ANALES
DE LA
FACULTAD DE MEDICINA

TOMO XXVIII 3 LIMA, TERCER TRIMESTRE DE 1945

METODOS ESTADISTICOS

Dr. ALBERTO HURTADO
Departamento de Fisiopatología
Facultad de Medicina

PROLOGO

El análisis, interpretación y publicación de datos cuantitativos, asi como su estudio comparativo con otros datos similares, requiere en Medicina, al igual que en otras ciencias, el empleo de métodos estadísticos. Cierta familiarización con estos métodos, y los términos usados, se ha hecho también conveniente para quien consulta la literatura médica, por la frecuencia de su inclusión en artículos de diversa índole. Esta labor presenta, a menudo, dificultades para el médico y estudiante, por la lejanía de los días de matemática escolar y por la falta de textos adecuados de consulta.

Hemos creído útil, por estas razones, reunir las instrucciones necesarias para el empleo de algunos de los métodos más usados en Medicina, desarrollándolos en una forma elemental, eliminando en lo posible toda discusión matemática e ilustrando su aplicación con ejemplos concretos. Además, se ha procurado que cada capítulo se relacione, en forma independiente con determinado aspecto estadístico, de tal manera que el lector pueda satisfacer una consulta sin necesidad de referirse a los otros capítulos. Estas consideraciones indican, claramente, que las páginas que siguen no pretenden corresponder a un manual de estadística.

En la preparación de este folleto hemos utilizado apuntes acumulados en este Departamento durante algunos años, notas referentes a charlas que sobre estadística médica hemos ofrecido y las siguientes obras a las que referimos al lector interesado por una más amplia discusión y desarrollo:

Introduction to Medical Biometry and Statistics.
R. Pearl. W. B. Saunders Co., Philadelphia, 1940.

An Outline of Statistical Methods.
H. Arkin and R. Colton. Barnes & Noble Co., New York, 1938.

Métodos Estadísticos.
F. C. Mills. Editorial M. Aguilar, Madrid, 1940.

Análisis Estadístico.
Sixto E. Trucco. Editorial El Ateneo, Buenos Aires, 1944.

Principles of Medical Statistics.
A. B. Hill. The Lancet Limited, London, 1939.

Nociones de Bio-Estadística.
Franz Schruefer. Caja Nacional de Seguro Social, Lima, 1941.

ALFABETO GRIEGO

Α	α	ALPHA	Ν	ν	NU
Β	β	BETA	Ξ	ξ	XI
Γ	γ	GAMMA	Ο	ο	OMICRON
Δ	δ	DELTA	Π	π	PI
Ε	ε	EPSILON	Ρ	ρ	RHO
Ζ	ζ	ZETA	Σ	σ	SIGMA
Η	η	ETA	Τ	τ	TAU
Θ	θ	THETA	Υ	υ	UPSILON
Ι	ι	IOTA	Φ	φ	PHI
Κ	κ	KAPPA	Χ	χ	CHI
Λ	λ	LAMBDA	Ψ	ψ	PSI
Μ	μ	MU	Ω	ω	OMEGA

CONTENIDO

Relación lineal (que es representada gráficamente por una línea recta): significado y cálculo del coeficiente de correlación.
Derivación de la ecuación de regresión para predecir una variable cuando el valor de la otra es conocido.
Relación no lineal (que es representada gráficamente por una línea curva): significado y cálculo de la razón de correlación.
Procedimiento para definir si una relación es lineal o no lineal: significado estadístico de la diferencia entre el coeficiente de correlación y la razón de correlación.
Correlación parcial: relación entre dos series de datos cuando una tercera es mantenida constante. Significado y cálculo del coeficiente de correlación parcial.
Presentación de los resultados obtenidos en el estudio estadístico de la relación entre dos o más series de datos.

Página: 153

CAPITULO IV *Significado estadístico de las diferencias. Cálculo de probabilidades.*

Procedimientos para averiguar si tienen significado estadístico las diferencias halladas entre dos o más:

Medias o promedios,
Desviaciones standards,
Porcentajes, y
Coeficientes de correlación.

Cálculo de probabilidades: probabilidad de encontrar por casualidad o azar una diferencia igual a la encontrada entre dos valores.

Página: 180

CAPITULO V *Representación gráfica.*

Características generales de los diagramas.
Construción de diagramas que representan la frecuencia, variabilidad y distribución de los datos analizados:

1—Diagrama de barras horizontales;

2—Histograma;
3—Polígono de frecuencia;
4—Diagrama de frecuencias acumuladas; y
5—Diagrama en coordenadas angulares.

Construcción de diagramas que representan tendencia y relación:

1—Diagrama en coordenadas con escala aritmética;
2—Diagrama en coordenadas con escala logarítmica o semi-logarítmica;
3—Diagrama de dispersión; y
4—Diagrama polar.

Nomograma.

Diagrama que representa comparativamente la variabilidad hallada en el análisis estadístico de dos o más series de datos: representación gráfica del coeficiente de variación.
Cálculo de la línea de regresión: representación gráfica de la relación expresada por el coeficiente de correlación.

Página: 189

CAPITULO VI *Construcción matemática de una línea recta o curva.*

Procedimiento para construir una línea recta o curva (parabólica o logarítmica) que represente la relación entre dos series datos o variables, cuando esta relación está dada en un diagrama por varios puntos aislados ('curve fitting').

Página: 249

CAPITULO VII *Bio-Estadística. Coeficientes o tasas; indices.*

Procedimientos para calcular los siguientes coeficientes e índices:

A—Mortalidad.

Coeficiente general de mortalidad:

APENDICES

INTRODUCCION

Aunque este folleto no corresponde a un manual de estadística, sino simplemente a una descripción elemental de algunos de los métodos más usados en Medicina, es quizás pertinente señalar, en forma breve, las características y limitaciones más importantes relacionadas con la aplicación de estos métodos.

La mentalidad médica es, por lo general, poco afecta y propicia a la consideración estadística. Es evidente, sin embargo, que conclusiones experimentales y clínicas reposan, con frecuencia, sobre datos cuantitativos cuya interpretación correcta sólo puede ser hecha por medio del análisis estadístico. Dicho análisis, para ser llevado a cabo satisfactoriamente, requiere, aparte del conocimiento de los simples procedimientos matemáticos relacionados, la comprensión de las limitaciones que tienen los resultados hallados. La ignorancia de esta segunda consideración es quizás el factor principal en el mal uso de la estadística, y de la aparente justificación de la frecuente crítica a su empleo.

Señalaremos algunas de las consideraciones principales que deben regir la aplicación de métodos estadísticos:

1—Métodos estadísticos sólo pueden ser aplicados al análisis de datos de carácter *cuantitativo*.

2—El significado estadístico no es siempre sinónimo de significado médico o biológico. Por ejemplo, un grupo de sujetos tiene un valor medio de 60 pulsaciones por minuto y en otro grupo dicho valor corresponde a 62 pulsaciones. La diferencia entre ambas medias puede tener significado estadístico (dependiente del número de observaciones y de la variabilidad observada), pero poco o ninguno desde un punto de vista fisiológico o clínico.

Con frecuencia, el corto número de observaciones analizadas no permite generalizar los resultados estadísticos hallados, por precisos que éstos sean desde un punto de vista matemático.

3—Los métodos estadísticos analizan, a menudo, cifras cuya magnitud está sujeta a múltiples influencias, las que son necesarias tomar en cuenta al interpretar el resultado, o resultados obtenidos, mediante el empleo de dichos métodos. Ajustar, con un criterio inflexible, una cifra o valor, determinado estadísticamente, a la interpretación de un fenómeno biológico o clínico puede conducir a errores lamentables.

4—El análisis estadístico presupone, como condición esencial, que los datos que se analizan representen observaciones precisas. No hay método estadístico que compense o remedie investigaciones carentes de técnica o una recopilación de datos que corresponden a observaciones inexactas.

El significado, pues, del análisis estadístico reposa, fundamentalmente, sobre la veracidad de los datos analizados.

CAPITULO I

ANALISIS Y PRESENTACION DE DATOS CUANTITATIVOS

El análisis, interpretación y presentación de datos cuantitativos requiere su ordenación y el cálculo de algunas constantes que expresen, en forma concisa, los valores centrales de frecuencia y el grado de variación o dispersión. Las constantes más frecuentemente calculadas con este fin son: *media aritmética* o *promedio, desviación standard* y *coeficiente de variación*. En algunos casos puede ser útil el cálculo de la *mediana* y el *modo,* pero por lo general estas dos últimas constantes no son incluídas en la presentación de los datos analizados. A las constantes mencionadas se les calcula el *error standard.*

El significado de estas constantes es como sigue:

Media aritmética o promedio (Símbolo: M).—Es la constante más común y la más fácilmente entendida. Representa, mecánicamente, el centro de gravedad de los datos analizados. Cuando no se emplean métodos estadísticos adecuados en el análisis de datos cuantitativos, la media generalmente constituye la única constante mencionada. Su valor aislado tiene, sin embargo, un significado muy limitado pues no indica la variabilidad o dispersión de los datos; además, el valor de esta constante puede no ser enteramente típico por la influencia que tienen en su cálculo los valores extremos.

Desviación standard (Símbolo: σ).—También denominada "desviación cuadrática media" y "desviación típica". Esta constante mide, en términos absolutos de la unidad de medida a que se refiere los datos analizados, el grado de variación o dispersión de estos datos. Si la distribución de éstos es simétrica (distribuídos proporcionalmente a ambos lados de la media), la resta y la adición de la desviación standard a la media incluye, entre las cifras mínima y máxima obtenidas, el 68.3% de los datos analizados; si las mismas operaciones se hacen con la desviación standard $\times$ 2 el 95.5% de los datos están incluídos, y este porcentaje alcanza al 99.7% si las operaciones se realizan con la desviación standard $\times$ 3.

Ejemplo: En el análisis estadístico de mediciones de estatura hechas en 90 sujetos se han obtenido las siguientes constantes:

Media = 165 centímetros
Desviación standard = 3 centímetros

(a)—68.3% de las 90 estaturas están comprendidas entre 162 — 168 centímetros:

165 — 3 = 162
165 + 3 = 168

(b)—95.5% de las 90 estaturas están comprendidas entre 159 — 171 centímetros:

165 — (3 X 2) = 159
165 + (3 X 2) = 171

(c)—99.7% de las 90 estaturas están comprendidas entre 156 — 174 centímetros:

165 — (3 X 3) = 156
165 + (3 X 3) = 174

Con fines comparativos es a veces necesario fijar los límites de variación encontrados en el análisis de una serie de datos. Para esto se utilizan los cálculos acabados de mencionar, y es costumbre fijar dichos límites restando y añadiendo a la media la desviación standard $\times$ 2, ya que estas operaciones proporcionan valores que incluyen aproximadamente el 96% de los datos analizados.

Ejemplo: Con el objeto de apreciar el aumento y disminución de la hemoglobina (gramos por 100 cc. de sangre) en casos de policitemia y anemia, respectivamente, se ha determinado, previamente, esta substancia en 200 sujetos sanos hallando los siguientes resultados:

Media = 16.00 gramos
Desviación standard = 0.70 gramos

Los límites de variación son:

16.00 — (0.70 X 2) = 14.60 gramos
16.00 + (0.70 X 2) = 17.40 gramos

Luego, si se acepta que 200 determinaciones de hemoglobina en sujetos sanos representan fielmente los valores correspondientes a toda la población sana, tenemos, que todo valor por debajo de 14.60 gramos, y por en encima de 17.40 gramos, pueden considerarse como anormales.

Coeficiente de variación (C. V. %).—Expresa, en porcentaje, la variación o dispersión de los datos analizados. Se calcula:

$$\frac{\text{Desviación standard}}{\text{Media}} \times 100$$

Es una medida útil de variación desde un punto de vista comparativo, ya que no es posible comparar directamente unidades correspondientes a medidas diferentes.

Ejemplo: En los 90 sujetos, citados en un ejemplo anterior, y en quienes se ha medido la estatura (en centímetros) se ha determinado también el peso (en kilos). Los resultados obtenidos son los siguientes:

Estatura— Media = 165 cms.
Desviación standard = 3 cms.

Peso— Media = 60 kilos
Desviación standard = 4 kilos

Puede uno preguntarse: ¿Qué varía más en esta serie de sujetos: la estatura o el peso? No siendo posible comparar directamente "centímetros" con "kilos" se emplean los coeficientes de variación respectivos:

$$\text{Estatura: Coef. de var.} = \frac{3}{165} \times 100 = 1.8\%$$

$$\text{Peso: Coef. de var.} = \frac{4}{60} \times 100 = 6.7\%$$

La respuesta es: el peso varía más puesto que su coeficiente de variación es más elevado que el correspondiente a la estatura.

Las constantes estadísticas que hemos descrito son las más usadas das y las únicas que generalmente se mencionan en publicaciones. Puede, sin embargo, interesar el cálculo de las siguientes:

Mediana (Símbolo: Md).—Es la constante por encima de la cual se encuentra el 50% de los datos analizados, hallándose el otro 50% por debajo. El cálculo de la mediana no es influenciado por los valores extremos, y, por consiguiente, puede, a veces, representarse más fielmente que la media el tipo o centro de variación de los datos. Su desventaja consiste en que no es una constante utilizada con frecuencia en publicaciones y su significado es, por lo tanto, menos comprendido.

Modo (Símbolo: Mo).—Es el valor más frecuente y típico de los datos analizados. En su cálculo no influyen los valores extremos de los datos y puede tener, por consiguiente, un significado más descriptivo que la media y la mediana. Al igual que esta última constante, el modo tiene la desventaja de no ser utilizado con frecuencia. Además, su valor es limitado cuando el número de datos que se analiza es pequeño, y en estas circunstancias puede aún no existir si ninguno de los datos es repetido. No es tampoco posible hacer un cálculo matemáticamente exacto de esta constante.

Error standard (E. S.).—El significado estadístico de las constantes descritas se juzga en relación al error standard calculado para cada una de ellas, y que se expresa a continuación precedido por el signo ±. Por ejemplo, el valor medio del ácido láctico en la sangre, determinado en una serie de sujetos, es 12.5 milígramos por 100 cc., y el error standard de esta media es 0.15. Esto se expresa: 12.5 ± 0.15 milígramos por 100 cc.

El error standard es una medida de precisión y también de variabilidad, aunque no es generalmente usado en este último sentido. El error standard es tanto más elevado cuanto mayor sea la variabilidad y menor el número de observaciones.

