ESTUDIO ESTADISTICO DEL GRADO DE CORRELACION ENTRE LOS VALORES DEL HEMATOCRITO Y VOLUMEN CORPUSCULAR MEDIO EN GRUPOS DE NIÑOS DE MENOS DE 1 AÑO A 7 AÑOS DE EDAD. IV COMUNICACION

Por los Dres. TEODORO ZENTENO y OSCAR CORREA B.

Hospital L. Calvo Mackenna, Cátedra de Clínica Pediátrica del Prof. Aníbal Ariztia.

Con alguna frecuencia, en estudios biométricos, se plantea el problema de establecer el grado de relación entre dos variables. Ello, que en ciertos casos es de la más alta importancia, puede ser resuelto con exactitud mediante el uso de algunas constantes, cuya derivación y medios de interpretación, si bien uno de los capítulos más interesantes de esta clase de metodología científica, no tocaremos en esta ocasión por quedar fuera de la índole de este artículo. A título de ejemplo, sin embargo, de este tipo de estudio, supongamos que nuestro interés se concentre en demostrar el grado de asociación que existe, si es que éste existe, entre las condiciones climatéricas de los meses calurosos del verano y la producción de trastornos nutritivos agudos en lactantes sanos. Independientemente del tipo de alimentación que se haya instituido en niños de una edad determinada y de las condiciones de higiene y cuidado que los rodee es posible que, fuera de éste y otros factores que deliberadamente no consideramos, existan condiciones climáticas especiales que influyan en mayor o menor grado, en la aparición de dichos trastornos.

Si para el ejemplo anterior, como para muchísimos otros de la misma o diferente categoría, dispusiésemos de un método que nos permita encarar el problema en el sentido de medir el grado de correlación que hay entre las variables en juego es indudable que nuestra posición para el estudio e interpretación de los resultados sería más ventajosa, puesto que contaríamos con un procedimien-

to que, basado en una escala común, nos llevaría a calcular un valor que podría ser directamente comparable con otros valores similares, que representen ya sea los mismos caracteres de la serie que se analiza o caracteres diferentes de otros pares de series similares.

Antes de entrar a referirnos al tema que hoy comunicamos, creemos fundamental repetir, con Pearl, las expresiones siguientes: Toda vez que pueda apreciarse una cierta correlación entre dos series de fenómenos y, por esa misma razón se intente llegar al cálculo de este grado de correlación, debemos recordar de una sola vez y para siempre que el coeficiente que así obtengamos no mide otra cosa y no significa otra cosa que una asociación de orden estadístico entre las variables en estudio, independientemente de lo atractiva y seductora que pueda aparecer cualquiera hipótesis de relación causal entre dos hechos. En otros términos, la correlación mide únicamente una asociación de hechos estadísticos y en ningún caso penetra en el terreno de la causalidad.

En comunicaciones anteriores aparecidas en esta misma Revista^{1 2 3}, hemos dado a conocer consideraciones estadísticas sobre valores de glóbulos rojos, glóbulos blancos, hematocrito, hemoglobina y volumen corpuscular medio de niños de menos de 1 año a 7 años de edad. Nuestro propósito en este artículo es presentar los resultados del estudio estadístico del grado de correlación entre los valores del hematocrito y volumen corpuscular medio de estos mismos niños,

tomados por grupos anuales sucesivos. En la tabla 1 aparecen registrados algunos de los valores de este estudio y con el fin de obtener aquellos que necesitamos reemplazar en la fórmula del Coeficiente de Correlación que se va a emplear, de su Desviación Standard y de la Desviación Standard de los Errores Residuales, debemos proceder al cálculo que sigue:

$\bar{X}$ = media aritmética de los promedios de X	=	39.54
$\bar{Y}$ = media aritmética de los promedios de Y	=	88.44
ΣX = suma de los valores de X	=	276.80
ΣY = suma de los valores de Y	=	619.10
ΣX^2 = suma de los cuadrados de las X	=	10958.06
ΣY^2 = suma de los cuadrados de las Y	=	54797.87
ΣXY = suma de los productos de las X por las Y	=	24503.92
$\bar{X}^2$ = cuadrado de la media de X	=	1563.41
$\bar{Y}^2$ = cuadrado de la media de Y	=	7821.63
D. S. x = desviación standard de X	=	1.53
D. S. y = desviación standard de Y	=	2.57
D. S. x D. S. y = producto de las D. S. de X y de Y	=	3.9321

La fórmula del coeficiente de correlación que emplearemos es:

$$r = \frac{\Sigma (XY) - N (\bar{M}_x) (\bar{M}_y)}{\sqrt{(\Sigma X^2 - N \bar{M}_x^2) (\Sigma Y^2 - N \bar{M}_y^2)}}$$

en donde

Σ = suma

N = número de observaciones

$\bar{M}_x$ = media de X

$\bar{M}_y$ = media de Y.

