
El empleo de la correlación lineal simple en Epidemiología Veterinaria (The use of simple lineal correlation in veterinarian epidemiology)

Ramírez, W.

Departamento de Sanidad Animal, Facultad de Medicina Veterinaria, Universidad de Granma. Cuba.

Contacto: wramirezs@udg.co.cu

Resumen

El objetivo del presente trabajo, consiste en demostrar la utilidad del empleo de correlación lineal simple en el análisis de un problema de salud.

A tal propósito se desarrolla un problema de posible presentación en la Epidemiología Veterinaria; se explica en qué consiste el procedimiento, cómo se puede comprobar gráfica y matemáticamente, así como su utilidad en la práctica.

Se concluye que la correlación lineal simple es una herramienta, empleada adecuadamente, puede proporcionar una información confiable y útil en el análisis de algún problema de la salud animal.

Palabras Claves: Correlación, lineal, simple, utilidad

Abstract

The objective of the present work consists on demonstrating the utility of the employment of simple lineal correlation in the analysis of a problem of health. To such a purpose a problem of possible presentation is developed in the Veterinary Epidemiology; it is explained on what it consists the procedure, how it can be proven graphic and mathematically, as well as their utility in the practice.

It was concluded that the simple lineal correlation is an appropriate tool, employed, it can provide a reliable and useful information, in the analysis of some problem of the animal health.

Keyword: Correlation, lineal, simple, utility

Introducción

La Matemática Estadística ha contribuido mucho al desarrollo de las ciencias (física, química, biología, agronomía, etc.); en nuestros días, se ha hecho inaceptable un trabajo de investigación que no cuente con una base sobre análisis matemático.

La Medicina Veterinaria, como parte de las ciencias biológicas, obviamente requiere de ese fundamento y por tanto, los médicos veterinarios, deben hacer uso de esa herramienta.

En cualquier estudio que se desee hacer o se haga, como puede ser en Epidemiología Veterinaria, la matemática estadística desempeña un papel indispensable y quizás la evaluación de la correlación lineal simple, sea uno de los primeros pasos a dar.

El objetivo del presente trabajo, consiste en demostrar la utilidad del empleo de correlación lineal simple en el análisis de un problema de salud.

Desarrollo

La variación simultánea en una población, puede mostrar una determinada relación entre ellas y a esa relación (en nuestro caso de dos variables), es lo que se conoce como correlación. La correlación mide en esencia la fuerza de esa relación.

El coeficiente de correlación, es una cifra que se suele representar por una (**r**) minúscula, precedida por los signos positivo y negativo ($r = \pm$), con un valor determinado.

Para aclarar mejor el fundamento de la relación, como su cálculo está basado en la distribución normal, su valor será entre -1 y +1, es decir **$r = \pm 1$** .

Características de la correlación.

Si el incremento de una variable va acompañado de un aumento de la otra, se dice que la correlación es positiva (directa). Si por el contrario, al aumentar una de ellas, la otra disminuye, se establece una correlación negativa (inversa).

Puede ocurrir que no haya correlación entre ellas o que sea poca, entonces se dice que no hay correlación o que es pobre.

Se supone que ambas variables son aleatorias; se manejan las dos simétricamente.

Como ejemplos clásicos tenemos:

Correlación positiva.- Es la que existe entre los cambios de temperatura y la longitud de una barra, en la que por cada grado que aumenta la temperatura, la extensión de la barra se incrementa en una cantidad fija; por tanto la correlación es positiva.

Correlación negativa.- Es la que existe entre la altura y la presión atmosférica, porque a mayor altura, menor presión.

Deberá tenerse en cuenta que en los estudios biológicos, las correlaciones no son semejantes a los ejemplos señalados, debido ello a la gran variabilidad presente en los caracteres de los seres vivos.

Finalmente para concluir con la caracterización de correlación lineal simple, es lineal porque si determinamos los valores de sus variables en un sistema de coordenadas rectangulares, los puntos estarán sobre una línea recta o muy cercana a ella; siendo simples por estar constituida por dos variables. (Figuras 1 y 2).

