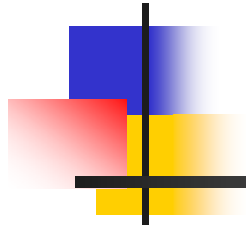




UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE MEDICINA
DEPARTAMENTO ACADÉMICO DE MEDICINA PREVENTIVA
Y SALUD PÚBLICA
Sección de Estadística y Epidemiología



CURSO: Bioestadística
EAP Enfermería 2006-II



Tema 14.- CORRELACIÓN Y REGRESIÓN

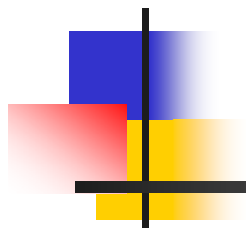
Dr Alvaro Withemburry Vlásica
Lima, Noviembre 2006



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE MEDICINA
DEPARTAMENTO ACADÉMICO DE MEDICINA PREVENTIVA
Y SALUD PÚBLICA
Sección de Estadística y Epidemiología



CURSO: Bioestadística
EAP Enfermería 2006-II



Tema 14.-
CORRELACIÓN Y REGRESIÓN

Dr. Álvaro Withemburry Vlásica
Lima, Noviembre 2006



Correlación y Regresión Lineales

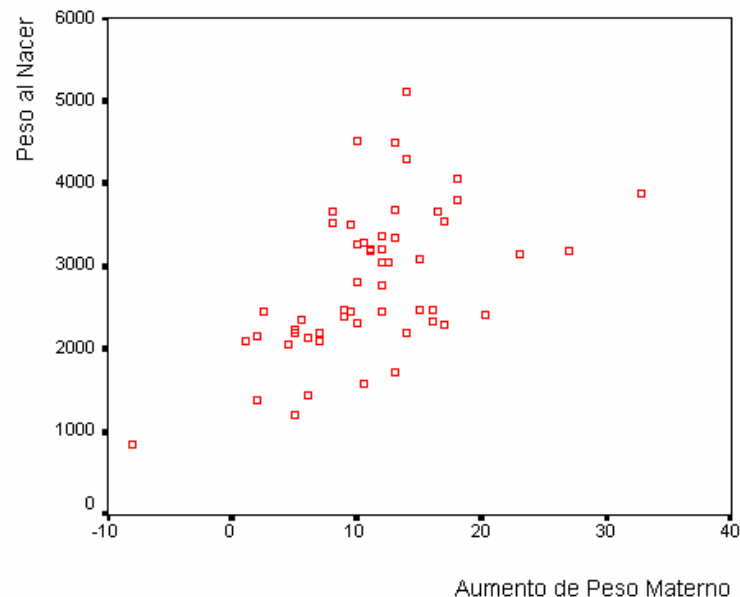
Cuando se tiene interés en analizar la relación entre dos variables cuantitativas como por ejemplo:

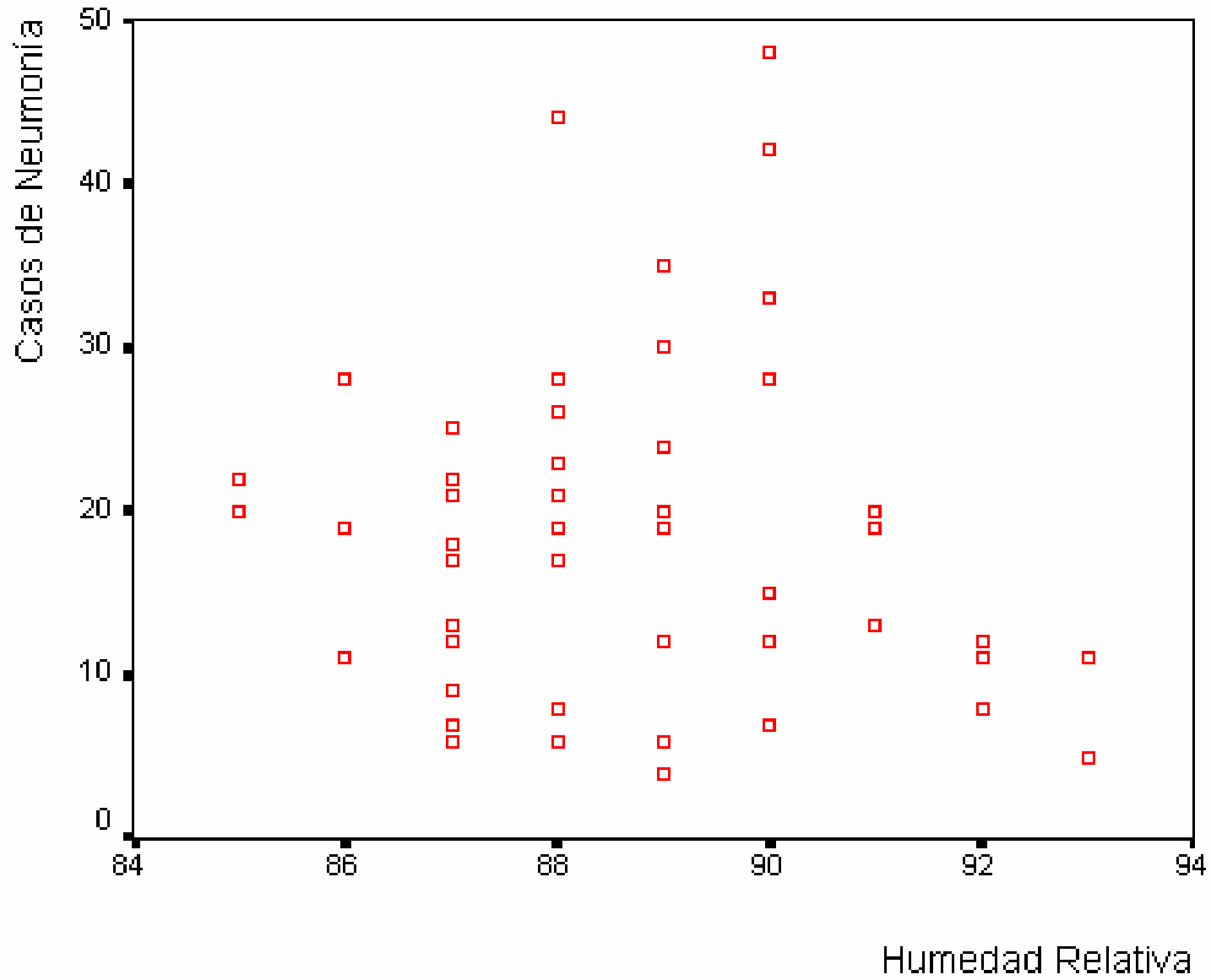
- **presión sanguínea y edad,**
- **estatura y peso,**
- **el ingreso familiar y los gastos médicos,**
- **los niveles de pobreza y las tasas de mortalidad infantil.**