Se ha hecho costumbre aceptar que una constante, o diferencia, menor a dos veces su error standard no tiene significado estadístico; en cambio éste es evidente si es dos o más veces mayor *.

Ejemplo: En dos grupos de sujetos se ha determinado la bilirrubina en el plasma sanguíneo.

En el primer grupo la media ± error standard obtenida es: 0.42 ± 0.32 milígramos por 100 cc. En este caso la media es menor a dos veces su error standard (0.32 X 2 = 0.64); luego no tiene significado estadístico.

En el segundo grupo la media ± error standard obtenida es: 0.53 ± 0.12 milígramos por 100 cc. En este caso la media es mayor a dos veces su error standard (0.12 X 2 = 0.24); en consecuencia sí tiene significado estadístico.

Los métodos que se emplean para calcular las constantes descritas varían según el número de datos por analizar es menor o mayor a 20. Ambos procedimientos serán descritos separadamente.

Cálculo de las constantes de frecuencia y variación cuando el número de datos por analizar es menor 20 *.—Los procedimientos que

---

* *Algunos autores usan, con la mismos fines y de idéntica manera, el llamado ERROR PROBABLE (E. P.), en lugar del error standard.*
El error probable equivale al error standard X 0.6745.
Para que una constante, o diferencia, tenga significado estadístico debe ser tres o más veces mayor que su error probable.

se emplean para determinar la media, desviación standard y el coeficiente de variación (la mediana y el modo son omitidos dado el corto número de observaciones) serán desarrollados utilizando un ejemplo:

Ejemplo: En 12 sujetos se ha determinado la hemoglobina (en gramos por 100 cc. de sangre). Los resultados obtenidos están dados en el Cuadro 1.

CUADRO 1

DETERMINACIONES DE HEMOBLOBINA
(gramos por 100 cc. de sangre)
EN 12 SUJETOS

13.20	12.12	13.10	13.50
13.88	14.80	14.60	11.90
14.00	12.10	13.40	13.50

(a)—Con estos datos se construye el Cuadro 2:

CUADRO 2

1	2
x	x^2
13.20	174.2400
13.88	192.6544
14.00	196.0000
12.12	146.8944
14.80	219.0400
12.10	146.4100
13.10	171.6100
14.60	213.1600
13.40	179.5600
13.50	182.2500
11.90	141.6100
13.50	182.2500
160.10	2145.5788

* *El número de 20 datos no es un límite estricto para usar los métodos descritos en esta sección; de igual manera puede procederse con 25, 30, o aún más datos, pero en estos casos es preferible aplicar los métodos indicados en la siguiente sección (página 138).*

Con respecto al número mínimo de datos que pueden ser sometidos a un análisis estadístico no hay una indicación precisa, pero en términos generales puede señalarse como 6 o 8. El criterio final, a este respecto, debe ser dado por la relación que se encuentre entre las diversas constantes y sus respectivos errores standard, que pueden ser elevados dado el corto número de datos analizados. Si el error standard X 2 excede en valor a la constante que lo precede hay que concluír que esta última no tiene significado estadístico.

Columna 1 (x)— Cifras de hemoglobina tal como se han obtenido.
Se suman las cifras de esta columna.

Columna 2 (x^2)— Corresponde al cuadrado de cada una de las cifras de la Columna 1 (x):

$$(13.20)^2 = 174.2400$$
$$(13.88)^2 = 192.6544$$

y así sucesivamente.

Se suman las cifras de esta columna.

(b)—Se hacen los siguientes cálculos:

$$V1 = \frac{\text{Suma: } x}{\text{N}^\circ \text{ datos}} = \frac{160{,}10}{12} = 13.3417$$

$$V2 = \frac{\text{Suma } x^2}{\text{N}^\circ \text{ datos}} = \frac{2145.6788}{12} = 178.8064$$

$$\pi_2 = V2 - (V1)^2 = 178.8064 - (13.3417)^2 = 0.8055$$

$$\sigma = \sqrt{\pi_2} = \sqrt{0.8055} = 0.8975$$

c)— Se aplican las siguientes fórmulas*:

Media = V1
Remplazando: Media = 13.34

Desviación standard = σ
Remplazando: Desviación standard = 0.90

$$\text{Coeficiente de variación} = \frac{\text{Desv. St.}}{\text{Media}} \times 100$$

* *El número de decimales que debe incluírse en los resultados hallados para las constantes depende de la unidad de medida a que se refieren estas constantes. Así, en nuestro ejemplo, las constantes se refieren a gramos de Hb por 100 cc. y esta determinación generalmente se expresa hasta el segundo decimal, o sea hasta el centigramo.*

En las operaciones que se emplean para obtener estas constantes (correspondientes a (b) en nuestro ejemplo) es costumbre calcular hasta el cuarto decimal. Para el coeficiente de variación solo es necesario incluir una cifra decimal en el resultado obtenido.

Cuando la cifra que sigue a la última decimal que se desea incluir es mayor a 5, se añade 1 a la que precede. Así, en nuestro ejemplo, la desviación standard que solo incluye dos decimales, se expresa: 0.90 porque corresponde a 0.8975.

Remplazando: Coef. de var. $= \dfrac{0.90}{13.34} \times 100 = 6.7\,\%$

(d)—Para calcular el error standard de la media y de la desviación standard se aplican las siguientes fórmulas:

$$\text{E. S. de la Media} = \frac{\text{Desv. St.}}{\sqrt{n-1}}$$

n = Nº de datos

Remplazando:

$$\text{E. S. de la Media} = \frac{0.90}{\sqrt{12-1}} = 0.27$$

$$\text{E. S. de la Desviación standard} = \frac{\text{Desv. St.}}{\sqrt{2\,(n-1)}}$$

Remplazando:

$$\text{E. S. de la Desv. St.} = \frac{0.90}{\sqrt{2\,(12-1)}} = 0.19$$

(e)—Los resultados finales obtenidos en el análisis estadístico de los 12 datos de hemoglobina son:

Media $\pm$ E. S. = 13.34 $\pm$ 0.27 gramos por 100 cc.
Desv. St. $\pm$ E. S. = 0.90 $\pm$ 0.19 gramos por 100 cc.
Coef. de var. = 6.7 %
Valores extremos* = 11.90 — 14.60 gramos por 100 cc.

Estos resultados pueden ser presentados en el siguiente cuadro:

---

* *Es conveniente mencionar los valores extremos (menor y mayor) de los datos analizados. En nuestro ejemplo estos son: 11.90 y 14.60 gramos, respectivamente.*

CUADRO 3

DETERMINACIONES DE HEMOGLOBINA EN 12 SUJETOS

	Media $\pm$ E. S.	Desv. St. $\pm$ E. S.	Coef. de var. (%)	Valores extremos
Hemoglobina (gms. por 100 cc.)	13.34 $\pm$ 0.27	0.90 $\pm$ 0.19	6.7	11.90 — 14.60

Cálculo de las constantes de frecuencia y variación cuando el número de datos por analizar es 20 o más.—Los procedimientos que se emplean para calcular la media, desviación standard y el coeficiente de variación (y también la mediana y el modo) serán desarrollados utilizando un ejemplo:

Ejemplo: En 138 sujetos se ha determinado la bilirrubina (en milígramos por 100 cc. de plasma). Los resultados obtenidos están dados en el Cuadro 4.

CUADRO 4

DETERMINACIONES DE BILIRRUBINA
(mgms. por 100 cc. de plasma)
EN 138 SUJETOS

0.59	0.81	0.97	1.88	0.66	0.26
0.75	0.57	1.11	0.73	0.38	0.43
0.45	0.49	0.43	0.86	1.26	0.96
0.75	0.78	0.50	1.36	0.47	0.68
0.90	0.52	0.68	0.59	0.35	0.36
0.61	0.66	0.74	0.42	0.61	0.49
0.54	0.86	0.81	0.54	0.35	0.61
0.70	0.56	0.72	1.44	0.75	0.82
0.97	0.35	1.20	0.56	0.95	0.43
0.47	0.90	0.61	1.03	1.20	0.63
0.86	0.90	1.22	0.61	0.97	0.76
0.64	0.61	0.56	0.68	1.24	1.26
1.20	0.72	1.12	0.61	0.86	0.97
0.38	0.86	0.65	0.54	0.75	0.70
0.47	1.13	0.48	0.52	0.59	0.56
0.64	0.48	0.61	0.48	0.55	0.58
0.54	0.56	0.45	1.23	0.97	0.71
0.77	0.81	0.60	0.49	0.52	1.19
1.20	0.46	0.41	0.62	0.49	0.91
0.60	0.47	0.43	1.61	1.02	0.71
1.50	0.66	1.56	1.21	0.54	0.56
0.47	0.48	0.44	0.49	0.49	0.62
1.26	0.56	0.85	0.96	0.69	0.72

(a)—Se anotan las cifras menor y mayor de los datos.

En este ejemplo son: Menor = 0.26
Mayor — 1.88

(b)—En seguida se construye el Cuadro 5.

CUADRO 5

1	2	3	4	5	6	7
Grupos (mgms por 100 cc.)	Punto medio	Z	*N*	*x*	*Zx*	Zx^2
0.10 — 0.29	0.20	1	1	0	0	0
0.30 — 0.49	...	31	32	1	31	31
0.50 — 0.69	...	45	77	2	90	180
0.70 — 0.89	...	26	103	3	78	234
0.90 — 1.09	...	14	117	4	56	224
1.10 — 1.29	...	15	132	5	75	375
1.30 — 1.49	...	2	134	6	12	72
1.50 — 1.69	...	3	137	7	21	147
1.70 — 1.89	...	1	138	8	8	64
		138			371	1327

Intervalo de grupo: 0.20

Columna 1—Corresponde a los '*grupos de frecuencia*' en los cuales se distribuyen los datos analizados; el valor más bajo de estos está incluído en el primer grupo y el más alto en el último grupo. El número de grupos de frecuencia es arbitrario, dependiendo, en parte, del número y naturaleza de los datos, pero es conveniente que no sean menos de 6 ni más de 25.. El número de grupos empleados no influye en los resultados que se obtienen para las diversas constantes.

La diferencia cuantitativa entre cada grupo de frecuencia se denomina 'intervalo de grupo' cuyo valor, también arbitrario, se anota en la parte inferior del cuadro, pues es utilizado en cálculos posteriores.

En nuestro ejemplo hemos distribuído arbitrariamente los datos en 9 grupos y el intervalo entre ellos es 0.20 (diferencia entre los grupos que principian en 0.10, 0.30, 0.50, etc.). El primero de los grupos que es 0.10 — 0.29 incluye el valor más bajo de los datos analizados: 0.26 y el último grupo que es 1.70 — 1.89 incluye el valor más alto: 1.88.*

---

* *Debe tenerse cuidado de anotar, claramente, los límites de cada grupo de frecuencia para evitar errores. Así, en nuestro ejemplo, el primer grupo incluye datos entre 0.10 y 0.29 miligramos INCLUSIVE; el segundo grupo entre 0.30 y 0.49 miligramos INCLUSIVE, etc.*
Si en cambio, se fijaran los límites de estos grupos como 0.10 — 0.30 y 0.30 — 0.50 miligramos, etc., ante un valor de 0.30 no se sabría si le correspondía el primer o segundo grupo.

Columna 2 (Punto medio)—Corresponde al valor medio del primer grupo de frecuencia. Se calcula añadiendo al límite inferior de este grupo la mitad del intervalo.

En nuestro ejemplo el límite inferior del primer grupo es 0.10 y la mitad del intervalo es también 0.10. Luego:

$$0.10 + 0.10 = 0.20$$

Columna 3 (Z)—Corresponde al número de datos que cae entre cada grupo de frecuencia. Si a uno de estos grupos no corresponde dato alguno se anota θ.

En nuestro ejemplo hay 1 determinación de bilirrubina cuyo valor está entre 0.10 y 0.29, luego se anota 1. Hay 31 determinaciones cuyo valor está entre 0.30 y 0.49, luego se anota 31, y así sucesivamente.

La suma de las cifras anotadas en esta columna debe corresponder al número total de datos analizados que en nuestro ejemplo es 138. Es importante, antes de seguir adelante, verificar esta correspondencia para estar seguro de no haber omitido dato alguno*.

Columna 4 (N)—Corresponde a la suma progresiva de las cifras anotadas en la Columna 3. La última cifra de esta columna debe corresponder al número total de datos analizados, que en nuestro ejemplo es 138.

Nota.—Esta columna es omitida si no se calcula la mediana.

Columna 5 (x)—Corresponde a la numeración de los grupos de frecuencia empezando por 0.

En nuestro ejemplo, en el que hemos usado 9 grupos de frecuencia, el último número de esta columna es naturalmente 8; si hubiéramos usado 16 grupos de frecuencia el último número sería 15, etc.

Columna 6 (Zx)—Cada cifra en esta columna corresponde a la multiplicación de los valores anotados en Columna 3 (Z) por los correspondientes en Columna 5 (x).

En nuestro ejemplo:

Columna 3 Columna 5
Z x

* *Un procedimiento conveniente para llenar esta columna es el siguiente: se tienen anotados los datos por analizar en hoja separada, y conforme se va contando los que corresponden a cada grupo de frecuencia se les tarja, o se les pone una señal al lado. De esta manera, al concluir, puede verificarse fácilmente si se ha omitido dato alguno.*

1	X	0	=	0
31	X	1	=	31
45	X	2	=	90
26	X	3	=	78

y así sucesivamente.

Los valores de esta columna se suman y se anota el resultado. En nuestro ejemplo es 371.

Columna 7 (Zx^2)—Cada cifra en esta columna corresponde a la multiplicación de los valores anotados en Columna 3 (Z) por el cuadrado de los que les corresponden en Columna 5 (x).

En nuestro ejemplo:

Columna 3		Columna 5		
Z		x		
1	X	$(0)^2$	=	0
31	X	$(1)^2$	=	31
45	X	$(2)^2$	=	180
26	X	$(3)^2$	=	234

y así sucesivamente.

Los valores de esta columna se suman y se anota el resultado. En nuestro ejemplo es 1327.