De donde, substituyendo los valores en la ecuación, tendremos

$$r = \frac{24503.92 - (7) (39.54) (88.44)}{\sqrt{[10958.06 - (7) (39.54)^2] [54797.87 - (7) (88.44)^2]}}$$

cuyo resultado es

$$r = 0.99$$

y su desviación standard, alrededor de cero

$$D. S. r = \frac{1}{\sqrt{N-1}}, \quad D. S. r = \frac{1}{\sqrt{6}} = 0.4$$

Lo que permite concluir que el grado de correlación entre estas dos variables, hematocrito y volumen corpuscular medio, es directo y extraordinariamente estrecho. En otros términos, que hay un grado de asociación evidente entre estos dos valores y, puesto que el coeficiente de correlación r es más del doble del valor de su desviación standard, podemos afirmar que es muy poco pro-

bable que esta asociación se halle determinada por obra del azar.

Otro de los aspectos fundamentales de este problema y, sin duda, uno de los más interesantes, es el relativo al estudio de la predicción de valores basados en el grado de asociación de las variables. Así, ¿sería posible, si conocemos el valor de una de las variables, hematocrito, por ejemplo, determinar con cierto grado de aproximación el valor de la otra variable, volumen corpuscular medio? Y si esta predicción fuera posible, ¿dentro de qué margen de error podríamos estimar como cierta esta me-

didada? No necesitamos insistir sobre la importancia de un hallazgo de esta índole y lo sorprendente que a veces resulta obtener por medio del cálculo matemático e independientemente de los procedimientos de laboratorio, un valor que así lo establezca.

Con el objeto, en nuestro caso, de resolver este aspecto del estudio, debemos acudir al empleo de la Ecuación de Regresión, la misma que buscara Galton en sus estudios de genética y que desarrollaran de una manera admirable, años más tarde, Pearson y otros biómetras. Ya que, como antes dijimos, la

En esta ecuación, a y b , son las constantes o parámetros cuyos valores debemos determinar. Si reemplazamos en esta fórmula los valores encontrados en el cálculo anterior, tendremos

$$\begin{aligned} 7a + 276.8b &= 619.1 \\ 276.8a + 10958.06b &= 24503.92 \end{aligned}$$

Tabla 1

Edad en años	Nº de obs. (2)	X (3)	Y (4)	X ² (5)	Y ² (6)	XY (7)
—1	482	42.6	84.2	1.814.76	8.873.64	4.012.92
2	70	38.4	86.3	1.474.56	7.447.69	3.313.92
3	63	38.6	86.8	1.489.96	7.534.24	3.350.48
4	71	38.9	87.1	1.513.21	7.586.41	3.388.19
5	70	38.9	88.1	1.513.21	7.761.61	3.427.09
6	70	40.0	88.8	1.600.00	7.885.44	3.552.00
7	67	39.4	87.8	1.552.36	7.708.84	3.459.32
	893	276.8	619.1	10.958.06	54.797.87	24.503.92

En la columna 1 se halla consignada la edad.

La columna 2 contiene el número de observaciones.

X representa los valores de la media aritmética de las observaciones de hematocritos para cada grupo de edad.

Y representa los valores de la media aritmética de las observaciones de volúmenes corpusculares medios para cada grupo de edad.

X² los valores al cuadrado de las medias de hematocritos para cada grupo de edad.

Y² los valores al cuadrado de las medias de volúmenes corpusculares medios para cada grupo de edad.

XY el producto de las medias aritméticas de los valores del hematocrito y del volumen corpuscular medio para cada grupo de edad.

índole de este artículo no nos permite entrar en el análisis de los fundamentos de esta ecuación, debemos decir que parte importante de esta metodología comprende el uso de las llamadas Ecuaciones Normales, las que damos a continuación con el fin de que pueda conocerse el procedimiento que hemos seguido en este caso.

$$Na + b \sum X = \sum Y$$

$$a \sum X + b \sum X^2 = \sum XY$$

Si calculamos a (uno de los parámetros de la ecuación), tendremos

$$a = \frac{(619.1 \times 10958.06) - (24503.92 \times 276.8)}{(7 \times 10958.06) - (276.8)^2}$$

lo que, una vez resuelto, nos da

$$a = 16.4423$$

y, por substitución en la primera de las

ecuaciones, nos permite obtener el valor de b .