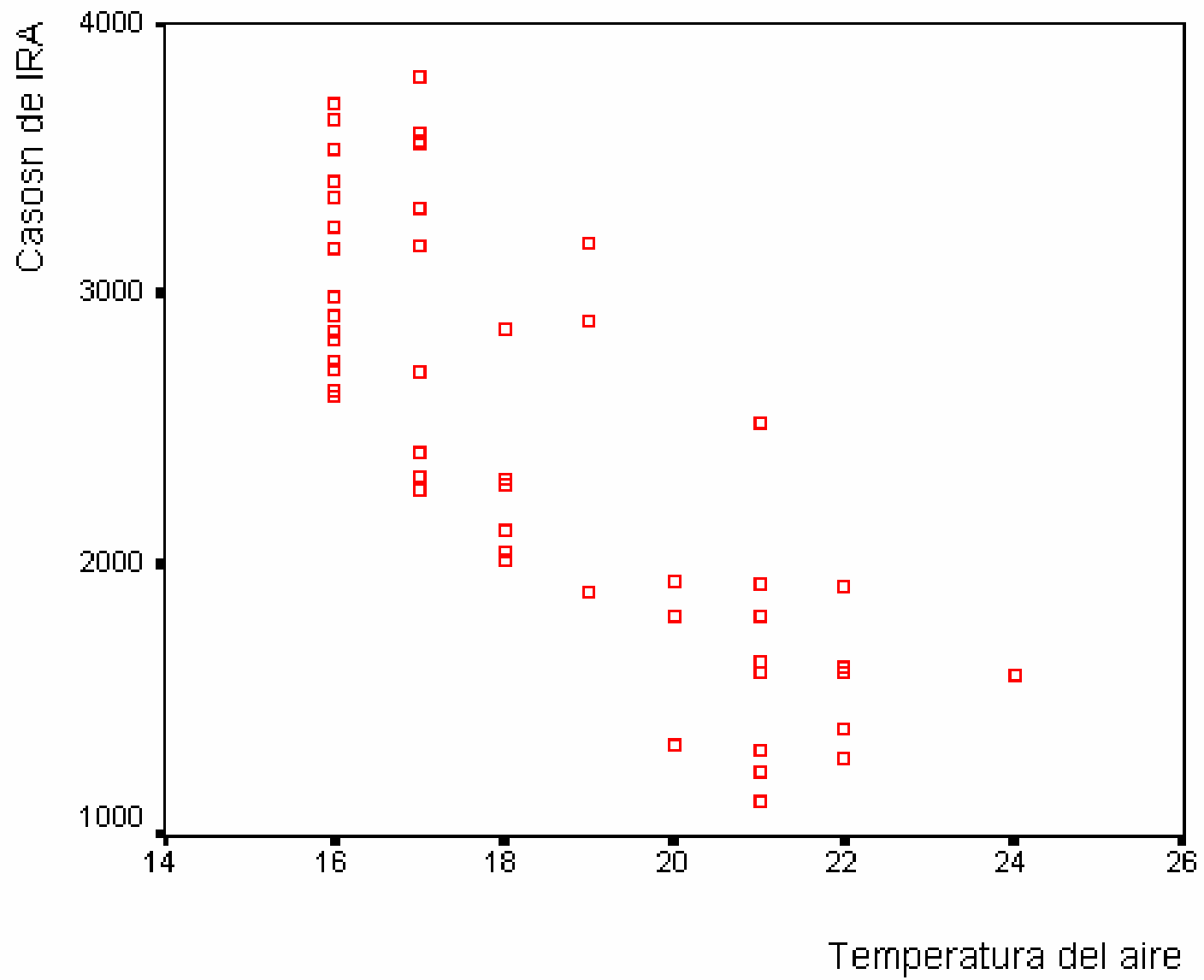
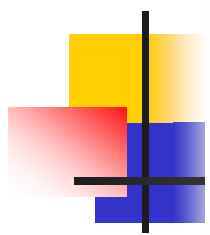
La naturaleza y la intensidad de la relación entre este tipo de variables puede ser examinada por medio del análisis de **CORRELACIÓN y REGRESIÓN LINEALES.**

