donde

$$Y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \quad B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Dando valores a la ecuación del hiperplano tenemos:

$$\left.\begin{array}{l} y_1^* = b_0 + b_1 x_{11} + b_2 x_{21} + \cdots + b_k x_{k1} \\ y_2^* = b_0 + b_1 x_{12} + b_2 x_{22} \cdots + b_k x_{k2} \\ \cdots \\ y_n^* = b_0 + b_1 x_{1n} + b_2 x_{2n} + \cdots + b_k x_{kn} \end{array}\right\}$$

Matricialmente sería $Y^* = XB$ donde

$$Y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \quad B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Si A es una matriz denoteremos por A^t la matriz traspuesta de A. La matriz de errores será:

$$E = Y - Y^* = \begin{pmatrix} e_1 = y_1 - y_1^* \\ e_2 = y_2 - y_2^* \\ \vdots \\ e_n = y_n - y_n^* \end{pmatrix}$$

De nuevo nuestro objetivo es minimizar

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_i^*)^2 = E^t E$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Si A es una matriz denoteremos por A^t la matriz traspuesta de A. La matriz de errores será:

$$E = Y - Y^* = \begin{pmatrix} e_1 = y_1 - y_1^* \\ e_2 = y_2 - y_2^* \\ \vdots \\ e_n = y_n - y_n^* \end{pmatrix}$$

De nuevo nuestro objetivo es minimizar

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_i^*)^2 = E^t E$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Por otro lado

$$E = Y - XB$$

luego tenemos que minimizar

$$E^t E = (Y - XB)^t (Y - XB) = Y^t Y - 2Y^t XB + B^t X^t XB$$

y derivando e igualando a cero tenemos:

$$\frac{\partial E^t E}{\partial B} = -2X^t Y + 2X^t XB = 0$$

de donde deducimos que

$$X^t XB = X^t Y$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Por otro lado

$$E = Y - XB$$

luego tenemos que minimizar

$$E^t E = (Y - XB)^t (Y - XB) = Y^t Y - 2Y^t XB + B^t X^t XB$$

y derivando e igualando a cero tenemos:

$$\frac{\partial E^t E}{\partial B} = -2X^t Y + 2X^t XB = 0$$

de donde deducimos que

$$X^t XB = X^t Y$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Por tanto si existe la matriz inversa de X^tX tenemos que:

$$B = (X^tX)^{-1}X^tY$$

Regresión Lineal Múltiple

Ajuste a un Hiperplano por Mínimos Cuadrados

Por tanto si existe la matriz inversa de X^tX tenemos que:

$$B = (X^tX)^{-1}X^tY$$

Coeficiente de Correlación Multiple

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple

Con la misma filosofía que en el caso de la regresión lineal definimos el coeficiente de correlación múltiple como:

$$R^2_{y,12\ldots k} = \frac{s^2_{y^*}}{s^2_y} = 1 - \frac{s^2_e}{s^2_y}$$

además se verifica que

$$0 \leq R^2_{y,12\ldots k} \leq 1$$

siendo mejor el ajuste cuando mas cercano esté a 1.

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple

Con la misma filosofía que en el caso de la regresión lineal definimos el coeficiente de correlación múltiple como:

$$R^2_{y,12\ldots k} = \frac{s^2_{y^*}}{s^2_y} = 1 - \frac{s^2_e}{s^2_y}$$

además se verifica que

$$0 \leq R^2_{y,12\ldots k} \leq 1$$

siendo mejor el ajuste cuando mas cercano esté a 1.

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple en Forma Matricial

Como $E = Y - XB$ y $B = (X^tX)^{-1}X^tY$ tenemos que

$$s_e^2 = \frac{1}{n}(Y^tY - B^tX^tY)$$

$$\left(\textbf{Dem:} \quad \begin{array}{l} ns_e^2 = E^tE = (Y-XB)^t(Y-XB) = Y^tY - B^tX^t - Y^tXB + B^tX^tXB = \\ = Y^tY - B^tX^tY - Y^tXB + (X^tX)^{-1}X^tY^tX^tXB = \\ = Y^tY - B^tX^tY - Y^tXB + Y^tXB = Y^tY - B^tX^tY. \end{array} \right)$$

