



Fundación

J. L. Castaño

Para el desarrollo del Laboratorio Clínico

ESTADÍSTICA BÁSICA APLICADA AL LABORATORIO CLÍNICO

Ed Cont Lab Clín; 30: 37 - 41

SEQC^{ML}

Sociedad Española de Medicina de Laboratorio

2016-2017

ASOCIACIÓN DE VARIABLES CUANTITATIVAS.

Inmaculada Pérez de Algaba Fuentes.

UGC Intercentros de Laboratorio. Hospital Regional Universitario de Málaga. Málaga.

Ana Peña Cabia.

Servicio de Análisis Clínicos. Hospital Virgen de la Luz. Cuenca.

En este tema vamos a abordar la regresión lineal. Ya hemos visto como asociar variables cualitativas con cualitativas, así como cualitativas con cuantitativas. Las variables cuantitativas las manejamos con mucha frecuencia en el laboratorio clínico, ya que muchas de las magnitudes que procesamos las expresamos en forma de valores numéricos. La asociación de variables cuantitativas la haremos mediante la regresión lineal. Si queremos ver cómo están asociadas dos variables cuantitativas, lo haremos mediante la regresión lineal univariante.

El grado de asociación, o lo que es lo mismo, lo que puede explicar el aumento o disminución de una variable respecto a la variación de la otra, se mide mediante el coeficiente de correlación, el cual puede tomar valores entre 1 y -1. Si el coeficiente de correlación (R) es positivo, quiere decir que si una variable aumenta de valor, la otra también. Por el contrario, el aumento de una variable se asocia con la disminución de la otra cuando el coeficiente de correlación es negativo.

Si el coeficiente de correlación toma el valor de 1 o de -1 implica que el aumento de una variable depende exclusivamente del valor de la otra. Por el contrario, si el valor que toma el coeficiente de correlación es 0 quiere decir que el aumento o disminución de la variable no explica cualquier cambio en la otra variable.

La regresión lineal univariante se expresa matemáticamente mediante la siguiente fórmula:

$$y = a + bx$$

Siendo "a", la ordenada en el origen y "b" la pendiente de la recta. La ordenada en el origen nos indica en qué punto la recta de regresión corta el eje "y". Si toma valor cero, la recta pasa por el punto donde el eje "x" corta al eje "y". Si es positivo, corta al eje "y" por encima del eje "x", y, si es negativo corta al eje "y" por debajo del eje "x". La pendiente nos da una idea de la inclinación de la recta. Si es positiva, la recta es ascendente y, si es negativa, la recta es descendente.

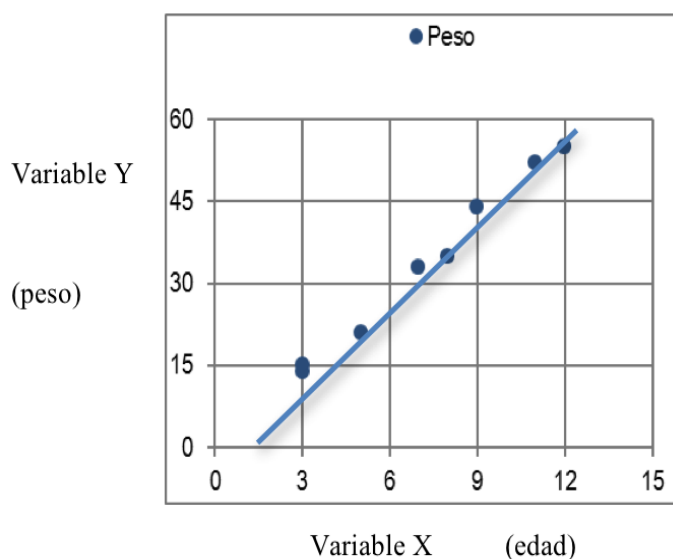
A las variables de interés en regresión lineal se denominan "x" e "y". La variable "x" suele ser la variable independiente o covariable y la variable "y" es la variable dependiente o variable resultado. La variable "x" se suele representar en el eje de abscisas (x) y la variable "y" en el eje de ordenadas (y). La representación gráfica se realiza mediante puntos de coordenadas (x, y). Los datos se suelen recoger en una tabla de doble entrada, donde a cada valor de "x" corresponde un valor de "y".

Para entender mejor lo que hemos comentado vamos a recurrir a un ejemplo:

En un colegio se eligieron, al azar, ocho niños de entre 3 y 12 años de edad. Se pesó a cada uno de ellos y se anotó el peso en kilogramos. A partir de estos datos podemos caracterizar una ecuación que nos permita ver la relación que existe entre ambas variables (peso y edad).