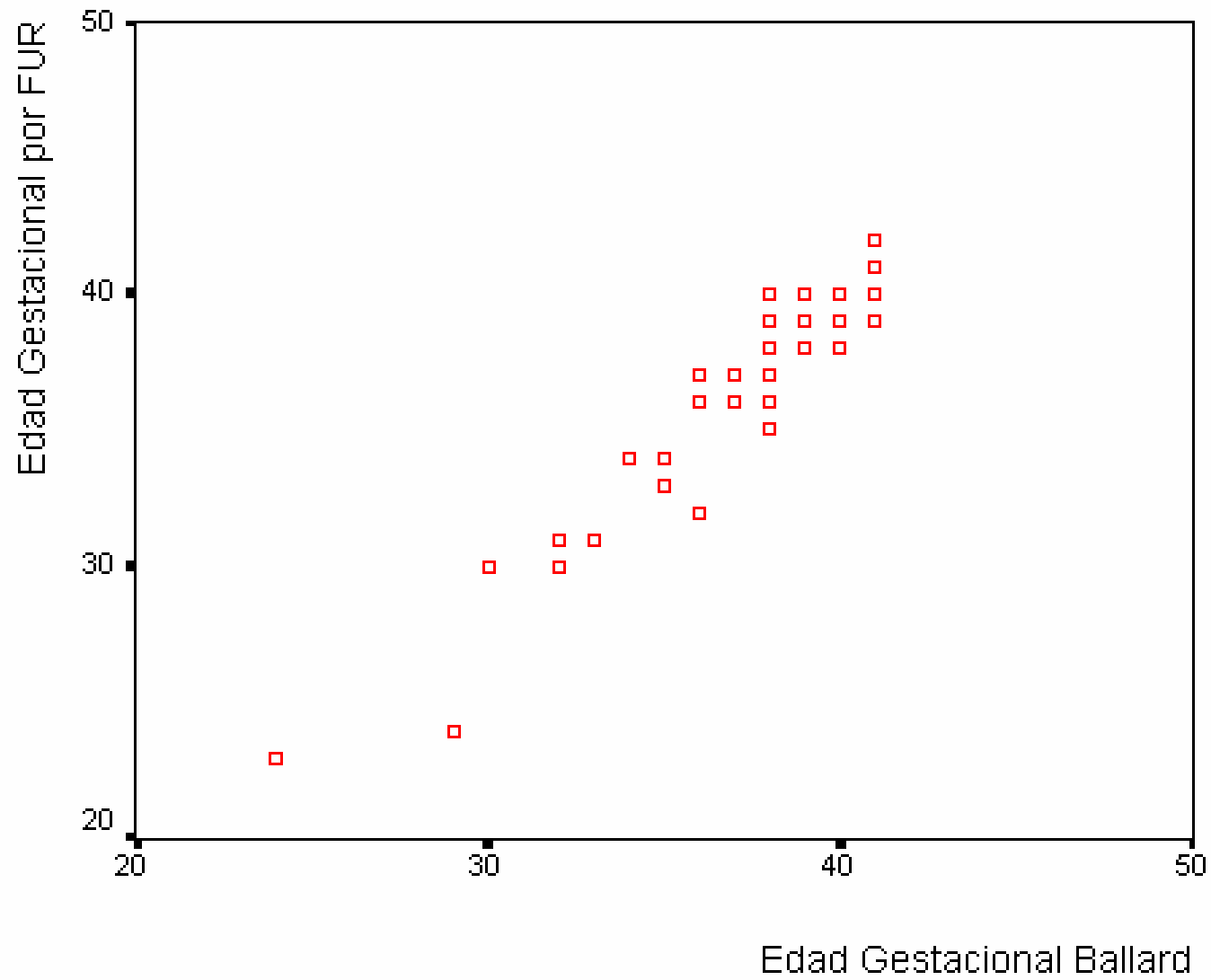
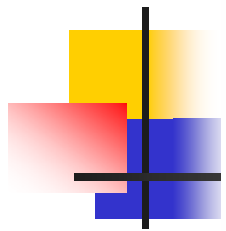
Correlación Lineal Simple

Tiene por objetivo medir y evaluar el grado de asociación lineal entre dos variables CUANTITATIVAS (medición de la intensidad de la relación). Una manera de visualizar la (posible) correlación entre las observaciones de dos variables X e Y, es a través de un diagrama de dispersión, en el cual los valores que toman estas variables son representados por puntos.











Medida de Correlación Lineal Simple

Se realiza a través del Coeficiente de Correlación Lineal de Pearson (r).

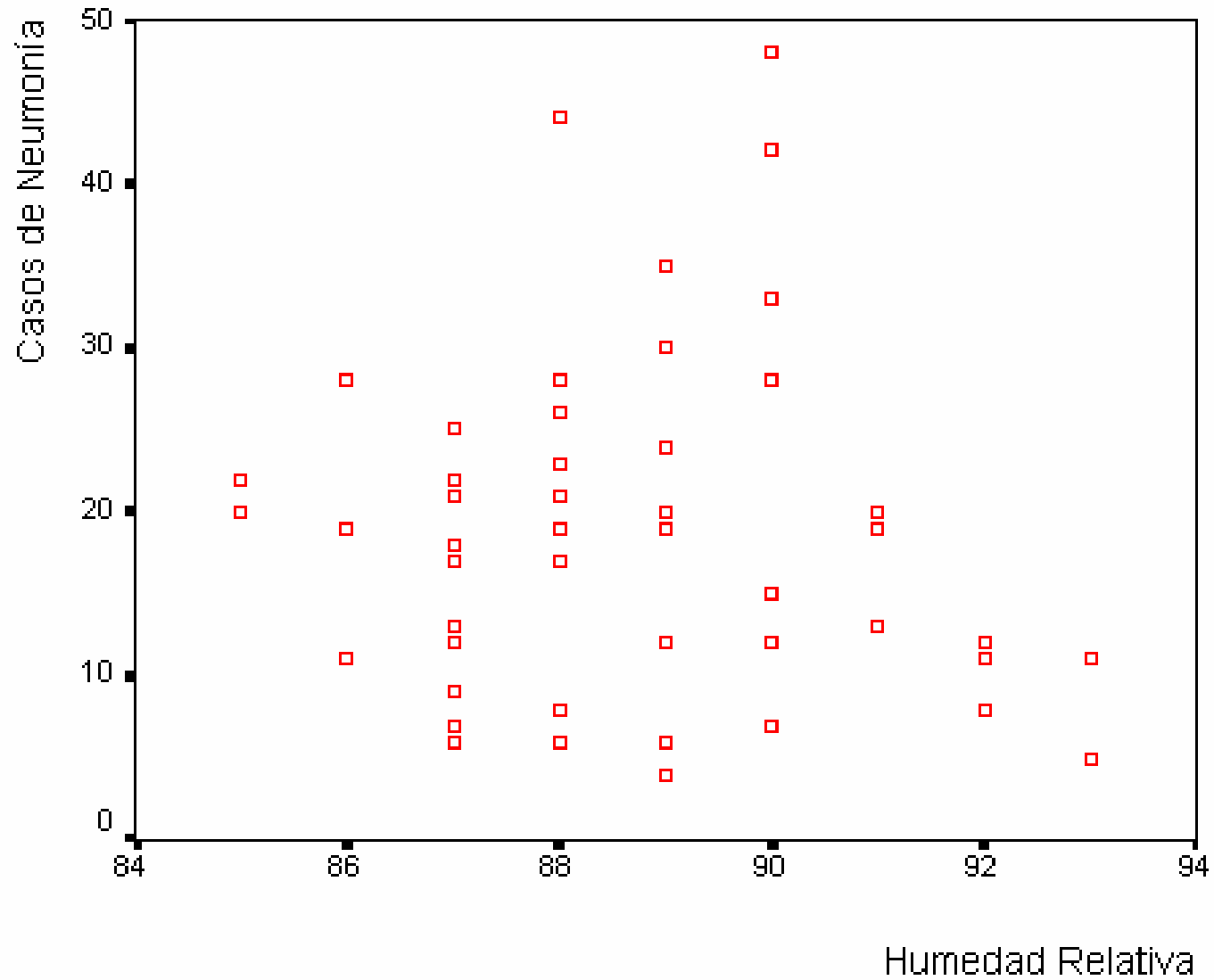
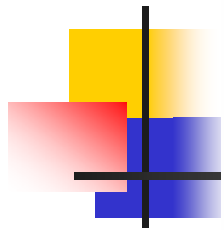
Se utiliza cuando las variables de estudio son cuantitativas medidas en escala de intervalo o razón.