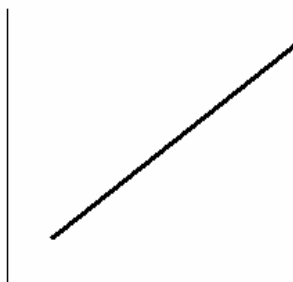


Figura 1: Correlación Positiva



Figura 2: Correlación Negativa

Las variables, identificadas como **x** e **y**, independiente y dependiente, respectivamente, también se denominan explicativa y dependiente.

La designación de sus nombres o identificación, posee un sentido de índole lógico; es decir, establecer ¿cuál variable depende de la otra?; como ejemplos, se pueden mencionar la longitud del tallo y el número de hojas de una planta, la relación entre la prevalencia de una enfermedad y su incidencia en una población animal, etc.

Determinación matemática del coeficiente de correlación.

Para medir matemáticamente y con mayor precisión el grado de relación entre dos variables, es necesario expresarlo numéricamente que es lo que recibe el nombre de **Coefficiente de Correlación**, ya mencionado con anterioridad.

El coeficiente de correlación se puede determinar mediante la fórmula siguiente:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}} \quad (1)$$

r = coeficiente de correlación lineal

n = número de pares de valores

Σxy = suma de cada par de valores

Σx = suma de todos los valores de la variable **x**

Σy = suma de todos los valores de la variable **y**

(Σx)² = cuadrado de la suma total de los valores de la variable **x**

(Σy)² = cuadrado de la suma total de los valores de la variable **y**

Σx² = suma de los cuadrados de los valores de la variable **x**

Σy^2 = suma de los cuadrados de los valores de la variable **y**

Ejemplo:

Determinar la relación existente entre las variables que a continuación se plantean.

En un territorio se han notificado: 12, 18, 24, 30, 36, 42, 48, **y** 5.27, 5.68, 6.25, 7.21, 8.02, 8.71, 8.42 enfermos y prevalencia de Anaplasmosis Bovina, respectivamente. Si el propósito fuera, supuestamente, evaluar la eficiencia de un tratamiento, se puede comprobar lo siguiente:

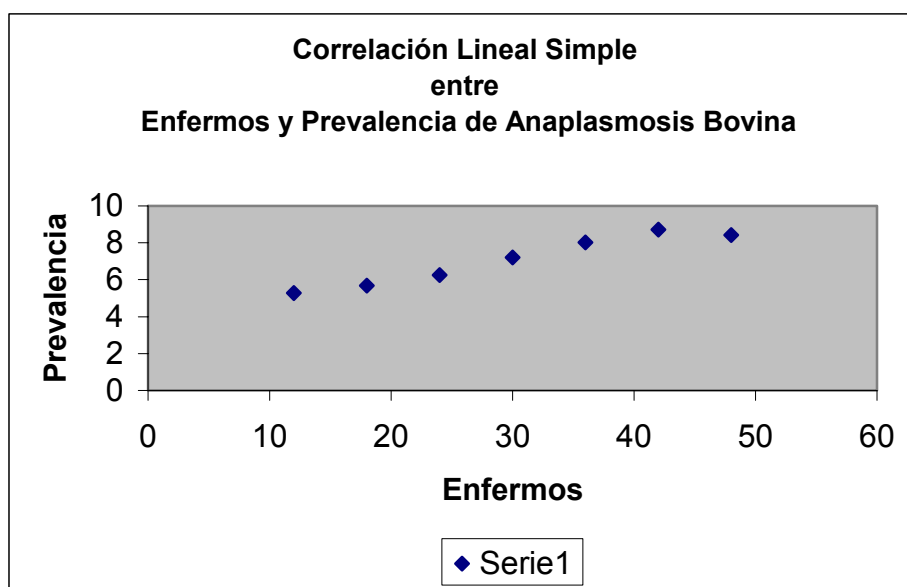


Fig. 1 Diagrama de dispersión

El diagrama de dispersión visualmente nos hace concebir que entre las variables pueda existir, con bastante certeza, una relación lineal.