En primer lugar, creamos la tabla de doble entrada:

Edad (Variable X)	Peso (Variable Y)
3	14
3	15
5	21
7	33
8	35
9	44
11	52
12	55



Una vez que hemos construido nuestra tabla, las representamos de tal manera que a cada valor de "x" en el eje de abscisas, le asignamos su correspondiente valor de "y" en el eje de ordenadas y lo representamos por puntos.

Ahora podríamos trazar una línea que pasará por el mayor número de puntos y dejará, en la medida de lo posible, el mismo número de puntos a un lado y otro de la recta:

Como ya hemos comentado, la relación entre estas variables viene "cuantificada" por el coeficiente de correlación (R) que mide la intensidad de la relación lineal entre dos variables cuantitativas, en nuestro caso, la edad y el peso. Las características del coeficiente de correlación son las siguientes:

R tiene un valor comprendido entre -1 y 1

Para $R = \pm 1$ hay una relación lineal perfecta entre "x" e "y", es decir, todos los puntos se encontrarían en una línea recta.

Un valor positivo de R indica que, a medida que aumenta una variable, lo hace la otra, o a medida que disminuye una, también lo hace la otra (directamente proporcional), mientras que un coeficiente de correlación negativo indica que, a medida que disminuye una variable, aumenta la otra o viceversa (inversamente proporcional)

- $R = 0$ indica que no hay relación lineal entre las dos variables.

El valor R es independiente de cualquier unidad usada para medir las variables.

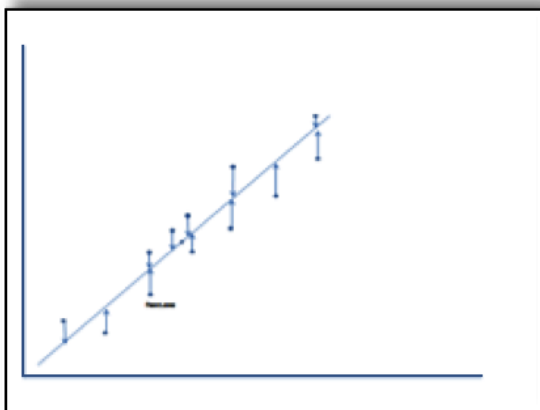
La correlación no implica necesariamente causalidad.

El cuadrado del coeficiente de correlación (R^2) se denomina coeficiente de determinación y va a representar el porcentaje de cambio de la variable "y" que es explicado por la variable "x".

TIPOS DE REGRESIÓN

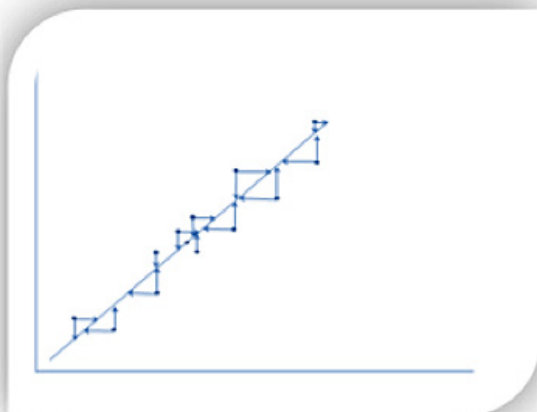
La ecuación se interpreta como el cambio en promedio de la variable "y" por el incremento en una unidad de la variable "x". La estimación de "a", es decir, la estimación de la ordenada en el origen, se interpreta como el valor promedio cuando la variable independiente "x" es igual a 0.

Para llevar a cabo una regresión lineal simple entre dos variables siempre se recomienda representar gráficamente a los pares de puntos en el denominado diagrama de dispersión y trazar una recta de regresión sobre este gráfico. Los modelos de regresión lineal más frecuentes son: el método de mínimos cuadrados, regresión de Deming y regresión de Passing Bablock.



Método de mínimos cuadrados

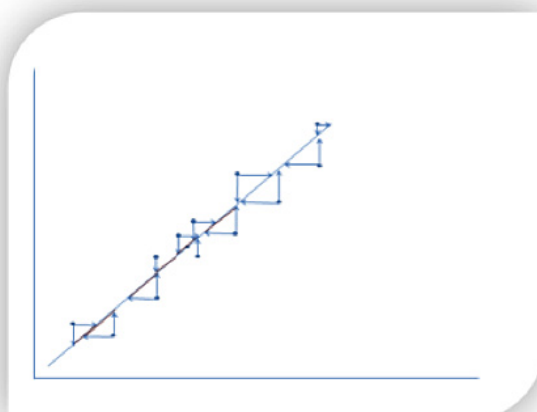
Este método trata de encontrar el mejor ajuste posible de tal forma que el cuadrado de la distancia de cada punto a la recta de regresión sea mínimo, es decir, que la distancia en el eje de ordenadas entre un punto y la recta sea mínima. Esta distancia entre la recta y el punto es lo que se conoce como residuales.