Con estos valores en la mano podemos construir la Ecuación de Regresión.

$$y = a + bx$$

y, reemplazando, tendremos

$$y = 16.44.233 + 1.8208 x$$

Prácticamente

$$y = 16.4 + 1.8 x$$

En esta ecuación el parámetro "a" nos revela el valor que tendría Y (volumen corpuscular medio) cuando X sea = 0 y el cambio que se determina en Y por cada unidad de cambio que se opera en X. Supongamos, por ejemplo, que nos interesemos por conocer cuál sería el valor del volumen corpuscular medio que queda determinado por un valor en el hematocrito de 38.9. La ecuación en este caso nos dice que el valor de Y (volumen corpuscular medio) sería igual al que se obtiene mediante el uso de la ecuación de regresión que sigue

$$Y = 16.4 + 1.8 \cdot 38.9$$

(puesto que a X le hemos asignado de antemano el valor de 38.9).

El resultado de la ecuación nos da, una vez hecho el cálculo, un valor de 86.4, es decir, en otros términos, el valor del volumen corpuscular medio obtenido mediante el cálculo para un valor de 38.9 en el hematocrito sería igual a 86.4.

Los valores obtenidos en la forma antes señalada (bastan en general dos para fijar la posición de una línea), nos permite construir un diagrama, como el que presentamos (figura 1). En este diagrama los valores del hematocrito se encuentran dibujados en la abscisa y los valores del volumen corpuscular medio en la ordenada correspondiente. Los promedios de las series de

valores X e Y, por pares de valores, aparecen también indicados y corresponden a los puntos dispersos del diagrama. La proyección de estos valores parece dibujar y hasta disponerse en el sentido de una línea recta. Mediante la ecuación de regresión hemos calculado la línea que mejor representa esta tendencia una vez suavizadas las diferencias atribuibles al azar y ella se halla representada por la línea continua.

Sin necesidad del empleo de la ecuación de regresión para calcular un valor cualquiera de volumen corpuscular medio que corresponda a un valor predeterminado de hematocrito, el diagrama así construido se presta para hacerlo. Basta para ello fijar el valor del hematocrito en cuestión en el punto correspondiente de la abscisa y levantar una perpendicular a escuadra, a partir de este punto, hasta alcanzar aquel otro en que la línea así trazada corte a la recta (línea de regresión de Y de X). Ahora, si a partir de la ordenada correspondiente al volumen corpuscular medio trazamos otra perpendicular que alcance a la recta en el punto anteriormente encontrado, el pie de dicha perpendicular nos permitirá hallar, de una manera inmediata, el valor del volumen corpuscular medio correspondiente al valor del hematocrito que de antemano conocemos.

Como puede apreciarse, el empleo del diagrama simplifica considerablemente nuestro problema y evita, hasta cierto punto, recordar los valores de los parámetros "a" y "b" que nos sirvieron para calcular nuestra línea de regresión. La ventaja de esta última se halla en que nos permite calcular cualquier valor, dentro de los límites para los cuales ella ha sido construida y, además, en que posibilita al estadístico rehacer exactamente el diagrama, puesto que invariablemente la ecuación le suministrará los datos necesarios para dibujar esta línea, que, en definitiva, es la única que puede ser así trazada.

Para contestar, ahora, la segunda de las preguntas que nos hemos hecho,

FIG. 1.

¿dentro de qué margen de error pueden establecerse estos hechos como ciertos?, necesitamos conocer la desviación standard de los valores de la línea y ésta se obtiene mediante la fórmula

$$S_{yx} = D. S. y \sqrt{1 - r^2}$$

o mediante el procedimiento que hemos descrito en ocasión anterior, para determinar la desviación standard corriente, aplicada a los residuos o diferencias

Tabla 2

Edad en años	X (2)	Y (3)	Y' (4)	Y—Y' (5)	(Y—Y') ² (6)
—1	42.6	84.2	84.0	+0.2	0.04
2	38.4	86.3	86.4	—0.1	0.01
3	38.6	86.8	86.7	+0.1	0.01
4	38.9	87.1	87.3	—0.2	0.04
5	38.9	88.1	87.3	+0.8	0.64
6	40.0	88.8	89.3	—0.5	0.25
7	39.4	87.8	88.1	—0.3	0.09
				0.0	1.08

La columna 1 corresponde a la edad.

X media aritmética de los valores del hematocrito para cada grupo de edad.

Y media aritmética de los valores del volumen corpuscular medio para cada grupo de edad.

Y' los valores del volumen corpuscular medio obtenidos por el empleo de las ecuaciones normales para establecer la línea de regresión.

Y—Y' las diferencias entre los valores reales obtenidos de las observaciones objeto de este estudio y los valores que suponemos en un Universo normal.

(Y—Y')² el cuadrado de estas diferencias cuya raíz cuadrada del promedio de dichos valores nos dará el error o límite de nuestra estimación.

de los valores de Y (volúmenes corpusculares medios encontrados en la observación directa de los casos) e Y prima (Y'), que son los valores obtenidos por medio de la aplicación de la línea de regresión. La tabla 2 ilustra este último procedimiento.