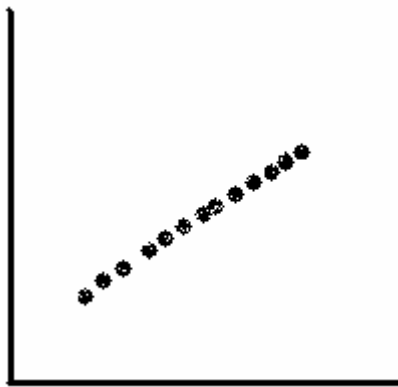
Para la determinación matemática se prepara una hoja de trabajo como sigue:

x	y	xy	X ²	Y ²
12	5.27	63.24	144	27.77
18	5.68	102.24	324	32.26
24	6.25	150.00	576	39.06
30	7.21	216.30	900	51.98
36	8.02	288.72	1296	64.32
42	8.71	365.82	1764	75.86
48	8.42	404.16	2304	70.90
Σ 210	49.56	1590.48	7308	362.15

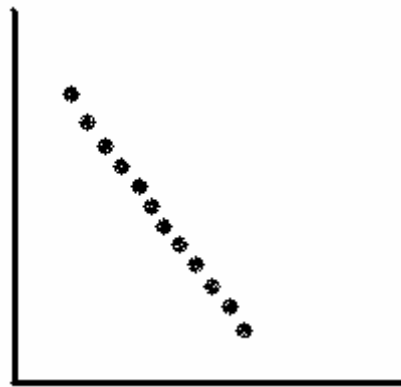
Cuadro 1 Hoja de trabajo.

Sustituyendo en (1):

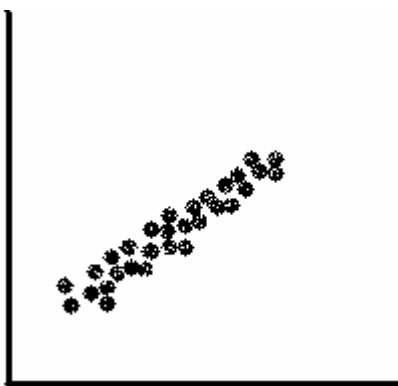
$$r = \frac{7 \cdot 1590.48 - 210 \cdot 49.56}{\sqrt{7 \cdot 7308 - (210)^2}} \frac{\sqrt{7 \cdot 362.15 - (49.56)^2}}{\sqrt{7 \cdot 7308 - (210)^2}}$$
$$= 0.973$$