(c)—En seguida se procede a las operaciones siguientes:

$$V1 = \frac{\text{Suma } Zx}{\text{N}^{\circ}\text{ datos}} = \frac{371}{138} = 2.6884$$

$$V2 = \frac{\text{Suma } Zx^2}{\text{N}^{\circ}\text{ datos}} = \frac{1327}{138} = 9.6159$$

$$\pi_2 = V2 - (V1)^2 = 9.6159 - (2.6884)^2 = 2.3884$$
$$\mu_2 = \pi_2 - 0.0833 = 2.3884 - 0.0833 = 2.3051$$

(0.0833 = constante)

$$\sigma = \sqrt{\mu_2} = \sqrt{2.3051} = 1.5182$$

(d)—Se aplican las siguientes fórmulas:

Media = (V1 X intervalo de grupo) + Punto medio del primer grupo

Remplazando: Media = (2.6884 X 0.20) + 0.20 = 0.74

Desviación standard = σ X intervalo de grupo

Remplazando: Desv. st. $= 1.5182 \times 0.20 = 0.30$

$$\text{Coeficiente de variación} = \frac{\text{Desviación standard}}{\text{Media}} \times 100$$

$$\text{Remplazando: Coef. de var.} = \frac{0.30}{0.74} \times 100 = 40.5\%$$

(e)—Para calcular el error standard de la media y de la desviación standard se aplican las siguientes fórmulas:

$$\text{E. S. de la Media} = \frac{\text{Desv. st.}}{\sqrt{n}}$$

(n = Nº de datos)

Remplazando:

$$\text{E. S. de la Media} = \frac{0.30}{\sqrt{138}} = 0.03$$

$$\text{E. S. de la Desviación standard} = \frac{\text{Desv. st.}}{\sqrt{2 \times n}}$$

Remplazando:

$$\text{E. S. de la Desv. st.} = \frac{0.30}{\sqrt{2 \times 138}} = 0.02$$

(f)—Los resultados finales obtenidos en el análisis estadístico de los 138 datos de bilirrubina son:

Media $\pm$ E. S. $= 0.74 \pm 0.03$ miligramos por 100 cc. de plasma.
Desv. St. $\pm$ E. S. $= 0.30 \pm 0.02$ miligramos por 100 cc. de plasma.
Coef. de var. $= 40.5\%$
Valores extremos $= 0.26 — 1.88$ miligramos por 100 cc. de plasma.

Estos resultados pueden ser presentados en el siguiente cuadro:

CUADRO 6

Determinaciones de Bilirrubina en 138 sujetos

	Media ± E. S.	Desv. St. ± E.S.	Coef. de var. (%)	Valores extremos
Bilirrubina (mgms por 100 cc. plasma) . .	0.74 ± 0.03	0.30 ± 0.02	40.5	0.26 — 1.88

Las constantes (media, desviación standard y coeficiente de variación), cuyo cálculo hemos descrito, son generalmente las únicas que se incluyen en la presentación de datos cuantitativos sometidos a un análisis estadístico. Pero puede ser útil, en ciertos casos, la determinación de la mediana y el modo. Los cálculos que se emplean con este fin los aplicaremos al mismo ejemplo.

Ejemplo: 138 determinaciones de Bilirrubina. *(Ver el Cuadro 5).*

A—Para calcular la *mediana* se procede de la manera siguiente:

(1)—El nº total de datos se divide entre 2 : $\dfrac{\text{Suma Z}}{2}$

En nuestro ejemplo : $\dfrac{138}{2} = 69$

(2)—Se busca en la Columna 4 (N) la cifra próxima inferior a la obtenida en (1):

En nuestro ejemplo es 32

(3)—Réstese (2) de (1):

En nuestro ejemplo : 69 — 32 = 37

(4)—Se busca en la Columna 4 (N) la cifra próxima superior a la obtenida en (2):

En nuestro ejemplo es 77

(5)—Réstese (2) de (4):

En nuestro ejemplo : 77 — 32 = 45

(6)—Divídase el resultado obtenido en (3) entre el resultado obtenido en (5):

En nuestro ejemplo : $\frac{37}{45} = 0.8222$

(7)—Multiplíquese el resultado obtenido en (6) por el intervalo de grupo:

En nuestro ejemplo : $0.8222 \times 0.20 = 0.1644$

(8)—Añádase el resultado obtenido en (7) al límite inferior del grupo de frecuencia a que corresponde la cifra encontrada en (4):

En nuestro ejemplo : la cifra de (4) es 77 y el grupo de frecuencia que le corresponde en la Columna 1 es 0.50 — 0.69 . Luego :

$0.1644 + 0.50 = 0.5644$ que es el valor de la mediana.

La fórmula del error standard de la mediana es :

$$\text{E. S. de la Mediana} = 1.2533 \times \frac{\text{Desv. st}}{\sqrt{n}}$$

(1.2533 = constante)

Reemplazando:

$$\text{E. S. de la Mediana} = 1.2533 \times \frac{0.30}{\sqrt{138}} = 0.04$$

B—Para calcular el *modo* se procede de la manera siguiente (*Ver el Cuadro 5*):

Se aplica la siguiente fórmula:

$$\text{Modo} = \text{Lmo} + \left(\frac{\text{fa}}{\text{fa} + \text{fb}} \times \text{c} \right)$$

Lmo	=	límite inferior del grupo de frecuencia que incluye el mayor número de datos (llamado grupo modal) ;*
fa	=	número de datos en el grupo de frecuencia próximo superior al grupo modal;
fb	=	número de datos en el grupo de frecuencia próximo inferior al grupo modal;
c	=	intervalo de grupo

Remplazando en nuestro ejemplo:

Lmo	=	0.50 Límite inferior del grupo de frecuencia 0.50 — 0.69 que incluye el mayor número de datos: 45 (grupo modal) ;
fa	=	26 Número de datos en el grupo de frecuencia próximo superior al grupo modal;
fb	=	31 Número de datos en el grupo de frecuencia próximo inferior al grupo modal;
c	=	0.20 Intervalo de grupo.

Luego:

$$\text{Modo} = 0.50 + \left(\frac{26}{26 + 31} \times 0.20 \right) = 0.59$$

C—Los resultados finales obtenidos son:

Mediana ± E. S.	=	0.66	±	0.04	miligramos por 100 cc. de plasma,
Modo	=	0.59			miligramos por 100 cc. de plasma

* *Si en el cuadro de frecuencia de los datos que se analizan (en nuestro ejemplo :Cuadro 5) hay dos grupos modales, es decir, dos grupos de frecuencia que presentan idéntica concentración máxima de datos, es necesario, para aplicar la fórmula descrita, hacer un cuadro nuevo, variando el número de grupos de frecuencia y la magnitud del intervalo de grupo hasta que se consiga un solo grupo modal.*

CAPITULO II

MEDIDA DE LA ASIMETRIA

Los datos de una serie analizada estadisticamente pueden distribuirse de una manera uniforme alrededor del valor central. En este caso la media y la mediana tiene un valor igual y la serie es *simétrica.* Si la distribución no es uniforme la serie analizada es *asimétrica.* Esta asimetría puede ser *positiva* (tendencia a desplazarse hacia valores bajos), presentando la mediana un valor menor al de la media; o *negativa* (ten-

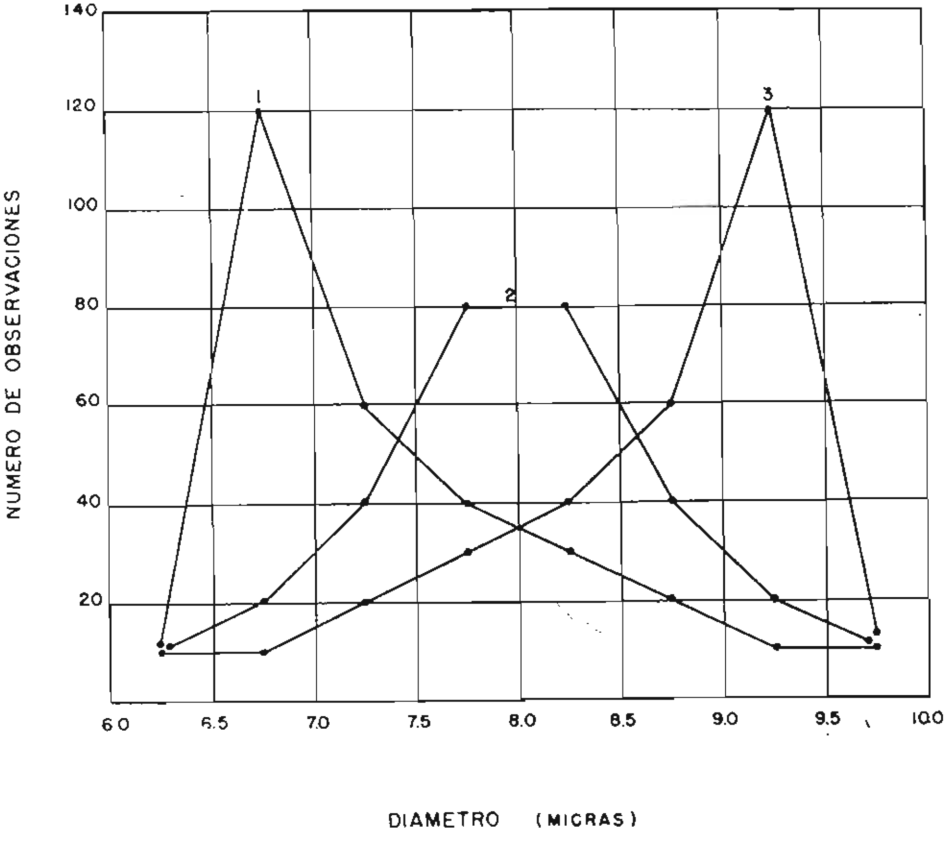

FIGURA 1— Medición del diámetro de 300 hematíes en 3 sujetos.
Curva 1— Corresponde a una distribución asimétrica positiva (construída con los datos del Cuadro 7);
Curva 2— Corresponde a una distribución simétrica;
Curva 3— Corresponde a una distribución asimétrica negativa (construída con los datos del Cuadro 8).

dencia a desplazarse hacia valores altos), teniendo la mediana un valor más alto que la media.

La simetría y asimetría, positiva o negativa, de una serie de datos pueden ser apreciadas gráficamente en la Figura 1.

La diferencia pues entre la media y la mediana permite apreciar la existencia y clase de asimetría, pero, con fines comparativos, puede ser interesante, en deteminados casos, medir cuantitativamente el grado de asimetría, y esto es posible calculando el llamado *coeficiente de asimetría*, representado por el símbolo: a. El coeficiente de asimetría puede variar entre +1 (asimetría máxima positiva) y —1 (asimetría máxima negativa). Cuando es igual a 0 la simetría es perfecta. Un coeficiente de 0.1 indica una asimetría moderada;si es mayor a 0.3 el grado de asimetría es marcado.

Desarollaremos los procedimientos necesarios para el cálculo del coeficiente de asimetría utilizando dos ejemplos uno de los cuales se refiere a un coeficiente de asimetría positivo y el otro a uno negativo.

Ejemplo: En una lámina de sangre se ha determinado el diámetro (en micras) de 300 hematíes. Con los resultados obtenidos se construye el Cuadro 7*.

CUADRO 7

1	2	3	4	5	6	7
Grupos	Punto medio	Z	N	x	Zx	Zx^2
6.00 — 6.49	6.25	10	10	0	0	0
6.50 — 6.99		120	130	1	120	120
7.00 — 7.49		60	190	2	120	240
7.50 — 7.99		40	230	3	120	360
8.00 — 8.49		30	260	4	120	480
8.50 — 8.99		20	280	5	100	500
9.00 — 9.49		10	290	6	60	360
9.50 — 9.99		10	300	7	70	490
		300			710	2550

Intervalo de grupo : 0.50

* *La manera como se construye este cuadro, utilizando los datos originales, está descrita en el Capítulo I (página 140).*

Utilizando los datos del Cuadro 7:

(a)—Se calculan la media y la mediana * obteniéndose los resultados siguientes:

Media	=	7.43 micras
Mediana	=	7.17 micras

(b)—Se calcula el *primer cuartil (Q1)*:

(1)—El nº total de datos se divide entre 4: $\frac{\text{Suma Z}}{4}$

En nuestro ejemplo : $\frac{300}{4} = 75$

(2)—Búsquese en Columna 4 (N) la cifra próxima inferior a la obtenida en (1):

En nuestro ejemplo es 10

(3)—Réstese (2) de (1):

En nuestro ejemplo : 75 — 10 = 65

(4)—Búsquese en Columna 4 (N) la cifra próxima superior a la obtenida en (2). Anótese el grupo de frecuencia en Columna 1 y el número de datos en Columna 3 (Z) que corresponden a esta cifra.

En nuestro ejemplo la cifra próxima superior a 10 (obtenida en (2)) es 130 en la Columna 4 (N), y los valores que corresponden a esta cifra son:

En Columna 1	=	6.50 — 6.99
En Columna 3 (Z)	=	120

(5)—Divídase el resultado obtenido en (3) entre el segundo resultado obtenido en (4):

En nuestro ejemplo : $\frac{65}{120} = 0.5416$

(6)—Multiplíquese el resultado obtenido en (5) por el intervalo de grupo:

---

* *Los procedimientos que se emplean para calcular la media y la mediana, con los datos contenidos en el Cuadro 7, están descritos en el Capítulo I (páginas 138 y 144, respectivamente).*

En nuestro ejemplo : 0.5416 X 0.50 = 0.2708

(7)—Añádase el resultado obtenido en (6) al límite inferior del grupo de frecuencia anotado en (4).
En nuestro ejemplo: la cifra de (6) es 0.2708 y el grupo de frecuencia anotado en (4) es 6.50 — 6.99. Luego:

0.2708 + 6.50 = 6.7708

Q1 = 6.7708

(c)—Se calcula el tercer *cuartil (Q3)*:

(1)—Multiplíquese el número total de datos X 3 y divídase el resultado entre 4 : $\dfrac{\text{Suma Z} \times 3}{4}$

En nuestro ejemplo : $\dfrac{300 \times 3}{4} = 225$

(2)—Búsquese en Columna 4 (N) la cifra próxima inferior a la obtenida en (1):
En nuestro ejemplo es 190

(3)—Réstese (2) de (1):

En nuestro ejemplo : 225 — 190 = 35

(4)—Búsquese en Columna 4 (N) la cifra próxima superior a la obtenida en (2). Anótese el grupo de frecuencia en Columna 1 y el número de datos en Columna 3 (Z) que corresponden a esta cifra.