De donde

$$S_y = \sqrt{\frac{1.08}{7}} = 0.4$$

El procedimiento

$$S_y = D. S. y \sqrt{1 - r^2}$$

da exactamente el mismo resultado (en nuestro cálculo hay una diferencia en la segunda decimal, explicable por el reducido número de cifras decimales con que hemos operado). En ambos casos los resultados, redondeados, dan una desviación standard de 0.4, la interpretación de la cual aquí (D. S. de los residuos de la línea) es en todo igual a la que hemos dado a conocer en publicaciones anteriores.

Resumen y conclusiones.

Se presenta un estudio estadístico sobre el grado de correlación entre los valores del hematocrito y volumen corpuscular medio en niños de menos de 1 año a 7 años de edad, tomados por grupos anuales sucesivos. Se demuestra que existe una asociación muy marcada ($r = 0.99$) entre los valores del hematocrito y volumen corpuscular medio, la que es positiva y directa, de donde el Coeficiente de Correlación puede ser escrito $+ 0.99$. La desviación standard de este coeficiente es 0.4.

Con los valores arriba mencionados se determina la Ecuación de Regresión siguiente: $Y = 16.4 + 1.8 X$, en donde Y representa los valores correspondientes al volumen corpuscular medio y X a los valores del hematocrito. La ecuación de regresión permite predecir, dentro de ciertos límites, los valores que debemos atribuir a la variable Y (volumen corpuscular medio) cada vez que conozcamos los valores de la variable X (hematocrito). El límite de error puede deducirse de la desviación standard de los residuos de la línea, que es igual a 0.4.

La línea de regresión determina los valores de Y en función de X.

Summary and conclusions.

A measurement of correlation in presented between values of volume of packed erythrocytes and mean corpuscular volume of infants and children from under 1 year to seven years of age. It was found a high association ($r = 0.99$) between values of volume of packed erythrocytes and mean corpuscular, which is positive and direct, so that the Coefficient of Correlation can be writted $+ 0.99$. The standard deviation of this coefficient around a value of "0" is 0.4.

With the values above mentioned the following Regression Equation is determined: $Y = 16.4 + 1.8 X$, where Y represents the values of mean corpuscular volume and X those of volume of packed erythrocytes. The regression equation permits to predict, within certain limits, the values which one must assign to the variable Y (mean corpuscular volume) when is known the values of the variable X (volume of packed erythrocytes). The limit of error may be derived from the standard deviation of the residuals of the line (standard

error of estimate), which is equal to 0.4.

The regression line determines the values of Y in function of X.

Bibliografía.

- 1.—ZENTENO, T.; CORREA, O. y GRUNEWALD, O. — Recuento de glóbulos rojos y blancos y determinación de hematocrito, hemoglobina y volumen corpuscular medio en 893 lactantes y niños desde la primera semana de la vida hasta los siete años. I Comunicación. Rev. Ch. de Pediatría 21: 99-110, marzo de 1950.
- 2.—ZENTENO, T. y CORREA, O. — Idem. II Comunicación: Coeficiente de variación. Rev. Ch. de Pediatría 21: 111-113, marzo de 1950.
- 3.—ZENTENO, T. y CORREA, O. — Idem. III Comunicación: Estudio de la significación estadística de las diferencias entre promedios. Rev. Ch. de Pediatría 21: 163-169, abril de 1950.
- 4.—PETERS, C. y VAN VOORHIS, W. — Statistical procedures and their mathematical bases. Mc Graw-Hill, New York, 1940.
- 5.—SNEDECOR, G. — Statistical Methods. Iowa State College Press, Ames, Iowa, 1948.
- 6.—TRELOAR, A. — Elements of statistical reasoning. John Wiley and Sons, New York, 1939.
- 7.—ZENTENO, T. y CORREA, O. — Estudio del peso del lactante. I Comunicación. Rev. Ch. de Pediatría 19: 270-284, marzo de 1948.

Corrección.

En el número anterior, III Comunicación, aparecen los siguientes errores:

Pág. 166. Donde dice:

Edad. N° obs. Promedio obs., etc., debe decir:
(en millones)

1) Edad. 2) N° obs. 3) Promedio obs. 4) D. S. obs. 5) Rango obs. 6) D. S. Promedio.
7) Rango promedio.

TABLA 3. Debe decir:

1) Edad. 2) N° obs. 3) Promedio obs. 4) D. S. obs. 5) Rango obs. 6) D. S. Promedio.
7) Rango promedio.

TABLA 4. Debe decir:

1) Edad. 2) N° obs. 3) Promedio obs. 4) D. S. obs. 5) Rango obs. 6) D. S. Promedio.
7) Rango promedio.

TABLA 5. Debe decir:

1) Edad. 2) N° obs. 3) Promedio obs. 4) D. S. obs. 5) Rango obs. 6) D. S. Promedio.
7) Rango promedio.