El Coeficiente puede tomar cualquier valor entre -1 y 1

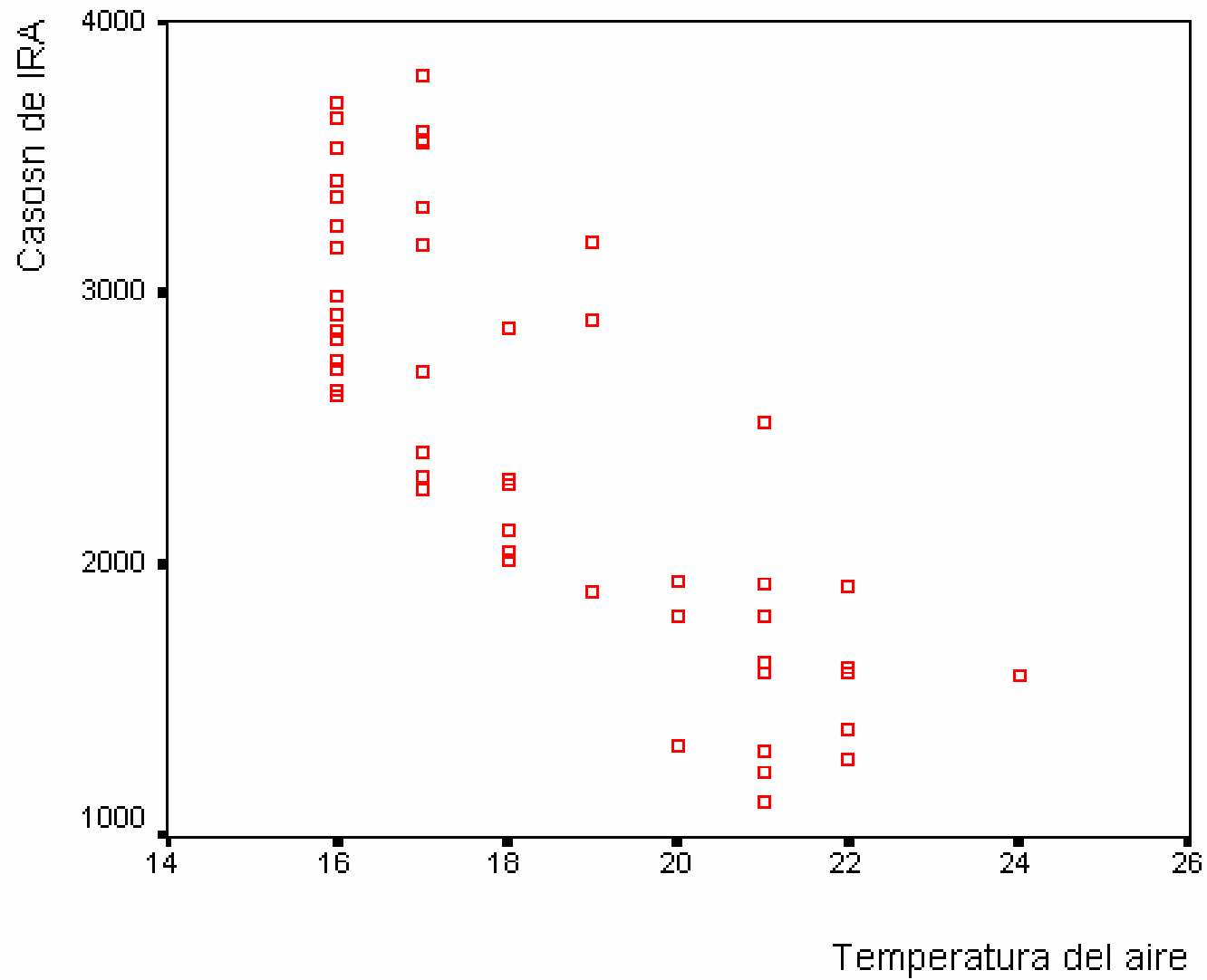
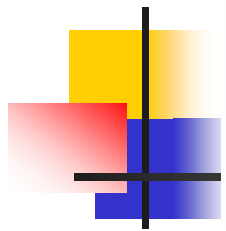
$r = 1$ Indica correlación lineal directa perfecta