En nuestro ejemplo la cifra próxima superior a 190, (obtenida en (2)), es 230 en la Columna 4 (N), y los valores que corresponden a esta cifra son:

En Columna 1 = 7.50 — 7.99
En Columna 3 (Z) = 40

(5)—Divídase el resultado obtenido en (3) entre el segundo resultado obtenido en (4):

En nuestro ejemplo: $\dfrac{35}{40} = 0.8750$

(6)—Multiplíquese el resultado obtenido en (5) por el intervalo de grupo:

En nuestro ejemplo: 0.8750 X 0.50 = 0.4375

(7)—Añádase el resultado obtenido en (6) al límite inferior del grupo de frecuencia anotado en (4).

En nuestro ejemplo: la cifra de (6) es 0.4375 y el grupo de frecuencia anotado en (4) es 7.50 — 7.99. Luego:

$$0.4375 + 7.50 = 7.9375$$

$$Q3 = 7.9375$$

(d)—En seguida se aplica la siguiente fórmula:

$$\text{Coeficiente de asimetría} = \frac{Q3 + Q1 - (2 \times \text{Mediana})}{Q3 - Q1}$$

Remplazando en nuestro ejemplo:

$$\text{Coef. de asimetría} = \frac{7.9375 + 6.7708 - (2 \times 7.17)}{7.9375 - 6.7708} = + 0.3157$$

Interpretación—El coeficiente de asimetría de + 0.3157 indica un grado de asimetría positiva marcado en la distribución de los datos correspondientes a 300 diámetros de hematíes.

NOTA.—La curva Nº. 1 de la Figura 1 representa gráficamente la distribución de los datos analizados.

Ejemplo: En una lámina de sangre se ha determinado el diámetro (en micras) de 300 hematíes. Con los resultados obtenidos se construye el Cuadro 8 *

CUADRO 8

1	2	3	4	5	6	7
Grupos	Punto medio	Z	N	x	Zx	Zx^2
6.00 — 6.49	6.25	10	10	0	0	0
6.50 — 6.99		10	20	1	10	10
7.00 — 7.49		20	40	2	40	80
7.50 — 7.99		30	70	3	90	270
8.00 — 8.49		40	110	4	160	640
8.50 — 8.99		60	170	5	300	1500
9.00 — 9.49		120	290	6	720	4320
9.50 — 9.99		10	300	7	70	490
		300			1390	7310

Intervalo de grupo : 0.50

* *La manera como se construye este cuadro, utilizando los datos originales, está descrita en el Capítulo I (página 140).*

Utilizando los datos del Cuadro 3:

(a)—Se calcula la media y la mediana * obteniéndose los resultados siguientes:

Media = 3.57 micras
Mediana = 3.33 micras

(b)—Se calcula el primer cuartil (Q1) ** obteniéndose el resultado siguiente:

Q1 = 8.0625

(c)—Se calcula el tercer cuartil (Q3) obteniéndose el resultado siguiente:

Q3 = 9.2292

(d)—Se aplica la siguiente fórmula:

$$\text{Coeficiente de asimetría} = \frac{Q3 + Q1 - (2 \times \text{Mediana})}{Q3 - Q1}$$

Remplazando en nuestro ejemplo:

$$\text{Coef. de asimetría} = \frac{9.2292 + 8.0625 - (2 \times 8.83)}{9.2292 - 8.0625} = -\,0.3157$$

Interpretación—El coeficiente de asimetría de — 0.3157 indica un grado de asimetría negativa marcada en la distribución de los datos correspondientes a 300 diámetros de hematíes.

NOTA.—La curva Nº 3 de la Figura 1 representa gráficamente la distribución de los datos analizados.

---

* *Los procedimientos que se emplean para calcular la media y la mediana, con los datos contenidos en el Cuadro 8, están descritos en el Capítulo I (páginas 138 y 144 respectivamente).*

** *Los procedimientos que se emplean para calcular el primer cuartil (Q1) y el tercer cuartil (Q3) han sido ya descritos en este capítulo.*

CAPITULO III

MEDIDA DE LA RELACION ENTRE DOS O MAS VARIABLES

Constituye un problema frecuente en el análisis estadístico de dos o más series de datos averiguar si existe alguna relación o asociación entre ellos. Por ejemplo: se ha determinado la capacidad vital (en litros) y la estatura (en centímetros) en 150 sujetos, y se desea averiguar si existe o no relación alguna entre estas dos características; o en 250 muestras de sangre, tomadas en igual número de sujetos, se ha hecho determinaciones del número de hematíes (en millones por mm^3.) y del volumen de los hematíes (en micras^3) y se desea investigar si existe o no relación alguna entre estas dos variables: número y tamaño de los hematíes.

Cuando es necesario relacionar una variable con otras varias, y averiguar con cuál de estas últimas está más relacionada la primera, es aún más importante aplicar métodos estadísticos. Por ejemplo: en los 150 sujetos citados en el párrafo anterior, no sólo se ha determinado la capacidad vital (en litros) y la estatura (en centímetros), sino también el peso (en kilos), el área de superficie corporal (en metros cuadrados) y la circunferencia torácica (en centímetros), y es conveniente establecer, como parte del objetivo de la investigación realizada, con cual de estas medidas físicas (estatura, peso, área de superficie y circunferencia torácica) tiene mayor relación la capacidad vital. Utilizando procedimientos estadísticos es posible determinar, en forma precisa y cuantitativa, la relación de la capacidad vital con cada una de las características físicas citadas para concluír: (a) - si existe relación en grado significativo, y (b) - en caso afirmativo, con cuál es más alta.

Si entre dos series de datos existe una relación ésta puede ser (1) - *lineal* o (2) - *no lineal.*

Relación lineal.—En este caso la relación puede ser representada gráficamente por una línea recta, porque al aumento o disminución en una de las variables corresponde variación porporcional en la otra. Si la variación ocurre en el mismo sentido en ambas variables la relación es *directa;* si, al aumento en una variable corresponde una disminución proporcional en la otra, o vice versa, la relación es *inversa.*

La existencia de una relación lineal entre dos series de datos o variables se determina calculando el llamado *coeficiente de correlación,* representado por el símbolo: *r*, y cuyo valor puede fluctuar entre + 1 (correlación máxima directa) y — 1 (correlación máxima inversa). La correlación, directa o inversa, será tanto mayor cuanto más se acerque el valor de este coeficiente a + 1 o — 1, respectivamente, siendo perfecta si adquiere uno de estos valores. Cuando el coeficiente es 0 no existe correlación alguna entre las dos variables. En general, si *r* es menor a 0.3000, el grado de correlación es muy limitado o ausente; si es 0.5000, o mayor, la correlación existe.

El significado estadístico del coeficiente de correlación es juzgado en relación a su error standard (E. S.), y para tener significado estadístico tiene que ser dos o más veces mayor que este último.

Cuando el coeficiente de correlación entre dos variables es elevado y significativo es posible derivar una fórmula (*ecuación de regresión*), por medio de la cual, conocido el valor de una de las variables se predice el que le debe corresponder en la otra. Esta fórmula puede ser representada gráficamente por una línea recta (*línea de regresión*).

Relación no lineal.—En este caso la relación entre dos variables no puede ser representada gráficamente por una línea recta, sino por una línea curva, porque al aumento o disminución en una de ellas no corresponde variación proporcional, directa o inversa, en la otra.

La existencia de una relación no lineal se averigua calculando la llamada *razón de correlación*, representada por el símbolo: *n*. El valor de la razón de correlación fluctúa entre 0 y 1, siendo la relación expresada tanto más elevada cuanto más se acerque a la unidad.

Cuando se tienen dos series de datos cuya relación se desea averiguar no es posible, a priori, saber si se trata de una relación lineal o no lineal, y por lo tanto es necesario calcular *r* y *n*, y determinar, en seguida, para concluir sobre la clase de relación existente, si la diferencia entre ambos valores tiene significado estadístico. En la práctica, y en la gran mayoría de las publicaciones médicas, para determinar el grado de relación entre dos variables sólo se utiliza el cálculo del coeficiente de correlación (*r*). Sin embargo, en determinado caso, puede ser interesante, o necesario, determinar tanto *r* como *n* por lo que indicaremos, utilizando ejemplos, las operaciones necesarias para calcular ambas medidas de relación.

Significado general del coeficiente de correlación y de la razón de correlación.—Es muy importante tener en cuenta que tanto el coeficiente como la razón de correlación representan una medida de la relación o asociación estadística entre dos variables, pero que en el caso de ser significativos *no indican necesariamente que tal relación tiene un carácter causal.*

Un ejemplo ilustrará fácilmente, este criterio: en el estudio sanitario de varias poblaciones se ha encontrado un coeficiente de correlación significativo entre el número de casos de tuberculosis y el número de sujetos desocupados. Dicho coeficiente, expresando cierto grado de asociación entre ambas variables, no significa que la causa de la enfermedad es la desocupación.

Correlación parcial.—Si se analizan estadísticamente varias series de datos puede ser conveniente establecer la relación entre dos variables manteniendo constante una o varias de las otras. Este procedimiento, denominado *correlación parcial*, es útil para establecer los factores que influyen en determinar tal relación. Por ejemplo: en los 150 sujetos, ya mencionados en párrafos anteriores y en quienes se han hecho diversas mediciones, se ha encontrado cierta relación entre la capacidad vital (en litros) y la estatura (en centímetros), pero se desea averiguar si la relación encontrada es la misma en el caso de que todos los sujetos tuvieran el mismo peso. Es posible, por medio de la correlación parcial, establecer el grado de relación entre la capacidad vital y la estatura, manteniendo un peso idéntico en todos los sujetos estudiados. De esta manera, puede obtenerse cierto conocimiento de la influencia del peso corporal sobre la relación que existe entre la capacidad vital y la estatura.

La aplicación de los métodos correspondientes a la correlación parcial sólo es posible cuando previamente se ha hallado un coeficiente de correlación significativo entre las variables cuya relación se estudia.

Cálculo del coeficiente de correlación (r) entre dos series de datos. Los métodos que se utilizan varían según que el número de observaciones por relacionarse sea mayor o menor a 25. Ambos procedimientos serán descritos utilizando ejemplos separados *.

A—Procedimiento a seguir cuando el número de observaciones es menor a 25 — :

Ejemplo: — Se ha determinado en 12 sujetos el número de hematíes (en millones por mm^3.) y la cantidad de hemoglobina (en gramos por 100 cc.) en la sangre circulante. Se trata de determinar si existe relación alguna entre ambas variables (hematíes y hemoglobina). Los resultados obtenidos en las determinaciones hechas están dados en el Cuadro 9.

* *En los cálculos relacionados con el coeficiente de correlación es frecuente realizar operaciones aritméticas de cantidades con signo diferente. Como guía para efectuar estas operaciones pueden consultarse los ejemplos dados en el Apéndice B.*

CUADRO 9

DETERMINACIONES DE HEMATIES (mill. por mm^3.) Y HEMOGLOBINA (gms. por 100 cc.) EN 12 SUJETOS.

Sujeto	Hematíes (mill. por mm^3)	Hemoglobina (gms. por 100 cc.)
1	4.15	13.20
2	3.89	12.12
3	4.20	13.10
4	4.40	13.50
5	4.45	13.88
6	4.38	14.80
7	4.72	14.60
8	3.97	11.90
9	4.56	14.00
10	3.95	12.10
11	4.31	13.40
12	4.26	13.50

(a)—Con estos datos se construye el Cuadro 10:

CUADRO 10

1	2	3	4	5
x	x^2	y	y^2	xy
4.15	17.2225	13.20	174.2400	54.7800
3.89	15.1321	12.12	146.8944	47.1468
4.20	17.6400	13.10	171.6100	55.0200
4.40	19.3600	13.50	182.2500	59.4000
4.45	19.8025	13.88	192.6544	61.7660
4.38	19.1844	14.80	219.0400	64.8240
4.72	22.2784	14.60	213.1600	68.9120
3.97	15.7609	11.90	141.6100	47.2430
4.56	20.7936	14.00	196.0000	63.8400
3.95	15.6025	12.10	146.4100	47.7950
4.31	18.5761	13.40	179.5600	57.7540
4.26	18.1476	13.50	182.2500	57.5100
51.24	219.5006	160.10	2145.6788	685.9908

Columna 1 (x)—Corresponde a los resultados obtenidos en la numeración de hematíes en los 12 sujetos.
Se suman las cifras de esta columna.

Columna 2 (x^2)—Corresponde al cuadrado de cada una de las cifras de la Columna 1 (x).

$(4.15)^2 = 17.2225$
$(3.89)^2 = 15.1321$
y así sucesivamente.
Se suman las cifras de esta columna.

Columna 3 (y)—Corresponde a los resultados obtenidos en el dosaje de he-hemoglobina en los 12 sujetos.
Se suman las cifras de esta columna.

NOTA—Cada uno de los valores en esta columna, y su correspondiente en Columna 1. deben referirse al mismo sujeto. Es decir, que en la anotación de los datos en ambas columnas hay que seguir el mismo orden del Cuadro 9.

Columna 4 (y^2)—Corresponde al cuadrado de cada una de las cifras de la Columna 3 (y).

$(13.20)^2 = 174.2400$
$(12.12)^2 = 146.8944$
y así sucesivamente.
Se suman las cifras de esta columna.

Columna 5 (xy)—Cada cifra en esta columna corresponde a la multiplicación de los valores anotados en Columna 1 (x) por los correspondientes en Columna 3 (y).

Columna 1		*Columna 2*		*Columna 5*
x		y		xy
4.15	X	13.20	=	54.7800
3.89	X	12.12	=	47.1468
4.20	X	13.10	=	55.0200

y así sucesivamente.

Se suman las cifras de esta columna.