Como también $\overline{y} = \overline{y}^*$ tenemos que

$$s_{y^*}^2 = \frac{1}{n}B^tX^tY - \overline{y}^2$$

$$\left(\textbf{Dem:} \quad \begin{array}{l} ns_{y^*}^2 = \sum_{i=1}^n (y_i^* - \overline{y}^*)^2 = \sum_{i=1}^n y_i^{*2} - n\overline{y}^2 = Y^{*t}Y^* - n\overline{y}^2 = (XB)^t(XB) - n\overline{y}^2 = \\ B^tX^tXB - n\overline{y}^2 = B^tX^tX(X^tX)^{-1}X^tY - n\overline{y}^2 = B^tX^tY - n\overline{y}^2. \end{array} \right)$$

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple en Forma Matricial

Como $E = Y - XB$ y $B = (X^tX)^{-1}X^tY$ tenemos que

$$s_e^2 = \frac{1}{n}(Y^tY - B^tX^tY)$$

$$\left(\textbf{Dem:} \quad \begin{array}{l} ns_e^2 = E^tE = (Y-XB)^t(Y-XB) = Y^tY - B^tX^t - Y^tXB + B^tX^tXB = \\ = Y^tY - B^tX^tY - Y^tXB + (X^tX)^{-1}X^tY^tX^tXB = \\ = Y^tY - B^tX^tY - Y^tXB + Y^tXB = Y^tY - B^tX^tY. \end{array} \right)$$

Como también $\overline{y} = \overline{y}^*$ tenemos que

$$s_{y^*}^2 = \frac{1}{n}B^tX^tY - \overline{y}^2$$

$$\left(\textbf{Dem:} \quad \begin{array}{l} ns_{y^*}^2 = \sum_{i=1}^{n}(y_i^* - \overline{y}^*)^2 = \sum_{i=1}^{n} y_i^{*2} - n\overline{y}^2 = Y^{*t}Y^* - n\overline{y}^2 = (XB)^t(XB) - n\overline{y}^2 = \\ B^tX^tXB - n\overline{y}^2 = B^tX^tX(X^tX)^{-1}X^tY - n\overline{y}^2 = B^tX^tY - n\overline{y}^2. \end{array} \right)$$

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple en Forma Matricial

Además

$$s_y^2 = \frac{1}{n} Y^t Y - \overline{y}^2$$

$\left(\textbf{Dem:} \quad ns_y^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = Y^t Y - n\overline{y}^2. \right)$

Luego

$$\mathbf{R^2 = 1 - \frac{s_e^2}{s_y^2} = 1 - \frac{Y^tY - B^tX^tY}{Y^tY - n\overline{y}^2} = \frac{B^tX^tY - n\overline{y}^2}{Y^tY - n\overline{y}^2}}.$$

Regresión Lineal Múltiple

Coeficiente de Correlación Multiple en Forma Matricial

Además

$$s_y^2 = \frac{1}{n} Y^t Y - \overline{y}^2$$

$\left(\textbf{Dem:} \quad ns_y^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = Y^t Y - n\overline{y}^2. \right)$

Luego

$$\mathbf{R^2 = 1 - \frac{s_e^2}{s_y^2} = 1 - \frac{Y^tY - B^tX^tY}{Y^tY - n\overline{y}^2} = \frac{B^tX^tY - n\overline{y}^2}{Y^tY - n\overline{y}^2}}.$$

Coeficiente de Correlación Parcial

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

Problema

Estudiar el grado de asociación lineal entre las variables Y y X_i.

Solución 1

Una posible solución sería calculando el coeficiente de correlacion lineal

$$r = \frac{S_{YX_i}}{s_y s_{x_i}}$$

Pero de esta manera no se tiene en cuenta que el grado de correlación obtenido pueda ser fruto de la influencia que ejercen el resto de variables $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

Problema

Estudiar el grado de asociación lineal entre las variables Y y X_i.

Solución 1

Una posible solución sería calculando el coeficiente de correlacion lineal

$$r = \frac{S_{YX_i}}{s_y s_{x_i}}$$

Pero de esta manera no se tiene en cuenta que el grado de correlación obtenido pueda ser fruto de la influencia que ejercen el resto de variables $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

Solución 2

Eliminar la influencia de $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$

¿Cómo?

En tres pasos.