$r = -1$ Indica correlación lineal inversa perfecta

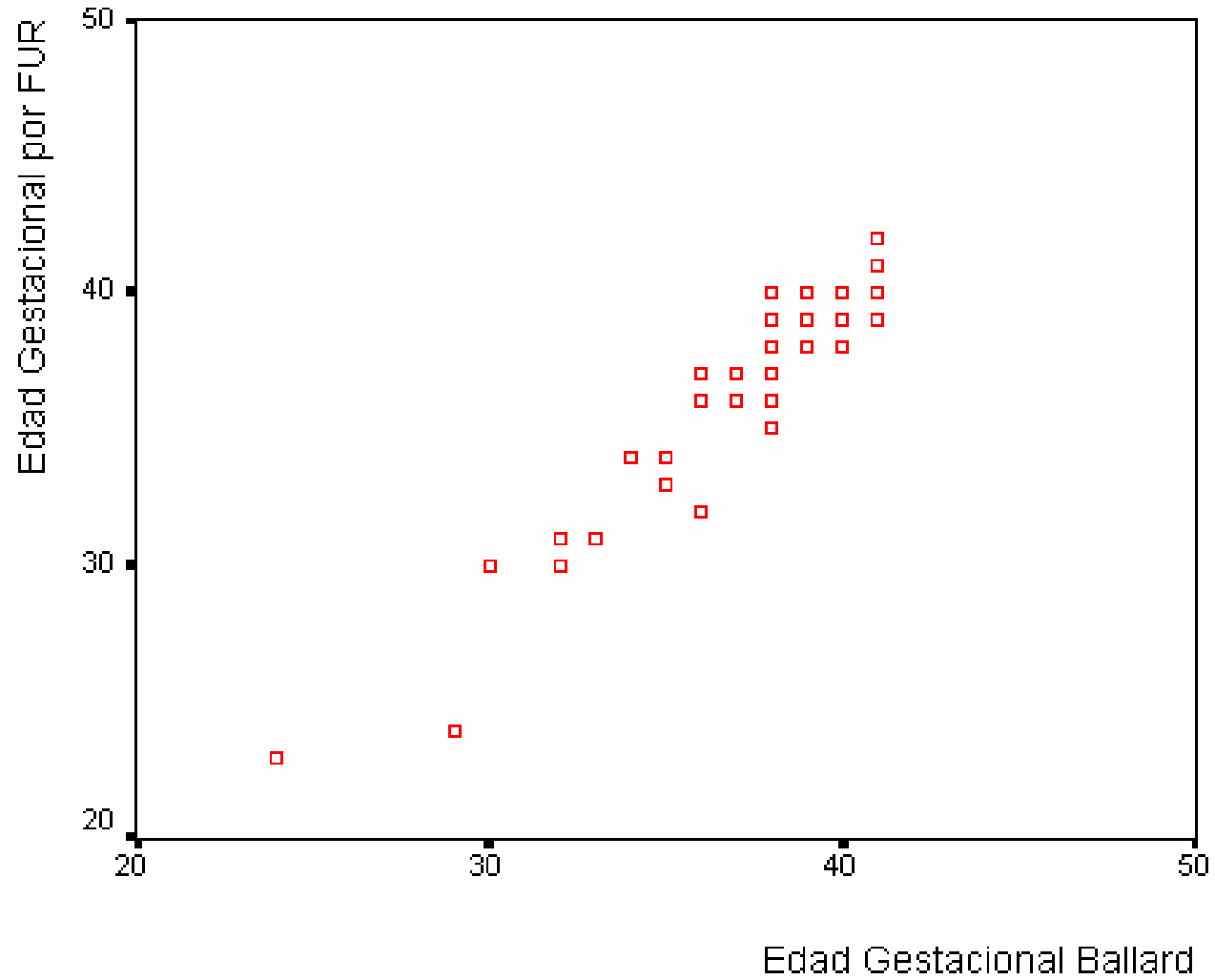
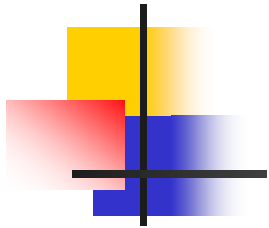
$r = 0$ Indica que las dos variables no están asociadas



$$r = -0.067$$



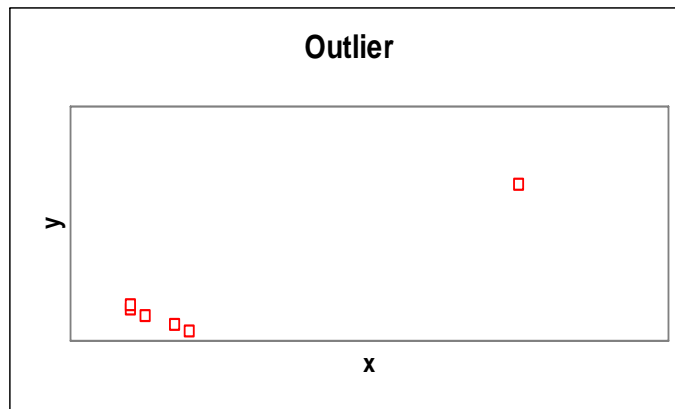
$$r = -0.798$$



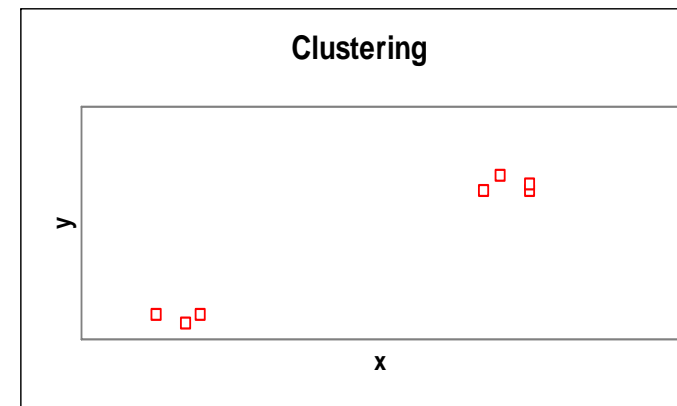
$$r = 0.955$$

Desventajas

- No puede ser usada para hacer predicciones ya que no nos dice nada acerca de la forma que toma la relación.
- En algunos casos los resultados del coeficiente de correlación pueden ser engañosos:

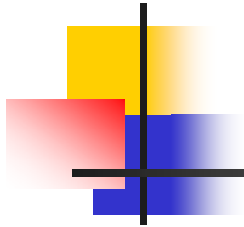


$$r = 0.97$$



$$r = 0.99$$

Regresión Lineal Simple

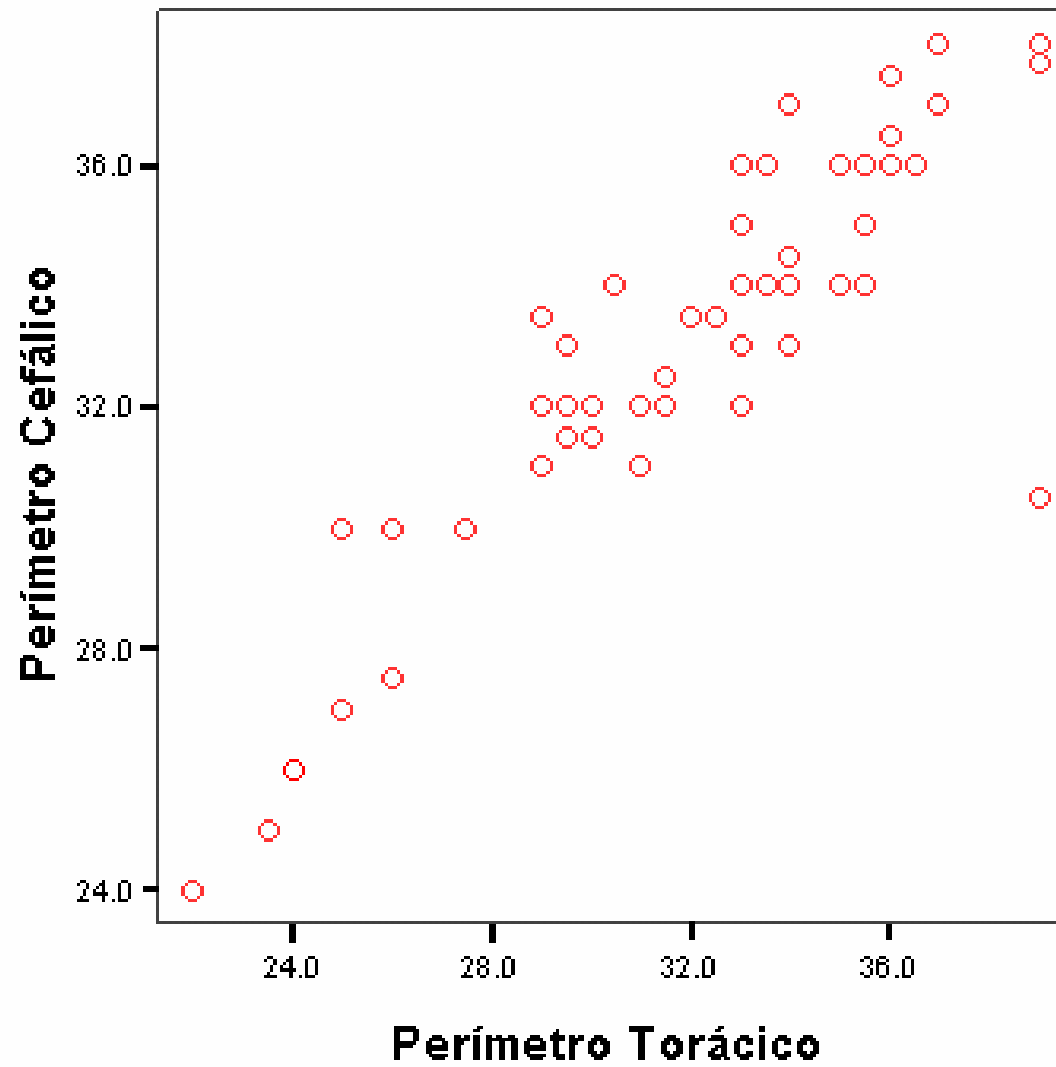
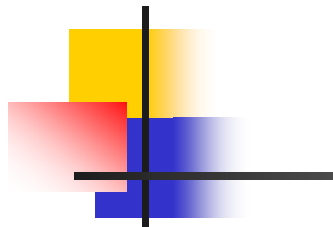


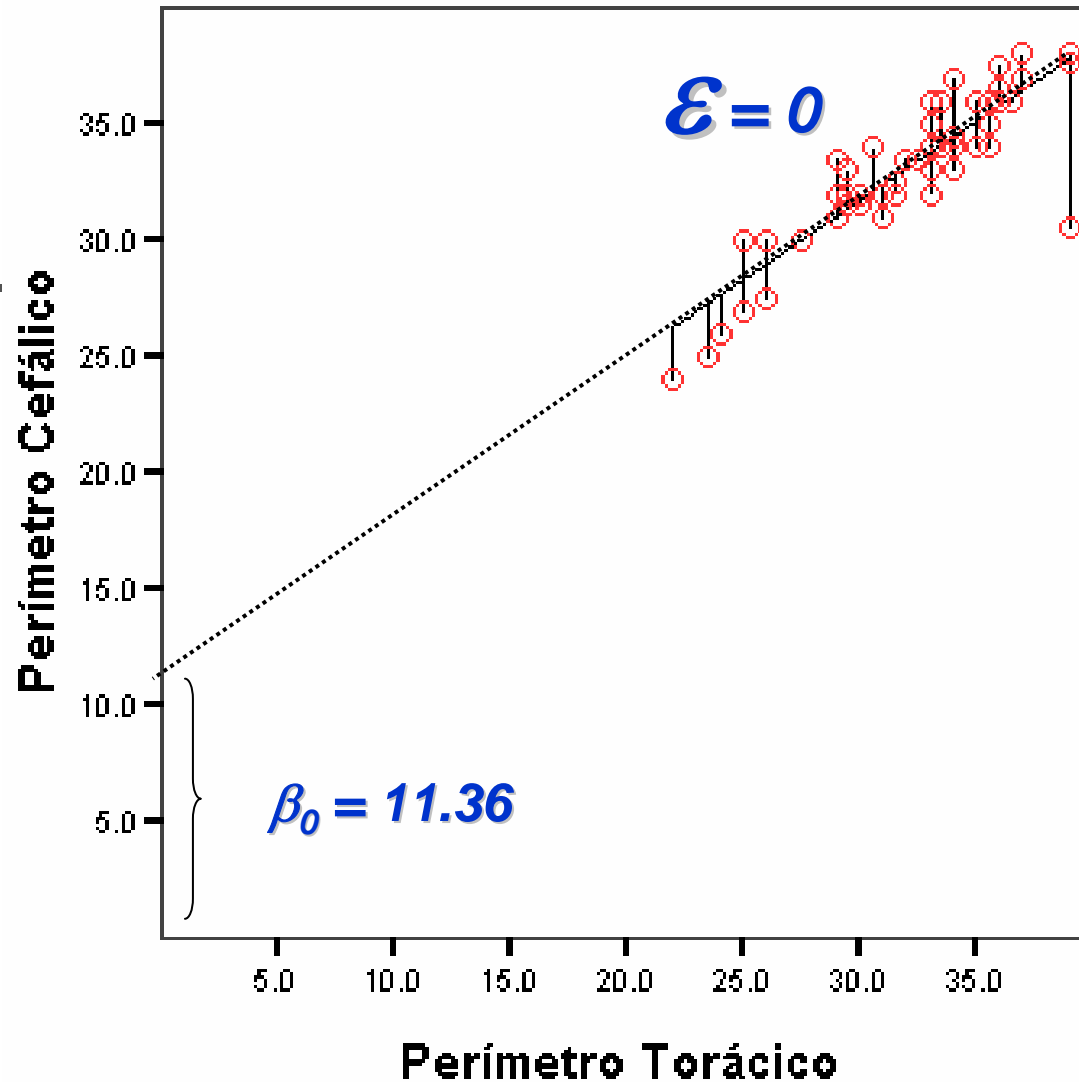
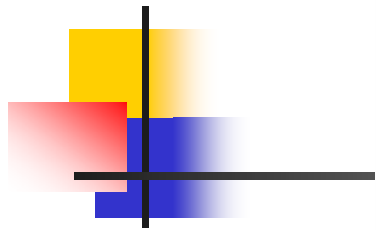
Estima la forma o naturaleza de la relación entre dos variables cuantitativas, siendo su objetivo final el PREDECIR o ESTIMAR el valor de una variable en función al valor dado de otra variable.