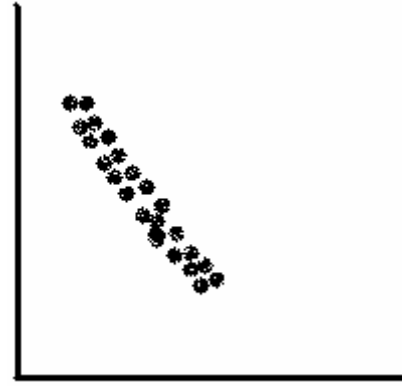
a) $r = 1$



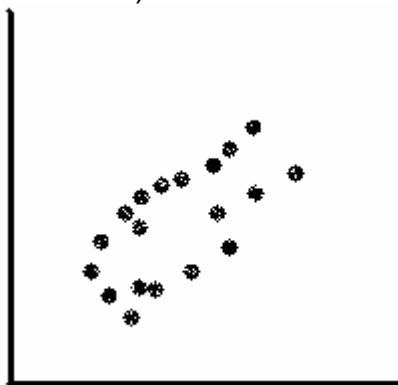
b) $r = -1$



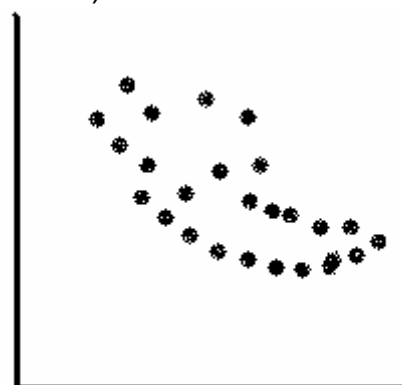
c) $r = \text{cerca de } 1$



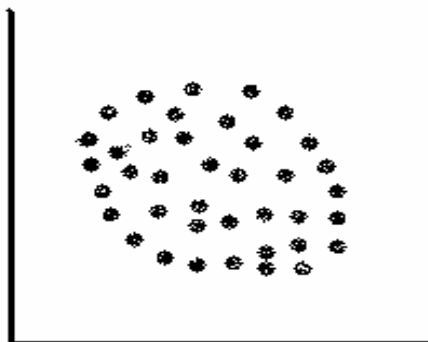
d) $r = \text{cerca de } -1$



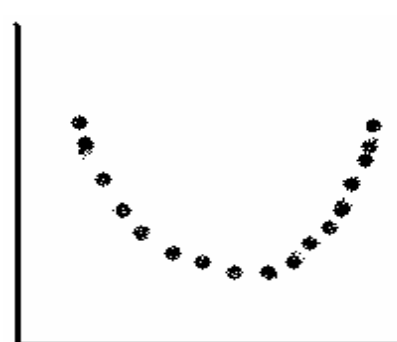
e) $r = \text{positivo pero cerca de } 0$



f) $r = \text{negativo pero cerca de } 0$



g) $r = 0$



h) $y = x^2$ pero $r = 0$

Figura 2 Representación de correlaciones.

La figura anterior (a, b, c, d, e, f, g, y h), demuestran diversas posibilidades.

r	Correlación	Figura
0	Ninguna correlación lineal, puede existir otro tipo. Las variables x e y son completamente independientes.	g
Cerca de 0	Correlación débil. Poca dependencia entre las variables x e y .	e
Cerca de 1	Correlación fuerte. Dependencia muy estrecha entre las dos variables x e y .	d
1	Correlación completa (perfecta).	a

Cuadro 2 Interpretación del coeficiente de correlación.

Según el valor determinado del coeficiente de correlación y en correspondencia con la tabla de interpretación, en el problema ejemplo, significa que existe una fuerte correlación entre las variables (enfermos y prevalencia de Anaplasmosis Bovina) en un territorio y momento dado. El resultado, pone en evidencia, que el tratamiento, no está siendo suficientemente eficaz, pues a medida que aumenta el número de enfermos, se incrementa el valor de la prevalencia. Deberá tenerse presente, que se está suponiendo que las demás posibles variables que intervienen en el asunto, están controladas.

Confiabilidad del coeficiente de correlación.

Esto se puede comprobar matemáticamente, lo que garantiza su utilización y se realiza mediante la fórmula siguiente:

$$F_c = r^2 (n - 2) / 1 - (r)^2 \quad (2) \quad \text{GL: } 1/n - 2$$

Sustituyendo en (2):

$$F_c = (0.973)^2 * 5 / 1 - (0.973)^2 = 89.314$$

Criterios de significación de la F.

$F_c > F_t$: correlación significativa

$F_c < F_t$: no hay correlación

Comparación de los valores de F_c con F_t .

$F_c > F_t$: correlación significativa

$F_c < F_t$: no hay correlación

$$F_{t .05} (1,5) = 6.61$$

$$F_{t .01} (1,5) = 16.3$$

Por tanto:

$$r = 0.973^{**} \text{ (altamente significativo)}$$

Propiedades más importantes del coeficiente de correlación.