1.- Obtenniendo los hiperplanos de regresión:

$$\left.\begin{array}{l} \widehat{Y} = c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_{i-1} X_{i-1} + c_{i+1} X_{i+1} + \cdots + c_k X_k \\ \widehat{X}_i = d_0 + d_1 X_1 + d_2 X_2 + \cdots + d_{i-1} X_{i-1} + d_{i+1} X_{i+1} + \cdots + d_k X_k \end{array}\right\}$$

que resumen la influencia que ejercen las variables $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$ sobre Y y X_i respectivamente.

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

Solución 2

Eliminar la influencia de $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$

¿Cómo?

En tres pasos.

1.- Obtenniendo los hiperplanos de regresión:

$$\left.\begin{aligned} \widehat{Y} &= c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_{i-1} X_{i-1} + c_{i+1} X_{i+1} + \cdots + c_k X_k \\ \widehat{X}_i &= d_0 + d_1 X_1 + d_2 X_2 + \cdots + d_{i-1} X_{i-1} + d_{i+1} X_{i+1} + \cdots + d_k X_k \end{aligned}\right\}$$

que resumen la influencia que ejercen las variables $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$ sobre Y y X_i respectivamente.

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

2.- Se definen las variables:

$$\left.\begin{array}{l} U = Y - \widehat{Y} \\ \\ V = X_i - \widehat{X}_i \end{array}\right\}$$

que incorporan aquella parte de Y y X_i respectivamente, que queda libre de la influencia de $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$

3.- Se define el coeficiente de correlación parcial entre Y y X_i como la correlación simple entre U y V:

$$\mathbf{r_{yi\cdot 12\ldots i-1i+1\ldots k} = \frac{S_{uv}}{s_u s_v}}$$

Regresión Lineal Múltiple

Coeficiente de Correlación Parcial

2.- Se definen las variables:

$$\left.\begin{aligned} U &= Y - \widehat{Y} \\ V &= X_i - \widehat{X}_i \end{aligned}\right\}$$

que incorporan aquella parte de Y y X_i respectivamente, que queda libre de la influencia de $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$

3.- Se define el coeficiente de correlación parcial entre Y y X_i como la correlación simple entre U y V:

$$\mathbf{r_{yi\cdot 12\ldots i-1 i+1\ldots k}} = \frac{\mathbf{S_{uv}}}{\mathbf{s_u s_v}}$$

El Problema de la Multicolinealidad

Regresión Lineal Múltiple

El Problema de la Multicolinealidad

La multicolinealidad ocurre cuando existe una correlacion entre dos o más variables explicativas.
Ésto implica que en la matriz X existen columnas que son combinación lineal de otras.
Por esta razón la matriz (X^tX) no será invertible y por tanto el vector de coeficientes $B = (X^tX)^{-1}X^tY$ no se puede determinar.

Regresión Lineal Múltiple

El Problema de la Multicolinealidad

La multicolinealidad ocurre cuando existe una correlacion entre dos o más variables explicativas.
Ésto implica que en la matriz X existen columnas que son combinación lineal de otras.
Por esta razón la matriz $(X^t X)$ no será invertible y por tanto el vector de coeficientes $B = (X^t X)^{-1} X^t Y$ no se puede determinar.

Bibliografía

Bobliografía

- GARCÍA CÓRDOBA J. A. , LÓPEZ HERNÁNDEZ F. A., PALACIOS SÁNCHEZ Mª Á. y RUIZ MARÍN, M. (2000), *Introducción a la Estadística para la Empresa*. Horacio Escarabajal Editores, pp 95–124.
- MARTÍN PLIEGO LÓPEZ, F.J. (2004), *Introdución a la Estadística Económica Y Empresarial*. Ed. Prentice Hall. pp. 235–372.
- MONTIEL A.M., RIUS F. y BARÓN F.J., (1997), *Elementos Básicos De Estadística Económica Y Empresarial*. Ed. Prentice Hall. pp. 147–186.
- NOVALES, A., (1996), *Estadística y Econometría*, Madrid: Mc Graw-Hill, pp.475-516.
- SANZ J.A.; BEDATE, A.; RIVAS, A. y GONZÁLEZ, J., (1996), *Problemas De Estadística Descriptiva Empresarial*. Ed. Ariel Economía., pp. 131–284.