Regresión de Deming

Esta regresión se basa en un modelo que trata de encontrar la línea de mejor ajuste para un conjunto de datos en dos dimensiones. Se diferencia del modelo de mínimos cuadrados en que esta última se basa en los errores de los distintos puntos de la recta respecto al eje "x", mientras que el modelo de Deming trata de que los errores respecto al eje de las "x" y de las "y" sea mínimo.

El cálculo del modelo de Deming es más complejo que el de mínimos cuadrados. No obstante, este tipo de regresión se encuentra en la mayoría de los software de química clínica.



Modelo de Passing-Bablok

Es un modelo de regresión no paramétrica. Es robusto en presencia de uno o varios outliers. Ajusta los coeficientes a y b de forma no paramétrica calculando la media de las pendientes de la ecuación $y=a+bx$ de las líneas de regresión entre dos puntos. Este modelo ha sido criticado porque ignora las diferencias aleatorias entre los métodos.

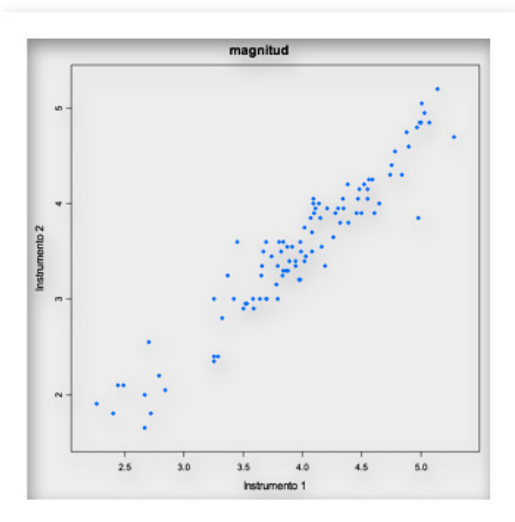
Regresión lineal en R

R en el entorno de Rcmdr sólo realiza la regresión lineal mediante el método de mínimos cuadrados. No obstante, existen paquetes para poder calcular los tres modelos que hemos descrito. En R los coeficientes del modelo a y b y el coeficiente de correlación se calculan mediante el ajuste de modelos. Os dejo un ejemplo de comparación de una magnitud en dos instrumentos diferentes:

```
Residuals:
  Min       1Q   Median       3Q      Max
-0.79775 -0.16436  0.00894  0.16845  0.64743

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.86990     0.14336  -6.068 2.44e-08 ***
Instrumeto1  1.10796     0.03549  31.218 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2367 on 98 degrees of freedom
Multiple R-squared:  0.9086, Adjusted R-squared:  0.9077
F-statistic: 974.6 on 1 and 98 DF, p-value: < 2.2e-16
```



La ordenada en el origen sería (Intercept) $-0,86990$, la pendiente $1,10796$ y el coeficiente de determinación $0,9077$ para obtener el coeficiente de correlación habría que hacer la raíz cuadrada.

Rcmdr nos da también una idea de la distribución de los residuales con los que se ha construido el modelo, así como un error estándar de los residuales.

```
> Confint(LinearModel.1, level=0.95)
              Estimate      2.5 %      97.5 %
(Intercept) -0.8699028 -1.154389 -0.5854166
Instrumetol  1.1079630  1.037533  1.1783934
```

Rcmdr puede calcularnos, además, los intervalos de confianza de la pendiente y de la ordenada en el origen.

La gráfica la haremos mediante un gráfico x-y.

Para hacer la regresión de Deming o Passing-Bablock hay que acudir a R estudio. Subiremos al foro el script correspondiente a este tipo de estudios.

La comisión de metrología de la SEQC recomienda cualquiera de los modelos alegando que no hay grandes diferencias entre ellos.

EDUCACIÓN CONTINUADA EN EL LABORATORIO CLÍNICO COMITÉ DE EDUCACIÓN

D. Balsells, B. Battikhi (*Residente*), R. Deulofeu, M. Gassó, N. Giménez, J.A. Lillo, A. Merino, A. Moreno, A. Peña (*Residente*), M. Rodríguez (*Presidente*), N. Rico, MC. Villà.

ISSN 1887-6463 – Febrero 2017 (recibido para publicación Noviembre 2016).