Es una regresión lineal ya que la relación X-Y está modelada para seguir una línea recta; y es simple ya que hay una sola variable X.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$$y = b_0 + b_1 x$$





$$\text{Perímetro Cefálico} = 11.36 + 0.68(\text{Perímetro Torácico})$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

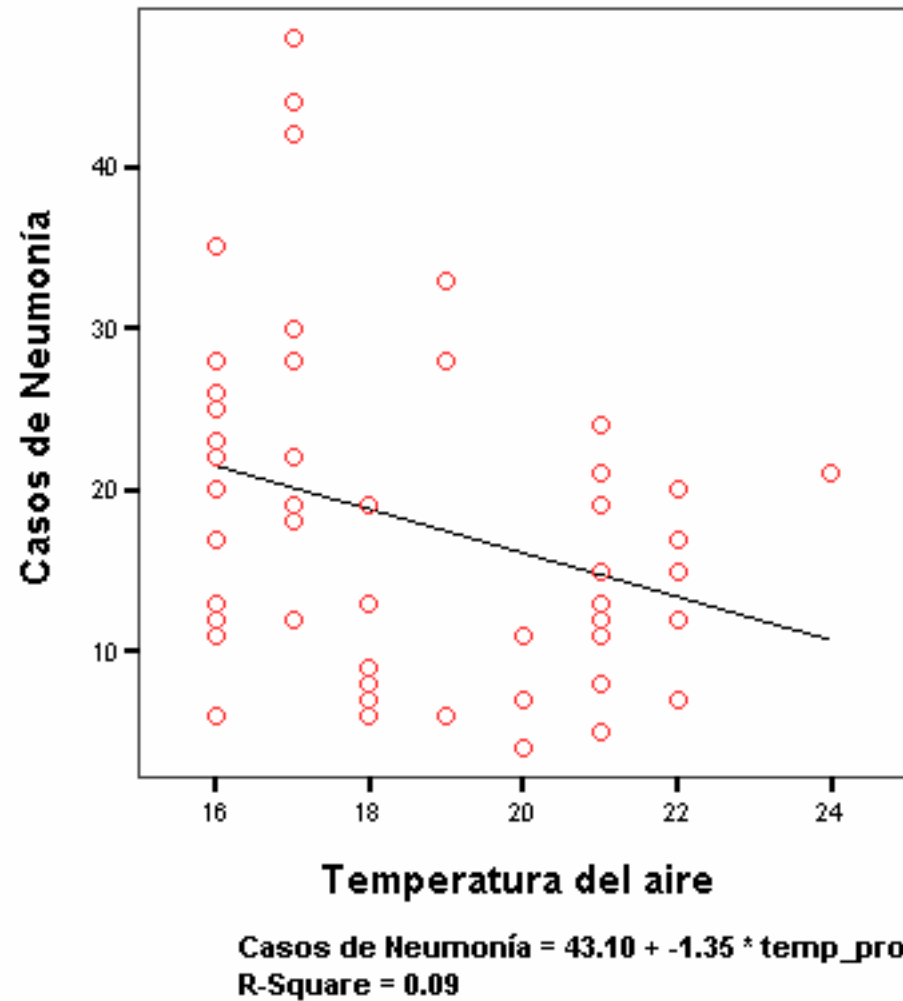
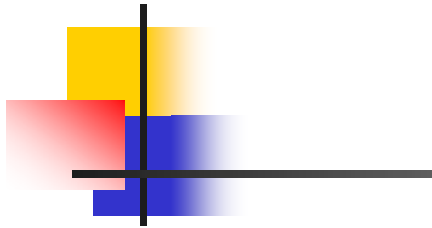


Coeficiente de Determinación - r^2

Mide la proporción o porcentaje de la variación total de Y que es explicada por el modelo de regresión.

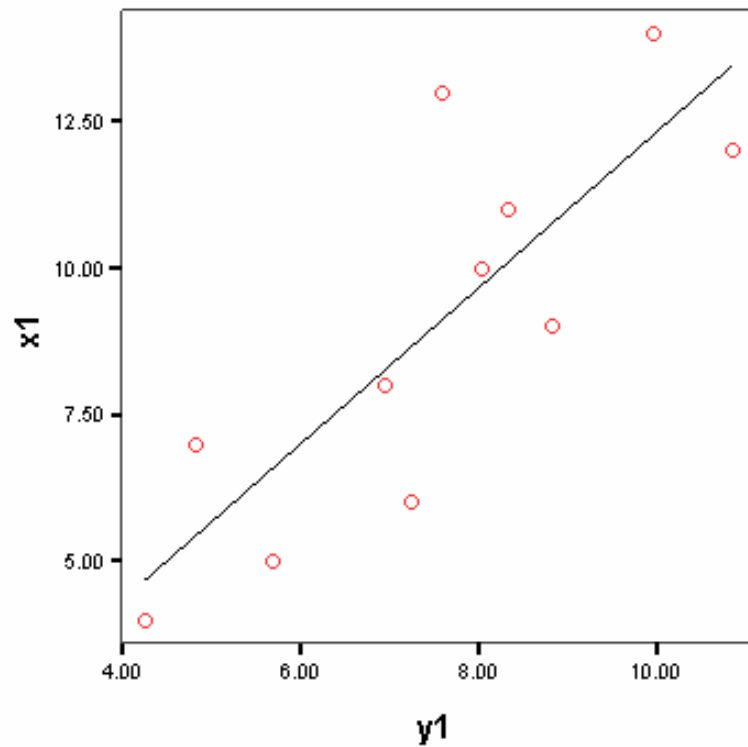
$$0 \leq r^2 \leq 1$$

Cuanto más se aproxime a 1, “mejor” será el ajuste del modelo a los puntos del diagrama de dispersión. Mide que tan exitoso es el modelo.