(b)—En seguida se procede a las operaciones siguientes*:

* *Las operaciones descritas en esta sección pueden también ser empleadas para el cálculo de la media, desviación standard y el coeficiente de variación, correspondientes a ambas series de datos: hematíes y hemoglobina (Ver Capítulo I, página 134).*

$$V1x = \frac{\text{Suma } x}{\text{N}^{\circ} \text{ datos}} = \frac{51.24}{12} = 4.2700$$

$$V2x = \frac{\text{Suma } x^2}{\text{N}^{\circ} \text{ datos}} = \frac{219.5006}{12} = 18.2917$$

$$\pi_{2x} = V2x - (V1x)^2 = 18.2917 - (4.2700)^2 = 0.0588$$

$$\sigma\, x = \sqrt{\pi_{2x}} = \sqrt{0.0588} = 0.2425$$

$$V1y = \frac{\text{Suma } y}{\text{N}^{\circ} \text{ datos}} = \frac{160.10}{12} = 13.3417$$

$$V2y = \frac{\text{Suma } y^2}{\text{N}^{\circ} \text{ datos}} = \frac{2145.6788}{12} = 178.8066$$

$$\pi_{2y} = V2y - (V1y)^2 = 178.8066 - (13.3417)^2 = 0.8057$$

$$\sigma\, y = \sqrt{\pi_{2y}} = \sqrt{0.8057} = 0.8976$$

(c)—Se aplica la siguiente fórmula:

$$\text{Coeficiente de correlación } (r) = \frac{A}{B}$$

$$\text{En la que } A = \frac{\text{Suma } x}{\text{N}^{\circ} \text{ datos}} - (V1x \times V1y)$$

$$B = \sigma\, x \times \sigma\, y$$

Remplazando en nuestro ejemplo:

$$A = \frac{685.9908}{12} - (4.2700 \times 13.3417) = 0.1968$$

$$B = 0.2425 \times 0.8976 = 0.2177$$

$$\text{Luego } r = \frac{0.1968}{0.2177} = +\ 0.9040$$

NOTA—El coeficiente de correlación lleva el signo — cuando el valor de $(V1x \times V1y)$ excede al de $\frac{\text{Suma } xy}{\text{N}^{\circ} \text{ datos}}$

Por ejemplo: en un caso dado tenemos los siguientes valores:

$$\frac{\text{Suma xy}}{\text{N}^{\circ}\text{ datos}} = 7.6375$$

$$(V1x \times V1y) = 8.5733$$
$$\sigma x \times \sigma y = 1.1341$$

Luego:

$$r = \frac{7.6375 - 8.5733}{1.1341} = -0.8251$$

(d)—Para calcular el error standard del coeficiente de correlación se aplica la siguiente fórmula:

$$\text{E. S. del Coef. de correlación} = \frac{1 - r^2}{\sqrt{n - 2}}$$

(n = N° de datos)

Remplazando en nuestro ejemplo:

$$\text{E. S. del Coef. de correlación} = \frac{1 - (0.9040)^2}{\sqrt{12 - 2}} = 0.0578$$

(e)—El resultado final obtenido en el cálculo del coeficiente de correlación (r) entre los datos de hematíes y hemoglobina, determinados en 12 sujetos, es el siguiente:

Coef. de correl. (*r*) $\pm$ E. S. = $+ 0.9040 \pm 0.0578$

Interpretación—El coeficiente de correlación de + 0.9040 es mayor a dos veces su error standard (2 X 0.0578 = 0.1156) y por lo tanto es significativo estadísticamente. Indica una elevada correlación positiva directa entre ambas variables: hematíes y hemoglobina.

B—Procedimiento a seguir cuando el número de observaciones es 25 o más:

Ejemplo: En 106 muestras de sangre obtenidas en casos de Enfermedad de Carrión se ha determinado el Volumen globular (en micras ^{3}) y la Hemoglobina globular (en micromicrogramos). Se trata de investigar si existe alguna relación entre el tamaño de los hematíes, representado por el volumen globular, y el contenido de hemoglobina de los hematíes, dado por la hemoglobina globular.

Los resultados obtenidos en las determinaciones están dados en el Cuadro 11.

CUADRO 11

Muestra Nº	Volumen globular (micras^3)	Hb globular (micrmicrgms)	Muestra Nº	Volumen globular (micras^3)	Hb globular (micrmicrgms)
1	114.7	31.6	54	104.4	31.1
2	148.2	44.2	55	95.9	30.5
3	114.7	30.7	56	98.7	31.2
4	118.9	33.1	57	102.0	32.9
5	103.4	27.7	58	90.9	28.5
6	106.3	28.6	59	99.1	31.4
7	92.1	26.3	60	111.3	32.2
8	105.5	30.3	61	130.4	36.6
9	93.4	26.9	62	111.4	30.3
10	85.3	26.1	63	145.0	38.2
11	89.5	27.3	64	152.9	41.8
12	134.4	43.6	65	127.1	36.5
13	123.2	34.2	66	162.7	44.5
14	125.3	36.2	67	143.3	41.8
15	137.1	38.0	68	108.1	29.1
16	138.1	40.1	69	119.1	34.2
17	111.4	34.0	70	76.2	24.3
18	150.3	41.1	71	115.3	34.0
19	122.4	39.2	72	97.4	29.7
20	130.9	34.5	73	98.8	30.5
21	138.1	38.3	74	98.5	30.0
22	125.9	37.6	75	140.3	41.2
23	112.2	32.8	76	108.8	33.2
24	128.0	36.0	77	132.2	39.2
25	120.1	37.2	78	135.5	42.2
26	137.3	42.2	79	103.8	34.3
27	106.9	34.9	80	131.3	39.9
28	94.2	29.5	81	96.0	30.8
29	103.9	32.1	82	100.0	32.1
30	105.8	31.2	83	96.4	30.4
31	112.8	32.3	84	100.4	33.5
32	97.0	30.6	85	108.8	34.7
33	84.1	26.8	86	94.7	31.3
34	110.1	31.9	87	108.4	33.2
35	176.4	42.5	88	101.4	31.0
36	113.8	36.1	89	97.4	28.7
37	102.3	30.3	90	94.4	28.0
38	122.5	33.8	91	101.5	31.8
39	97.5	34.0	92	81.2	26.0

(Continúa)

(Continuación)

Muestra Nº	Volumen globular (micras^3)	Hb globular (micrmicrgms)	Muestra Nº	Volumen globular (micras^3)	Hb globular (micrmicrgms)
40	91.8	31.7	93	87.1	31.4
41	90.1	31.8	94	78.0	24.0
42	117.1	39.0	95	95.8	26.8
43	122.2	36.6	96	85.8	25.8
44	96.7	31.2	97	69.3	22.4
45	110.8	34.1	98	83.5	30.4
46	112.3	29.2	99	86.2	28.0
47	100.5	26.9	100	76.0	24.8
48	112.0	31.4	101	111.2	31.5
49	105.5	29.6	102	91.7	32.1
50	104.0	31.1	103	91.7	29.1
51	80.8	23.6	104	99.7	31.7
52	156.7	42.8	105	87.4	29.4
53	108.7	40.6	106	89.2	29.7

(a)—Con estos datos se construye el Cuadro 12, denominado *Cuadro de Correlación.* Se procede de la manera siguiente:

Grupos de frecuencia de Volumen Globular y de Hemoglobina Globular *— Se anotan el valor más alto y el más bajo de ambas series de datos:

Volumen globular:	69.3	Hemoglobina globular:	22.4
	176.4		44.5

Tanto los datos de volumen globular como los de hemoglobina globular se dividen en cierto número de "grupos de frecuencia", el primero de los cuales incluye el valor más bajo y el último el valor más alto.
El número de grupos de frecuencia. y la magnitud del intervalo que los separa, es arbitrario, pero en un cuadro de correlación es conveniente que no sean menos de 6 ni más de 15.
En nuestro ejemplo, hemos dividido los datos de volumen globular en 12 grupos, el primero de los cuales: 65.0 — 74.9 incluye el valor más bajo de esta serie: 69.3, y el último grupo: 175.0 — 184.9 incluye el valor más alto: 176.4. El intervalo que separa los grupos es 10.0
Los datos de hemoglobina globular también se han dividido en 12 grupos, el primero de los cuales: 21.0 — 22.9 incluye el valor más bajo: 22.4, y el último grupo: 43.0 — 44.9 incluye el valor más alto: 44.5. El intervalo que separa los grupos es 2.0.

NOTA—En nuestro ejemplo es por simple coincidencia que el número de grupos de frecuencia es igual (12) en ambas series de datos. Esto no es necesario.

* *Las instrucciones para dividir una serie de datos en "grupos de frecuencia" están dadas en el Capítulo I, (página 140).*

CUADRO 12

	Hemoglobina globular																	
Volumen globular	21.0 — 22.9	23.0 — 24.9	25.0 — 26.9	27.0 — 28.9	29.0 — 30.9	31.0 — 32.9	33.0 — 34.9	35.0 — 36.9	37.0 — 38.9	39.0 — 40.9	41.0 — 42.9	43.0 — 44.9	1 Zy	2 y	3 Zyy	4 Zyy^2	5 Zxyx	6 Zxyxy
65.0— 74.9	1												1	—6	— 6	36	— 6	+ 36
75.0— 84.9		4	2		1								7	—5	—35	175	—30	+150
85.0— 94.9			4	4	4	5							17	—4	—68	272	—41	+164
95.0—104.9			2	2	8	11	3						26	—3	—78	234	—41	+123
105.0—114.9				1	6	8	6	1		1			23	—2	—46	92	—19	+ 38
115.0—124.9							5	1	1	2			9	—1	— 9	9	+ 9	— 9
125.0—134.9							1	4	1	2		1	9	0	0	0	+17	0
135.0—144.9									2	1	4		7	+ 1	+ 7	7	+23	+ 23
145.0—154.9									1		2	1	4	+ 2	+ 8	16	+15	+ 30
155.0—164.9											1	1	2	+ 3	+ 6	18	+ 9	+ 27
165.0—174.9													0	+ 4	0	0	0	0
175.0—184.9											1		1	+ 5	+ 5	25	+ 4	+ 20
1 Zx	1	4	8	7	19	24	15	6	5	6	8	3	106		—216	884	—60	+602
2 **x**	—6	— 5	— 4	— 3	— 2	—1	0	+ 1	+ 2	+ 3	+ 4	+ 5						
3 Zxx	—6	—20	—32	—21	—38	—24		+ 6	+ 10	+ 18	+ 32	+ 15	—60					
4 Zxx^2	36	100	128	63	76	24		6	20	54	128	75	710					

Los grupos de frecuencia de volumen globular se anotan en la primera columna vertical, a la izquierda, y los que corresponden a la hemoglobina globular en la primera columna horizontal, en la parte superior.

En seguida, en el espacio dividido en casilleros y limitado por líneas gruesas, se anota el número de datos de hemoglobina globular que corresponde a cada grupo de frecuencia de volumen globular. Para esto hay que consultar el Cuadro 11.

Correspondiendo a un volumen globular de 65.0 — 74.9 hay sólo un dato de hemoglobina globular: 22.4. Luego se anota 1 en el casillero correspondiente a 21.0 — 22.9 porque dicho valor corresponde a este grupo de frecuencia.

Correspondiendo a un volumen globular de 75.0 — 84.9 tenemos los siguientes valores de hemoglobina globular:

26.8
23.6
24.3
26.0
24.0
30.4
24.8

De estos datos hay 4 (23.6, 24.3, 24.0 y 24.8) que corresponden al grupo de frecuencia 23.0 — 24.9, luego se anota 4 en el casillero correspondiente a este grupo de frecuencia; hay 2 valores (26.8 y 26.0) que corresponden al grupo de frecuencia 25.0 — 26.9, luego se anota 2 en el casillero respectivo. Finalmente, hay 1 valor (30.4) que corresponde al grupo de frecuencia 29.0 — 30.9, luego se anota 1 en el casillero respectivo.

Y así, sucesivamente, se va anotando el número de datos en cada grupo de frecuencia de hemoglobina globular que corresponde a cada grupo de frecuencia de volumen globular.
En seguida se procede a calcular las otras columnas del mismo Cuadro 12:

Columnas horizontales:

Columna 1 (Σx)—Corresponde a la suma de las cifras anotadas en cada grupo de frecuencia de hemoglobina globular.

Columna (2 (x)—Si el número de grupos de frecuencia es *impar* se coloca el 0 correspondiendo al grupo central, es decir al grupo a cuya izquierda y derecha se encuentre igual número de grupos.
Si el número de grupos de frecuencia es *par* se coloca el 0 correspondiendo a $\frac{\text{N}^{\circ} \text{ grupos de frecuencia}}{2} + 1$.

En nuestro ejemplo el número de grupos de frecuencia de hemoglobina globular es par: 12; luego se coloca el 0 correspondiendo al grupo:

$$\frac{12}{2} + 1 = 7$$

Contando de izquierda a derecha el Grupo 7 de hemoglobina globular es el de 33.0 — 34.9; luego se coloca 0 correspondiendo a este grupo.

En seguida se numera, comenzando por 1 y en forma ascendente, todas las columnas situadas a la izquierda y a la derecha del 0; las de la izquierda llevan el signo — y las de la derecha el signo +

Columna 3 (Zxx)—Corresponde a la multiplicación de cada cifra en Columna 1 (Zx) por su correspondiente en Columna 2 (x).
En nuestro ejemplo:

$$1 \times -6 = -6$$
$$4 \times -5 = -20$$

y así sucesivamente.

Se suman los valores correspondientes a esta columna y se anota el resultado. En nuestro ejemplo es —60.

Columna 4 (Zxx^2)—Corresponde a la multiplicación de cada cifra en Columna 1 (Zx) por su correspondiente en Columna 2 (x), elevada al cuadrado.
En nuestro ejemplo:

$$1 \times (-6)^2 = 36$$
$$4 \times (-5)^2 = 100$$
$$8 \times (-4)^2 = 128$$

y así sucesivamente.

Se suman los valores obtenidos en esta columna y se anota el resultado. En nuestro ejemplo es 710.

Columnas verticales:

Columna 1 (Zy)—Corresponde a la suma de las cifras anotadas en cada grupo de frecuencia de volumen globular.
Se suman las cifras anotadas en esta columna. La suma debe corresponder el número total de observaciones. En nuestro ejemplo: 106.

Columna 2 (y)—Para colocar el 0 en esta columna se procede de idéntica manera a lo indicado en Columna 2 (x), página 163.

En nuestro ejemplo el 0 corresponde al grupo $\frac{12}{2} + 1 = 7$.

El Grupo 7 es, contando de arriba a abajo, el grupo de 125.0—134.9; luego se coloca el 0 correspondiendo a este grupo.

En seguida, se numera, comenzando por 1 y en forma ascendente, las columnas situadas arriba y abajo del 0. Las de arriba llevan el signo — y las de abajo el signo +

Columna 3 (Zyy)—Corresponde a la multiplicación de cada cifra en Columna 1 (Zy) por su correspondiente en Columna 2 (y).
En nuestro ejemplo:

$$1 \times -6 = -6$$
$$7 \times -5 = -35$$
$$17 \times -4 = -68$$

y así sucesivamente.

Se suman los valores correspondientes a esta columna y se anota el resultado. En nuestro ejemplo es —216.