1. Puede ser positivo o negativo; su signo dependerá del signo del numerador, que mide la covariación de las dos variables (Figuras a, b, c, d, e, f, g y h).
2. Tiene como límites -1 y $+1$; esto es, $-1 \leq r \leq 1$.
3. Es de naturaleza simétrica; es decir, que el coeficiente de correlación entre X e Y (r_{xy}) es igual al coeficiente de correlación entre Y y X (r_{yx}).
4. Es independiente del origen y de la escala; si definimos $X_i^* = aX_i + c$ y $Y_i^* = bY_i + d$, donde $a > 0$, $b > 0$, y c y d son constantes, entonces r entre X^* e Y^* es igual a r entre las variables originales X e Y .
5. Si X e Y son estadísticamente independientes, el coeficiente de correlación entre ellas es cero; pero si $r = 0$, no se puede inferir que las dos variables sean independientes; es decir, una correlación igual a cero, no implica necesariamente independencia (Figura h).
6. Es una medida de asociación lineal o dependencia lineal únicamente; por tanto, carece de sentido emplearlo para describir relaciones no lineales. En la figura h, $Y = X^2$ es una relación exacta pero r es cero.
7. Aunque mide la asociación lineal entre dos variables, no necesariamente implica una relación de causa y efecto.

Conclusión:

La correlación lineal simple, empleada adecuadamente, puede proporcionar una información confiable y útil en el análisis de algún problema de la salud animal.

Bibliografía consultada y recomendada:

- Bouza, C. N. Y Vivian Sistachas (2006). Estadística. Teoría básica y ejercicios. Editorial Félix Varela, La Habana.
- Carballo, Miriam y L. Prado. Bioestadística (1980). Editorial Pueblo y Educación, La Habana.
- Cué, J y col. (1987). Estadística. Facultad de Matemática Cibernética (T I - II) Editorial Pueblo y Educación, La Habana.
- Cramer, H. (1970). Métodos Matemáticos en la Estadística. Ediciones Aguilar, Madrid.

- Croxton, F. E. y D. J. Cowden (1972). Estadística General Aplicada. Fondo de Cultura Económica, México.
- Daniel, W. W. (1993). Bioestadística. Base para el análisis de las ciencias de la salud. UTEHA, México.
- Dixon, W. & F. Massey (1951). An Introduction to statistical analysis. McGraw – Hill Book Company, Inc., N.Y.
- Freund, J. E. (1960). Modern Elementary Statistics. Second Edition. Printice Hall, Inc. Englewood Cliffs, New Jersey.
- Guerra, Caridad y col. (1987). Estadística. Editorial Pueblo y Educación, La Habana.
- Kreyszic, E. (1974). Introducción a la Estadística Matemática. Principios y Métodos. Editorial Limusa, México.
- Lerch, G. (1977). La experimentación en las ciencias biológicas y agrícolas. Editorial Ciencia y Técnica, La Habana.
- Loma de la, J. (1981). Experimentación Agrícola. Edición Revolucionaria, La Habana (1ª edición).
- Mather, K. (1971). Análisis Estadístico en Biología. Paraninfo..
- Ostle, B. (1980). Estadística Aplicada. Editorial Ciencia y Técnica, La Habana.
- Sánchez, R y J. Torres (1989). Estadística Elemental. Editorial Pueblo y Educación, La Habana.
- Universidad de La Habana. (1986). Selección de Tablas Estadísticas. 1986.
- Snedecor, G W. & W. G. Cochran (1982). Métodos Estadísticos. Editorial Continental, México.
- Spiegel, M. (1977). Teoría y Problemas de Estadística. Editorial Revolucionaria, La Habana, (2ª Edición, 3ª reimpresión).
- Wilks, S. (1962). Mathematical Statistics. John Wiley and Sons, N. Y.

REDVET: 2009 Vol. 10, Nº 6

Recibido 27.04.09 - Aceptado 13.05.09 - Ref. 060906REDVET - Publicado: 01.06.09

Este artículo está disponible en <http://www.veterinaria.org/revistas/redvet/n060609.html>
concretamente en <http://www.veterinaria.org/revistas/redvet/n060609/060906.pdf>

REDVET® Revista Electrónica de Veterinaria está editada por Veterinaria Organización®.

Se autoriza la difusión y reenvío siempre que enlace con Veterinaria.org® <http://www.veterinaria.org>
y con REDVET® - <http://www.veterinaria.org/revistas/redvet> - <http://revista.veterinaria.org>