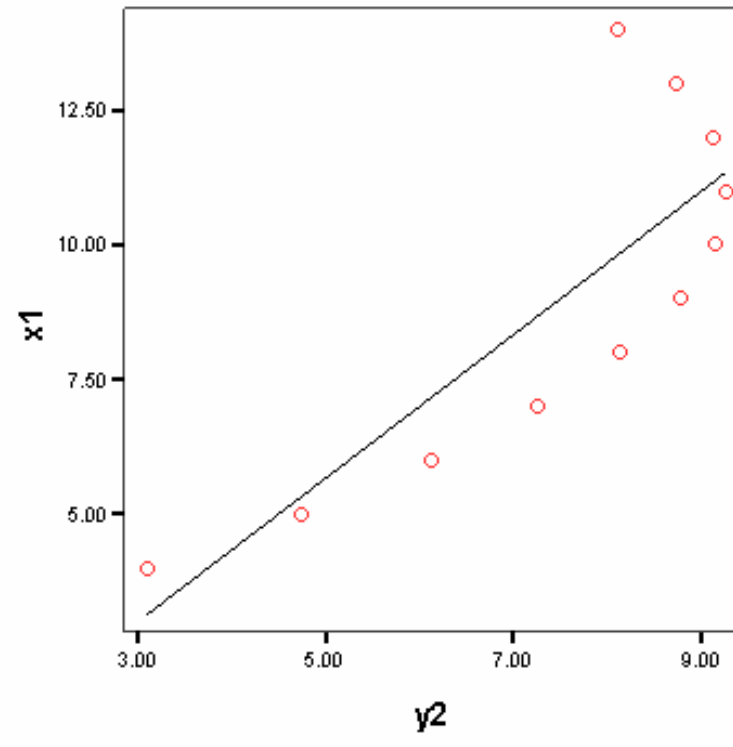


$$\text{Casos de Neumonía} = 43.1 + -1.35(\text{Temperatura del aire})$$

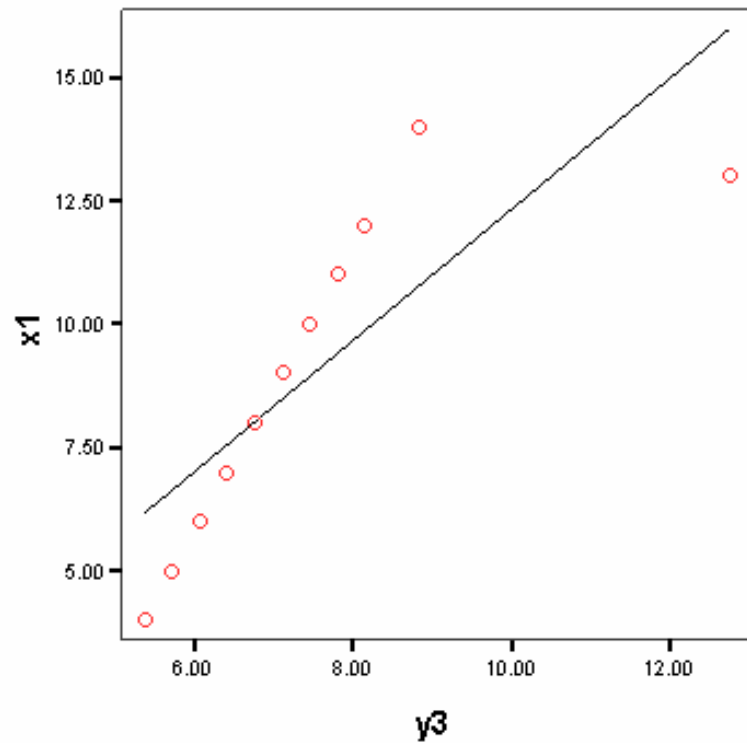
$$r = -0.3 \quad r^2 = 0.09$$



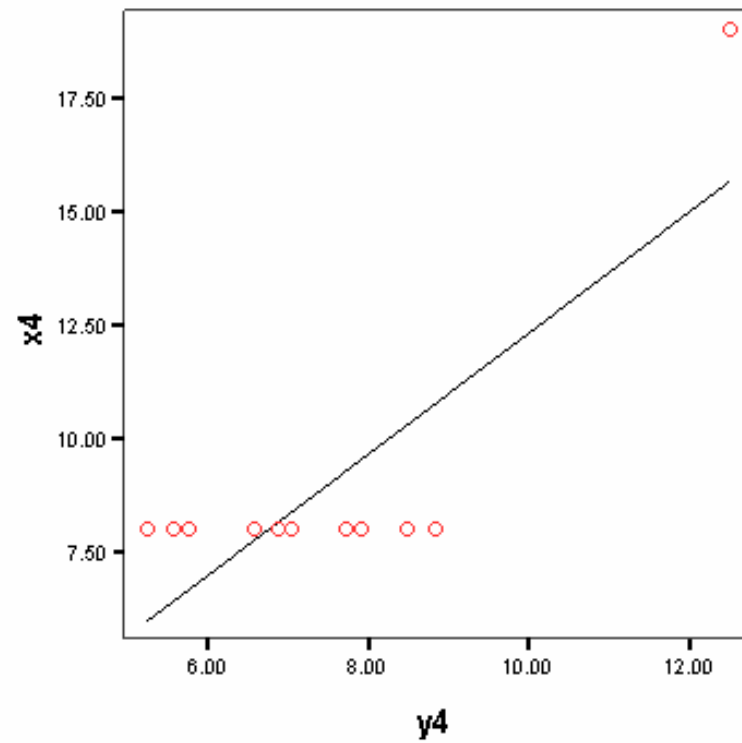
$x_1 = -1.00 + 1.33 * y_1$
 R-Square = 0.67



$x_1 = -0.99 + 1.33 * y_2$
 R-Square = 0.67



$x_1 = -1.00 + 1.33 * y_3$
 R-Square = 0.67



$x_4 = -1.00 + 1.33 * y_4$
 R-Square = 0.67



Suposiciones de la Regresión Lineal

- 1. Los valores de la variable independiente X son fijos (no aleatorios), medida sin error.**
- 2. La variable Y es aleatoria.**
- 3. Para cada valor de X , existe una población de valores de Y con distribución normal.**
- 4. Las varianzas de las poblaciones Y son todas iguales.**
- 5. Todas las medias de las poblaciones de Y están sobre la recta.**
- 6. Los valores de Y están normalmente distribuidos y son estadísticamente independientes.**



Bibliografía

1. **Daniel W. Bioestadística. Base para el análisis de las ciencias de la salud. 2002. Limusa SA. México DF.**
2. **Woodward M. Epidemiology, study design and data analysis. 1999. Chapman & Hall/CRC. USA.**