Columna 4 (Zyy2)—Corresponde a la multiplicación de cada cifra en Columna 1 (Zy) por su correspondiente en Columna 2 (y), elevada al cuadrado.
En nuestro ejemplo:

$$1 \times (-6)^2 = 36$$
$$7 \times (-5)^2 = 175$$
$$17 \times (-4)^2 = 272$$

y así sucesivamente.

Columna 5 (Zxyx)—Corresponde a la multiplicación de cada una de las cifras anotadas en los grupos de frecuencia de volumen globular por su correspondiente en Columna 2 (x).
Se procede separadamente con cada línea horizontal, incluída en el espacio limitado por líneas gruesas; se suman los resultados obtenidos y se anota el producto de esta suma en la Columna 5 (Zxyx).
En nuestro ejemplo:

Grupo 65.0—74.9	1											
	X											
Columna 2 (x)	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Resultado:	-6											= —6

Se anota —6 en Columna 5 (Zxyx).

Grupo 75.0—84.9		4	2		1								
		X	X		X								
Columna 2 (x)	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	
Resultado:		—20	—8		—2								= —30

Se anota —30 en Columna 5 (Zxyx)

Grupo 85.0—94.9			4	4	4	5							
			X	X	X	X							
Columna 2 (x)	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	
Resultado:			—16	—12	—8	—5							= —41

Se anota —41 en Columna 5 (Zxyx).

Y así sucesivamente.

Se suman los valores correspondientes a esta columna y se anota el resultado; este debe ser igual al obtenido en la suma de la Columna 3 (Zxx).
En nuestro ejemplo el total de la suma es —60, cifra igual a la obtenida en la suma de los valores correspondientes a la Columna 3 (Zxx).

Columna 6 (Zxyxy)—Corresponde a la multiplicación de cada cifra en Columna 5 (Zxyx) por su correspondiente en Columna 2 (y).
En nuestro ejemplo:

$$
\begin{array}{rcrcr}
-6 & \times & -6 & = & +\ 36 \\
-30 & \times & -5 & = & +\ 150 \\
-41 & \times & -4 & = & +\ 164
\end{array}
$$

y así sucesivamente.

Se suman los valores correspondientes a esta columna y se anota el resultado. En nuestro ejemplo es + 602.

(b)—En seguida se procede a las operaciones siguientes:

$$V1x = \frac{\text{Suma Zxx}}{\text{N}^{\circ}\text{ datos}} = \frac{-60}{106} = -0.5660$$

$$V2x = \frac{\text{Suma Zxx}^2}{\text{N}^{\circ}\text{ datos}} = \frac{710}{106} = 6.6981$$

$$\pi_{2x} = V2x - (V1x)^2 = 6.6981 - (-0.5660)^2 = 6.3777$$

$$\sigma x = \sqrt{\pi 2x} = \sqrt{6.3777} = 2.5254$$

$$V1y = \frac{\text{Suma Zyy}}{\text{N}^{\circ}\text{ datos}} = \frac{-216}{106} = -2.0377$$

$$V2y = \frac{\text{Suma Zyy}^2}{\text{N}^{\circ}\text{ datos}} = \frac{884}{106} = 8.3396$$

$$\pi_{2y} = V2y - (V1y)^2 = 8.3396 - (-2.0377)^2 = 4.1874$$

$$\sigma y = \sqrt{\pi 2y} = \sqrt{4.1874} = 2.0463$$

(c)—Se aplica la siguiente fórmula:

$$\text{Coeficiente de correlación } (r) = \frac{A}{B}$$

$$\text{En la que } A = \frac{\text{Suma Zxyxy}}{\text{N}^{\circ}\text{ datos}} - (V1x \times V1y)$$

$$B = \sigma x \times \sigma y$$

Remplazando en nuestro ejemplo:

$$A = \frac{+602}{106} - (-0.5660 \times -2.0377) = +4.5259$$

$$B = 2.5254 \times 2.0463 = 5.1677$$

$$\text{Luego } r = \frac{+4.5259}{5.1677} = +0.8758$$

(d)—Para calcular el error standard del coeficiente de correlación se aplica la siguiente fórmula:

$$\text{E. S. del Coef. de correlación} = \frac{1 - r^2}{\sqrt{n}}$$

(n = Nº de datos)

Remplazando en nuestro ejemplo:

$$\text{E. S. de Coef. de correlación} = \frac{1 - (0.8758)^2}{\sqrt{106}} = 0.0226$$

(e)—El resultado final obtenido en el cálculo del coeficiente de correlación (*r*) entre los datos de Volumen Globular y Hemoglobina Globular, determinados en 106 casos de Enfermedad de Carrión, es el siguiente:

Coef. de correl. (r) $\pm$ E. S. = 0.8758 $\pm$ 0.0226

Interpretación—El coeficiente de correlación de + 0.8758 es mayor a dos veces su error standard (2 X 0.0226 = 0.0452 y por lo tanto es significativo estadísticamente. Indica una elevada correlación positiva directa entre ambas variables: volumen globular y hemoglobina globular.

Cálculo de la razón de correlación (n) entre dos series de datos.— Los métodos que se emplean para este cálculo serán demostrados utilizando un ejemplo.

Ejemplo: En la sección anterior hemos hallado que el coeficiente de correlación entre el Volumen Globular y la Hemoglobina Globular, características hemáticas determinadas en 106 casos de Enfermedad de Carrión, es significativo, indicando una relación positiva directa entre ambas variables.

Utilizaremos el mismo ejemplo para calcular el valor de la razón de correlación (n) entre ambas características: Volumen Globular y Hemoglobina Globular.

Con los datos contenidos en los Cuadros 11 y 12 se construye el Cuadro 13.

CUADRO 13

1	2	3	4	5	6
Volumen globular Grupos	Hb globular Media	x	x^2	Zy	Zyx^2
65.0 — 74.9	22.4	— 10.5	109.20	1	109.20
75.0 - 84.9	25.7	— 7.2	51.84	7	362.88
85.0 — 94.9	29.0	— 3.9	15.21	17	258.57
95.0 — 104.9	30.9	— 2.0	4.00	26	104.00
105.0 — 114.9	32.3	— 0.6	0.36	23	8.28
115.0 — 124.9	35.7	2.8	7.84	9	70.56
125.0 — 134.9	37.8	4.9	24.01	9	216.09
135.0 — 144.9	40.5	7.6	57.76	7	404.32
145.0 — 154.9	41.3	8.4	70.56	4	282.24
155.0 — 164.9	43.6	10.7	114.49	2	228.98
165.0 — 174.9	—	—	—	0	—
175.0 — 184.9	42.5	9.6	92.16	1	92.16
				106	2137.28

Columna 1 *(Volumen globular-Grupos)*—Corresponde a la división de los datos de Volumen globular en 12 grupos de frecuencia. Estos son idénticos a los dados en la primera columna vertical del Cuadro 12.

Columna 2 *(Hb globular-Media)*—Corresponde a los valores medios o promedios de Hb globular que se refieren a cada grupo de frecuencia de Volumen globular. Para estos cálculos se utiliza el Cuadro 11 con los datos originales.

En nuestro ejemplo:

Correspondiendo a un Volumen globular de 65.0—74.9 hay un sólo dato de Hb globular: 22.4. En este caso, este es el valor medio o promedio de este grupo.

Correspondiendo a un Volumen globular de 75.0—84.9 tenemos los siguientes valores de Hb globular:

26.8
23.6
24.3
26.0
24.0
30.4
24.8

El valor medio de estos 7 datos de Hb globular es igual a la suma de estos datos dividida entre 7:

$$\frac{179.9}{7} = 25.7$$

Y así sucesivamente se van calculando los valores medios de Hb globular que corresponden a cada grupo de frecuencia de Volumen globular.

NOTA—Cuando hay que determinar el valor medio en un número considerable de datos se puede emplear, con el objeto de simplificar los cálculos, los procedimientos indicados en el Capítulo I, página 138.

Columna 3 (x)—A cada una de las cifras de la Columna 2 (Hb globular—Media) se le resta el valor medio que corresponde a todos los datos de Hb globular.

En nuestro ejemplo:

El valor medio de los 106 datos de Hb globular es 32.9. Luego:

22.4 — 32.9 = —10.5
25.7 — 32.9 = — 7.2
29.0 — 32.9 = — 3.9
y así sucesivamente.

NOTA—El valor medio de todos los datos de Hb globular se puede calcular con el Cuadro 11, utilizando los procedimientos indicados en el Capítulo I, página 138.
También es posible calcular este valor medio con los datos contenidos en el Cuadro 12 (cuadro de Correlación) y los correspondientes a la sección (b), página 166. En este caso se aplica la siguiente fórmula: Media de la Hb globular = (V]x X intervalo de grupo) + Punto medio del grupo a que corresponde 0 en x. Remplazando en nuestro ejemplo: Media de la Hb globular =
(—0.5660 X 2) + 34.0 = 32.9

Columna 4 (x^2)—Corresponde al cuadrado de cada una de las cifras de la Columna 3 (x).

En nuestro ejemplo:

(—10.5)^2 = 109.20
(— 7.2)^2 = 51.84
(— 3.9)^2 = 15.21
y así sucesivamente.

Columna 5 (Zy)—Indica el número de datos de Hb globular correspondiente a cada grupo de frecuencia de Volumen globular.
Esta columna es idéntica a la Columna 1 (Zy), del Cuadro 12.

Se suman las cifras anotadas en esta columna; el producto de la suma debe ser igual al número total de datos analizados. En nuestro ejemplo: 106.

Columna 6 (Zyx^2)—Corresponde a la multiplicación de cada una de las cifras en la Columna 5 (Zy) por la correspondiente en Columna 4 (x^2) .
En nuestro ejemplo:

1 X 109.20 = 109.20
7 X 51.84 = 362.88
17 X 15.21 = 258.57
y así sucesivamente

(a)—Se aplica la siguiente fórmula:

$$\text{Razón de correlación } (n) = \frac{A}{B}$$

$$\text{En la que } A = \sqrt{\frac{\text{Suma } Zyx^2}{\text{Suma } Zy}}$$

$$B = \sigma x \times \text{intervalo de grupo}$$

Remplazando en nuestro ejemplo:

$$A = \sqrt{\frac{2137.28}{106}} = 4.4903$$

$$B = 2.5254 \times 2 = 5.0508$$

NOTA—El valor de σx es obtenido de las operaciones referentes al cálculo de coeficiente de correlación (sección (B), página 159). El valor del intervalo de grupo es el que corresponde a los grupos de frecuencia de Hb globular (Cuadro 12).

$$\text{Luego } n = \frac{4.4903}{5.0508} = 0.8890$$

(b)—El valor obtenido para la razón de correlación tiene que ser corregido tomando en cuenta el número de grupos de frecuencia usado en su cálculo.

La fórmula que se usa es la siguiente:

$$n \text{ corregida} = \sqrt{\frac{n^2 - \dfrac{(k-3)}{\text{Suma } Zy}}{1 - \dfrac{(k-3)}{\text{Suma } Zy}}}$$

(n^2 = razón de correlación al cuadrado).
(k = N° de grupos de frecuencia)

Remplazando en nuestro ejemplo, en el que el número de grupos de frecuencia usado en el Cuadro 13 es 12:

$$n \text{ corregida} = \sqrt{\frac{(0.8890)^2 - \dfrac{(12-3)}{106}}{1 - \dfrac{(12-3)}{106}}} = 0.8779$$

(c)—Para que la razón de correlación (n) corregida tenga significado estadístico debe ser mayor a:

$$\sqrt{\frac{k-1}{\text{Suma Zy}}}$$

que representa el valor de la razón de correlación que se obtendría, de acuerdo con los datos analizados, en series prácticamente no relacionadas.

A su vez $\sqrt{\dfrac{k-1}{\text{Suma Zy}}}$ tiene que ser dos o más veces mayor que su error standard para poder ser tomado en cuenta. El error standard se calcula mediante la fórmula:

$$E.\,S. = \frac{1}{\sqrt{\text{Suma Zy}}}$$

Aplicando estas fórmulas a nuestro ejemplo tenemos:

$$\sqrt{\frac{k-1}{\text{Suma Zy}}} \pm \sqrt{\frac{1}{\text{Suma Zy}}} = \sqrt{\frac{12-1}{106}} \pm \frac{1}{\sqrt{106}} = 0.3222 \pm 0.0971$$

Interpretación.—El valor de 0.3222, que representa la razón de correlación que correspondería a series de datos de Volumen globular y Hb globular no relacionados, es significativo puesto que es mayor a dos veces su error standard ($2 \times 0.0971 = 0.1942$).

Pero la razón de correlación corregida: 0.8779 es mayor que 0.3222, y por consiguiente tiene significado estadístico e indica que existe una relación evidente entre ambas variables: Volumen globular y Hemoglobina globular.

Significado de la diferencia entre la razón de correlación (n) y el coeficiente de correlación (r).—Cuando se verifica la existencia de cierto grado de relación entre dos variables, mediante la determinación del coeficiente de correlación y de la razón de correlación, siempre se obtiene para esta última un valor más elevado que para el primero. Esto no significa, necesariamente, que la relación hallada no es lineal. Es preciso determinar si la diferencia entre *n* y *r* tiene significado estadístico; en caso de que la tenga la relación *no es lineal*, y por lo tanto está mejor expresada por *n;* si la diferencia no tiene valor estadístico la relación *es lineal*, es decir, debe ser expresada por *r*, y puede ser representada gráficamente por una línea recta.

Utilizaremos el mismo ejemplo que ha servido para el cálculo del coeficiente de correlación y de la razón de correlación para demostrar los procedimientos relacionados con el significado de la diferencia entre ambos.

Ejemplo: En el estudio de la relación existente entre Volumen globular y Hemoglobina globular, características hemáticas determinadas en 106 casos de Enfermedad de Carrión, hemos encontrado un coeficiente de correlación (*r*) de + 0.8758 y una razón de correlación (*n*) de 0.8890 (Páginas 168 y 171.

Se trata de investigar si la diferencia entre estos valores tiene significado estadístico para concluír acerca de la clase de relación (lineal o no lineal) entre ambas características: Volumen globular y Hemoglobina globular.

Los cálculos que se emplean son los siguientes:

(a)—La diferencia entre n y r está expresada por la fórmula:

$$\mathcal{J}\, n^2 - r^2$$

en la que $n^2 =$ (razón de correlación)^2
$r^2 =$ (coef. de correlación)^2

Remplazando en nuestro ejemplo:

$$\mathcal{J}\, n^2 - r^2 = (0.8890)^2 - (0.8758)^2 = 0.0233$$

(b)—Esta diferencia, para tener significado estadístico, tiene que ser dos o más veces mayor que su error standard.

El error standard se calcula mediante la fórmula:

$$\text{E. S. de } \mathcal{J}\, n^2 - r^2 = 2 \times \sqrt{\frac{\mathcal{J}\, n^2 - r^2}{n}} \times \sqrt{(1-n^2)^2 - (1-r^2)^2 + 1}$$

(n = nº de datos)

Remplazando en nuestro ejemplo:

$$\text{E. S. de } \mathcal{J}\ n^2 - r^2 = 2 \times \sqrt{\frac{0.0233}{106}} \times \sqrt{(1 - 0.8890^2)^2\ (1 - 0.8758^2)^2 + 1}$$

$$= 0.0281$$

(c)—El resultado final es:

$$\mathcal{J}\ n2 - r2 \pm \text{E. S.} = 0.0233 \pm 0.0281$$

Interpretación.—La diferencia de 0.0233 no tiene significado estadístico pues es menor a dos veces su error standard ($2 \times 0.0281 = 0.0562$). Luego a pesar de que n es mayor que r la relación entre Volumen globular, en la serie de datos analizados, tiene carácter *lineal* y debe ser expresada por el coeficiente de correlación (r) que es + 0.8758.

Presentación de *los resultados obtenidos en el estudio de la relación entre dos series de datos.*—Los resultados obtenidos mediante la aplicación de los métodos descritos en este capítulo, los que se refieren a la relación entre el Volumen globular y la Hemoglobina globular, características hemáticas determinadas en 106 casos de Enfermedad de Carrión, pueden ser presentados de la manera siguiente (Cuadro 14):

CUADRO 14

RELACION ENTRE EL VOLUMEN GLOBULAR Y LA HEMOGLOBINA GLOBULAR EN 106 DETERMINACIONES HECHAS EN CASOS DE ENFERMEDAD DE CARRION.

Características relacionadas	*Coeficiente de correlación ± E. S.*	*Razón de correlación*	*Razón de correlación corregida*	$\sqrt{\frac{k-1}{N}} \pm \frac{1}{\sqrt{N}}$	$\mathcal{J}\ n^2 - r^2 \pm$ E. S.
Volumen globular (micras^3) y Hb globular (micromicrogms.)	+ 0.8758 ± 0.0226	0.8890	0.8779	0.3222 ± 0.0971	0.0233 ± 0.0281

NOTA—En resumen, la interpretación estadística del Cuadro 14 es como sigue:

(1)—El coeficiente de correlación y la razón de correlación son significativos, y por consiguiente indican la existencia de una relación entre el Volumen globular y la Hb globular en la serie de casos estudiados;

(2)—La relación tiene carácter líneal, pues la diferencia entre la razón de correlación y el coeficiente de correlación no tiene valor estadístico, y por lo tanto debe ser expresada por el coeficiente de correlación;

(3)—El coeficiente de correlación es bastante elevado y llevando el signo + indica que la relación es directa: a un aumento o disminución en una de las características corresponde, respectivamente, un aumento o disminución proporcional en la otra;

(4)—Siendo líneal la relación entre ambas características, aquella puede ser representada gráficamente por una línea recta (Figura 31), y

(5)—Siendo el coeficiente de correlación significativo y elevado, puede derivarse una ecuación de regresión, por medio de la cual es posible predecir que Hb globular debe corresponder a un Volumen globular dado.

Cálculo de la ecuación de regresión.—Cuando la relación entre dos series de datos o variables es lineal, y está expresada por un coeficiente de correlación elevado, es posible derivar una ecuación, denominada *ecuación de regresión,* por medio de la cual, conocida una de las variables, es posible predecir el valor que le corresponde en la otra.

Ejemplo: En el estudio de la relación existente entre el Volumen globular y la Hemoglobina globular, características hemáticas determinadas en 106 casos de Enfermedad de Carrión, se ha encontrado que la relación entre ambas características es lineal, y está expresada por un coeficiente de correlación elevado: + 0.8758 (página 168).

La ecuación de regresión que permite predecir la Hemoglobina globular que debe corresponder a un Volumen globular dado se calcula de la manera siguiente:

(a)—La fórmula general es la siguiente:

$$X - \text{Media de la Hb globular} = \bar{x}\,(Y - \text{Media del Volumen globular})$$

en la que X = Hbglobular

Y = Volumen globular

$$\bar{x} = \text{Coef. de correl.} \times \frac{\sigma x \times \text{intervalo de grupo de Hb. glob.}}{\sigma y \times \text{intervalo de grupo de Vol. glob.}}$$

Remplazando en nuestro ejemplo en el que:

Media de la Hb globular = 32.9
Media del Volumen globular = 109.6

$$\bar{x} = + 0.8758 \times \frac{2.5254 \times 2}{2.0463 \times 10} = 0.2161$$

NOTA—Los valores medios de Hb globular y de Volumen globular se calculan con los datos contenidos en el Cuadro 11, utilizando los procedimientos indicados en el Capítulo I, página 131.
Los valores correspondientes a σx y σy están dados en la página 159, sección (B).

El intervalo de grupo de Hb globular y de Volumen globular corresponden a los usados en los grupos de frecuencia del Cuadro 12.

Tenemos:

$$X - 32.9 = 0.2161\ (Y - 109.6)$$

Resolviendo esta ecuación para hallar el valor de X:

(1) Pasando —32.9 con signo contrario:

$$X = 0.2161\ (Y - 109.6) + 32.9$$

(2) Multiplicando Y y — 109.6 por 0.2161:

$$X = 0.2161Y - 23.6846 + 32.9$$

(3) Restando 23.6846 de 32.9:

$$X = 0.2161Y + 9.2$$

(b)—Como X representa la Hb globular e Y el Volumen globular, tenemos que la ecuación de regresión es:

Hb globular = (0.2161 × Volumen globular) + 9.2

Por ejemplo, para un Volumen globular de 120 micras^3 debe corresponder una Hb globular de:

(0.2161 × 120) + 9.2 = 35.1 micromicrogramos

Correlación parcial.—Los procedimientos que se emplean para determinar la correlación parcial serán desarrollados utilizando un ejemplo.

Ejemplo: En 40 muestras de sangre se ha determinado el número de hematíes (en millones por mm^3), el Volumen globular (en $micras^3$) y la Hemoglobina globular (en micromicrogramos).

En el análisis estadístico de los resultados obtenidos, y utilizando los procedimientos indicados en este capítulo, se ha hallado coeficientes de correlación significativos y elevados entre las tres características hemáticas.

Representando, para abreviar, cada característica hemática por una cifra:

Número de hematíes = 1; Volumen globular = 2; Hb globular = 3, tenemos que los coeficientes de correlación hallados son los siguientes:

$$r^{12} \pm \text{E. S.} = +\ 0.8813 \pm 0.0362$$
$$r^{13} \pm \text{E. S.} = +\ 0.8685 \pm 0.0399$$
$$r^{23} \pm \text{E. S.} = +\ 0.8212 \pm 0.0528$$

r^{12} = coeficiente de correlación entre el número de hematíes y el Volumen globular;
r^{13} = coeficiente de correlación entre el número de hematíes y la Hemoglobina globular;
r^{23} = coeficiente de correlación entre el Volumen globular y la Hemoglobina globular.

Se desea averiguar cuál sería la relación entre el Volumen globular y la Hemoglobina globular en el caso de que las 40 muestras de sangre tuvieran igual número de hematíes, es decir, el problema por dilucidar es el cálculo de la correlación parcial entre el Volumen globular y la Hemoglobina globular manteniendo constante el número de hematíes

Para este cálculo se procede de la manera siguiente:

(a)—Se aplica la siguiente fórmula:

$$r^{23.1} = \frac{r^{23} - (r^{12} \times r^{13})}{\sqrt{1 - (r^{12})^2} \times \sqrt{1 - (r^{13})^2}}$$

$r^{23}.1$ = coeficiente de correlación parcial entre Volumen globular (2) y Hemoglobina globular (3) manteniendo constante el número de hematíes (1).

Remplazando en nuestro ejemplo:

$$r^{23.1} = \frac{0.8212 - (0.8813 \times 0.8685)}{\sqrt{1 - (0.8813)^2} \times \sqrt{1 - (0.8685)^2}} = 0.2382$$

(b)—El error standard del coeficiente de correlación parcial se calcula mediante la fórmula:

$$\text{E. S. del coef. de correl. parcial} = \frac{1 - (\text{coef. correl. parcial})^2}{\sqrt{n}}$$

(n = número de datos)

Remplazando en nuestro ejemplo:

$$\text{E. S. de } r^{23 \cdot 1} = \frac{1 - (0.2382)^2}{\sqrt{40}} = 0.1492$$

(c)—El resultado final obtenido para el coeficiente de correlación parcial entre el Volumen globular y la Hemoglobina globular, manteniendo constante el número de hematíes:

$$r^{23 \cdot 1} \pm \text{E. S.} = 0.2382 \pm 0.1492$$

Interpretación—El coeficiente de correlación parcial de 0.2382 entre el Volumen globular y la Hemoglobina globular, que se obtiene manteniendo constante el número de hematíes en las 40 determinaciones, es considerablemente menor a 0.8212 (coeficiente de correlación entre las mismas características sin mantener constante el número de hematíes), y, además, no tiene valor estadístico pues es menor a dos veces su error standard ($2 \times 0.1492 = 0.2584$).

Lo que significa que si en todas las muestras de sangre analizadas el número de hematíes fuera igual, no existiría relación alguna entre el Volumen globular y la Hemoglobina globular.

Los mismos procedimientos serían empleados si, utilizando el mismo ejemplo, quisiéramos calcular el coeficiente de correlación parcial entre otras dos variables manteniendo constante la tercera. Así para calcular tal coeficiente entre el número de hematíes (1) y la Hemoglobina globular (3), manteniendo constante el Volumen globular (2):

(a)—Se aplica la siguiente fórmula:

$$r^{13 \cdot 2} = \frac{r^{13} - (r^{12} \times r^{23})}{\sqrt{1 - (r^{12})^2} \times \sqrt{1 - (r^{23})^2}}$$

$r^{13 \cdot 2}$ = coeficiente de correlación parcial entre número de hematíes (1) y Hemoglobina globular (3) manteniendo constante el Volumen globular (2).

Remplazando en nuestro ejemplo:

$$r^{13.2} = \frac{0.8685 - (0.8813 \times 0.8212)}{\sqrt{1 - (0.8813)^2} \times \sqrt{1 - (0.8212)^2}} = 0.3406$$

(b)—El error standard del coeficiente de correlación parcial se calcula mediante la fórmula:

$$\text{E. S. del coef de correl. parcial} = \frac{1 - (\text{coef. correl. parcial})^2}{\sqrt{n}}$$

(n = número de datos)

Remplazando en nuestro ejemplo:

$$\text{E. S. de } r^{13.2} = \frac{1 - (0.3406)^2}{\sqrt{40}} = 0.1398$$

(c)—El resultado final obtenido para el coeficiente de correlación parcial entre el número de hematíes y la Hemoglobina globular, manteniendo constante el Volumen globular:

$$r^{13.2} \pm \text{E. S.} = 0.3406 \pm 0.1398$$

Interpretación—El coeficiente de correlación parcial de 0.3406 entre el número de hematíes y la Hemoglobina globular tiene valor estadístico pues es mayor a dos veces su error standard ($2 \times 0.1398 = 0.2796$), pero es considerablemente menor a 0.8685 (coeficiente de correlación entre las mismas características sin mantener constante el Volumen globular).

Lo que indica que si en todas las muestras de sangre analizadas el Volumen globular fuera el mismo la relación entre el número de hematíes y la Hemoglobina globular sería mucho menos evidente.

CAPITULO IV

SIGNIFICADO ESTADISTICO DE LA DIFERENCIA ENTRE MEDIAS, DESVIACIONES STANDARD, PORCENTAJES Y COEFICIENTES DE CORRELACION

CALCULO DE PROBABILIDADES

Constituye un problema frecuente concluir si la diferencia entre constantes obtenidas en el análisis estadístico de dos o más series de datos tiene significado estadístico, o si este no existe por razones de ser pocos los datos analizados, de haber gran variabilidad en la distribución de los datos, etc. Demostraremos con ejemplos concretos los procedimientos que se emplean con el objeto de determinar el significado estadístico de las diferencias.

Diferencia entre valores medios.—Para que una diferencia entre valores medios tenga significado estadístico debe ser dos o más veces mayor a su error standard. Dos ejemplos servirán para demostrar los cálculos correspondientes.

Ejemplo: En dos grupos de sujetos se ha determinado el número de hematíes en la sangre circulante (en millones por mm^3). Los valores medios, con sus respectivos errores standard, obtenidos en los dos grupos, son los siguientes:

Grupo I = 6.73 $\pm$ 0.13 millones de hematíes por mm^3.
Grupo II = 6.59 $\pm$ 0.11 „ „ „ „ „

Se trata de determinar si la diferencia entre los dos valores medios (6.73 — 6.59 = 0.14 millones de hematíes por mm^3) tiene significado estadístico. Para esto hay que calcular el error standard de la diferencia mediante la fórmula:

$$\text{E. S. de la diferencia} = \sqrt{(\text{E. S. de la Media I})^2 + (\text{E. S. de la Media II})^2}$$

en la que (E. S. de la Media I)^2 = el cuadrado del error standard correspondiente a la media del Grupo I.

(E. S. de la Media II)^2 = el cuadrado del error standard correspondiente a la media del Grupo II.

Remplazando en nuestro ejemplo:

$$\text{E. S. de la diferencia} = \sqrt{(0.13)^2 + (0.11)^2} = 0.17$$

Luego, el resultado final es:

$$\text{Diferencia} \pm \text{E. S.} = 0.14 \pm 0.17$$

Interpretación—La diferencia de 0.14 millones de hematíes por mm^3, entre los valores medios correspondientes a los dos grupos de sujetos, no tiene significado estadístico, pues es menor a dos veces su error standard ($2 \times 0.17 = 0.34$).

Lo que indica que desde un punto de vista estadístico no es justificado concluír que los dos valores medios comparados son diferentes.

Ejemplo: En dos grupos de sujetos se ha determinado la cantidad de CO2 en la sangre arterial (en centímetros cúbicos por 100 cc. de sangre). Los valores medios, con sus respectivos errores standard, obtenidos en los dos grupos, son los siguientes:

Grupo I = 45.68 ± 0.33 cc, de CO2 por 100 cc, de sangre
Grupo II = 33.50 ± 0.18 " " "

Se trata de determinar si la diferencia entre los dos valores medios (45.68 — 33.50 = 12.18) tiene significado estadístico. Como en el ejemplo anterior, se aplica la siguiente fórmula:

$$\text{E. S. de la diferencia} = \sqrt{(\text{E. S. de la Media I})^2 + (\text{E. S. de la Media II})^2}$$

en la que (E. S. de la Media I)^2 = el cuadrado del error standard correspondiente a la media del Grupo I.

(E. S. de la Media II)^2 = el cuadrado del error standard correspondiente a la media del Grupo II.

Remplazando en nuestro ejemplo:

$$\text{E. S. de la diferencia} = \sqrt{(0.33)^2 + (0.18)^2} = 0.38$$

Luego, el resultado final es:

$$\text{Diferencia} \pm \text{E. S.} = 12.18 \pm 0.38$$

Interpretación—La diferencia de 12.18 centímetros cúbicos de CO_2 por 100 cc. de sangre, entre los valores medios correspondientes a los dos grupos de sujetos, tiene significado estadístico, pues es mayor a dos veces su error standard ($2 \times 0.38 = 0.76$).

Lo que indica que desde un punto de vista estadístico hay justificación para concluír que los dos valores medios comparados son diferentes.

Diferencia entre desviaciones standard.—Al igual que en el caso de los valores medios, para que una diferencia entre desviaciones standard tenga significado estadístico debe ser dos o más veces mayor a su error standard.

Ejemplo: En dos grupos de sujetos se ha determinado el número de pulsaciones por minuto. Las desviaciones standard, con sus respectivos errores standard, obtenidos en los dos grupos, son las siguientes:

Grupo II $= 9 \pm 0.19$ pulsaciones por minuto
Grupo I $= 5 \pm 0.16$ „ „

Se trata de determinar si la diferencia entre las dos desviaciones standard ($9 - 5 = 4$) tiene significado estadístico. Para esto hay que calcular el error standard de la diferencia mediante la fórmula:

$$\text{E. S. de la diferencia} = \sqrt{(\text{E. S. de la Desv. st. I})^2 + (\text{E. S. de la Desv. st. II})^2}$$

En la que (E. S. de la Desv. st. I)^2 = el cuadrado del error standard correspondiente a la desviación standard del Grupo I;

(E. S. de la Desv. st. II)^2 = el cuadrado del error standard correspondiente a la desviación standard del Grupo II.

Remplazando en nuestro ejemplo:

$$\text{E. S. de la diferencia} = \sqrt{(0.19)^2 + (0.16)^2} = 0.25$$

Luego, el resultado final es:

$$\text{Diferencia} \pm \text{E. S.} = 4 \pm 0.25$$

Interpretación—La diferencia de 4 pulsaciones por minuto entre las desviaciones standard correspondientes a los dos grupos de sujetos tiene significado estadístico pues es mayor a dos veces su error standard ($2 \times 0.25 = 0.50$).

Lo que indica que desde un punto de vista estadístico hay evidencia para concluír que las dos desviaciones standard comparadas son diferentes.

Diferencia entre porcentajes.—Para que una diferencia entre porcentajes tenga significado estadístico debe también ser dos o más veces mayor a su error standard.

Ejemplo: Una serie de casos con Neumonía han sido tratados mediante la administración de una droga A, mientras que otra serie de casos con la misma enfermedad no ha recibido tal droga. Los resultados observados, en lo que concierne a la mortalidad, son los siguientes:

Serie I—En 76 casos tratados con la droga A han ocurrido 10 muertes:

$$\frac{10}{76} \times 100 = 13.16 \% \text{ de mortalidad}$$

Serie II—En 164 casos no tratados con la droga A han ocurrido 14 muertes:

$$\frac{14}{164} \times 100 = 8.54 \% \text{ de mortalidad}$$

Se trata de determinar si la diferencia entre los dos porcentajes de mortalidad ($13.16 - 8.54 = 4.62$) tiene significado estadístico.
Para concluír a este respecto hay que calcular: (a)— el error standard correspondiente al porcentaje de mortalidad de la Serie I; (b)— el error standard correspondiente al porcentaje de mortalidad de la Serie II; y (c)— el error standard de la diferencia entre los dos porcentajes.

(a)—El error standard del porcentaje de mortalidad de la Serie I se calcula mediante la fórmula:

$$\text{E. S.} = \sqrt{\frac{\text{Porcentaje} \times (100 - \text{porcentaje})}{\text{N}^{\text{o}} \text{ de casos}}}$$

Remplazando en nuestro ejemplo:

$$E.\ S. = \sqrt{\frac{13.16 \times (100 - 13.16)}{76}} = 3.8777$$

(b)—El error standard del porcentaje de mortalidad de la Serie II se calcula mediante la fórmula:

$$E.\ S. = \sqrt{\frac{\text{Porcentaje} \times (100 - \text{porcentaje})}{\text{N}^{\circ}\ \text{de casos}}}$$

Remplazando en nuestro ejemplo:

$$E.\ S. = \sqrt{\frac{8.54 \times (100 - 8.54)}{164}} = 2.1823$$

(c)—El error standard de la diferencia entre los dos porcentajes de mortalidad se calcula aplicando la siguiente fórmula:

$$E.\ S.\ \text{de la diferencia} = \sqrt{(E.\ S.\ \text{de la Serie I})^2 + (E.\ S.\ \text{de la Serie II})^2}$$

Remplazando en nuestro ejemplo:

$$E.\ S.\ \text{de la diferencia} = \sqrt{(3.8777)^2 + (2.1823)^2} = 4{,}4496$$

Luego, el resultado final es:

Diferencia $\pm$ E. S. = 4,62 $\pm$ 4,45

***Interpretación*—La diferencia de 4.62 % no tiene significado estadístico pues es menor a dos veces su error standard (2 × 4.45 = 8.90).**
Lo que indica que desde un punto de vista estadístico no es justificado concluír que el porcentaje de mortalidad en los casos de Neumonía tratados con la droga A sea verdaderamente más elevado que en los casos no tratados con tal droga.

Diferencia entre coeficientes de correlación.— La diferencia entre dos coeficientes de correlación tiene que ser dos o más veces mayor a su error standard para tener significado estadístico.

Ejemplo: En dos grupos de sujetos diferentes se ha relacionado la capacidad vital con el volumen torácico, obteniéndose los siguientes coeficientes de correlación, con sus respectivos errores standard:

Grupo I $= + 0.8754 \pm 0.0325$
Grupo II $= + 0.6353 \pm 0.0487$

Se trata de determinar si la diferencia entre los dos coeficientes de correlación (0.8754 — 0.6353 = 0.2401) tiene significado estadístico. Para esto se aplica la siguiente fórmula:

$$\text{E. S. de la diferencia} = \sqrt{(\text{E. S. de } r^1)^2 + (\text{E. S. de } r^2)^2}$$

en la que $(\text{E. S. de } r^1)^2$ = el cuadrado del error standard del coeficiente de correlación obtenido en el Grupo I;
$(\text{E. S. de } r^2)^2$ = el cuadrado del error standard del coeficiente de correlación obtenido en el Grupo II;

Remplazando en nuestro ejemplo:

$$\text{E. S. de la diferencia} = \sqrt{(0.0325)^2 + (0.0487)^2} = 0.0591$$

Luego, el resultado final es:

$$\text{Diferencia} \pm \text{E. S.} = 0.2401 \pm 0.0591$$

Interpretación—La diferencia de 0.2401, entre los dos coeficientes de correlación, tiene significado estadístico pues es mayor a dos veces su error standard ($2 \times 0.0591 = 0.1182$).
Lo que indica que desde un punto de vista estadístico hay justificación para concluír que los dos coeficientes son diferentes.

CALCULO DE PROBABILIDADES

En la sección anterior de este capítulo hemos indicado que para que una diferencia tenga significado estadístico debe ser dos o más veces mayor a su error standard. Tal procedimiento comparativo basta para concluir sobre tal significado. Sin embargo, si se desea comentar o profundizar aún más la conclusión alcanzada, es posible, conocido el error standard de una diferencia, calcular las probabilidades de que tal diferencia sea debida al azar o casualidad y no corresponda, por consiguiente, a una diferencia real. Tal cálculo de probabilidades se hace consultando el Cuadro 15.

CUADRO 15
CALCULO DE PROBABILIDADES

Diferencia / Error standard	*Probabilidades de encontrar tal diferencia por casualidad*	
	En 100 observaciones	*En relación a 1**
0.6745	50.00 %	1.00 a 1
0.7	48.39	1.07 a 1
0.8	42.37	1.36 a 1
0.9	36.81	1.72 a 1
1.0	31.73	2.15 a 1
1.1	27.13	2.69 a 1
1.2	23.01	3.35 a 1
1.3	19.36	4.17 a 1
1.4	16.15	5.19 a 1
1.5	13.36	6.48 a 1
1.6	10.96	8.12 a 1
1.7	8.91	10.22 a 1
1.8	7.19	12.92 a 1
1.9	5.74	16.41 a 1
2.0	4.55	20.98 a 1
2.1	3.57	26.99 a 1
2.2	2.78	34.96 a 1
2.3	2.14	45.62 a 1
2.4	1.64	60.00 a 1
2.5	1.24	79.52 a 1
2.6	0.932	106.3 a 1
2.7	0.693	143.2 a 1
2.8	0.511	194.7 a 1
2.9	0.373	267.0 a 1
3.0	0.270	369.4 a 1
3.1	0.194	515.7 a 1
3.2	0.137	726.7 a 1
3.3	0.0967	1,033 a 1
3.4	0.0674	1,483 a 1
3.5	0.0465	2,149 a 1
3.6	0.0318	3,142 a 1
3.7	0.0216	4,637 a 1
3.8	0.0145	6,915 a 1
3.9	0.00962	10,930 a 1
4.0	0.00634	15,770 a 1
5.0	0.0000573	1,744,000 a 1
6.0	0.0000002	500,000,000 a 1
7.0	0.00000000026	400,000,000,000 a 1

* 1 representa la probabilidad de encontrar por casualidad la diferencia, y la cifra previa a 1 representa el número correspondiente de probabilidades de no encontrarla.

En el Cuadro 15 encontramos que cuando una diferencia equivale a dos veces su error standard hay 4.55 probabilidades en 100 de encontrar tal diferencia por casualidad, y por cada probabilidad de encontrarla por azar hay casi 21 de no hallarla. Es por este motivo que generalmente se acepta que una diferencia debe ser por lo menos dos veces mayor a su error standard para tener significado estadístico. Hay autores que exigen para aceptar la significación estadística, que una diferencia debe ser por lo menos tres veces mayor a su error standard; en este caso sólo hay 0.27 probabilidades en 100 de encontrar tal diferencia por casualidad y 369 a 1 de no encontrarla.

Los siguientes ejemplos demostrarán la manera de consultar el Cuadro 15 en un caso dado.

Ejemplo: Se ha medido en radiografías el diámetro transverso del corazón en dos grupos de sujetos. Los valores medios, con sus respectivos errores standard, obtenidos son los siguientes:

Grupo I = 18.6 ± 1.04 centímetros
Grupo II = 15.3 ± 0.82 „

Empleando los procedimientos indicados en la página 180 encontramos que la diferencia entre estos valores medios, y el error standard de esta diferencia, son:

$$\text{Diferencia} \pm \text{E. S.} = 3.3 \pm 1.32$$

Esta diferencia es 2.5 mayor a su error standard: $\left(\dfrac{3.3}{1.32} = 2.5\right)$

En el Cuadro 15 hallamos que cuando una diferencia es 2.5 mayor a su error standard hay 1.24 probabilidades en 100 de encontrar tal diferencia por casualidad y por cada probabilidad de encontrarla por azar hay 79.52 probabilidades de no encontrarla.

Luego, basándose sobre este cálculo de probabilidades, hay justificación para considerar significativa la diferencia de 3.3 centímetros entre los dos valores medios.

Ejemplo: Se ha determinado la cantidad de glucosa sanguínea en dos grupos de sujetos. Los valores medios, con sus respectivos errores standard, obtenidos son los siguientes:

Grupo I = 1.16 ± 0.14 gramos por 100 cc. de sangre,
Grupo II = 0.94 ± 0.12 „ „ „

Empleando los procedimientos indicados en la página 180 encontramos que la diferencia entre estos valores medios, y el error standard de esta diferencia son:

$$\text{Diferencia} \pm \text{E. S.} = 0.22 \pm 0.18$$

Esta diferencia es 1.2 mayor a su error standard: $\left(\dfrac{0.22}{0.18} = 1.2\right)$

En el Cuadro 15 hallamos que cuando una diferencia es 1.2 mayor a su error standard hay aproximadamente 23 probabilidades en 100 de encontrar tal diferencia por casualidad, y por cada probabilidad de encontrarla por azar sólo 3.35 probabilidades de no encontrarla.
Luego, según este cálculo de probabilidades, hay justificación para concluír que la diferencia de 0.22 gramos entre los dos valores medios no tiene significado estadístico.

CAPITULO V

REPRESENTACION GRAFICA

Uno de los aspectos más importantes en el análisis estadístico de datos cuantitativos es su representación gráfica adecuada. Esta permite, con frecuencia, una rápida interpretación de los resultados obtenidos, de su variabilidad y tendencia, de la relación entre ellos y facilita, además, el estudio comparativo con otros datos relacionados o similares.

Existen muchos tipos de gráficos o diagramas. Su elección depende fundamentalmente, de la *clase de datos* por representar y del *objeto* del diagrama. Esta segunda consideración tiene importancia especial; el diagrama debe tener un fin demostrativo y no corresponder a un simple adorno de carácter gráfico.

Características generales de los diagramas.—Todos los diagramas estadísticos son representaciones de puntos, líneas, superficies o sólidos, cuya posición en el espacio es definida cuantitativamente por un sistema de coordenadas. Las coordenadas pueden ser de diferente carácter pero las más usadas son las rectangulares, en las cuales la línea horizontal se denomina *abcisa* y la línea vertical *ordenada* (Figura 2).

En esta clase de diagramas generalmente, pero no siempre, la escala en la ordenada se refiere a la frecuencia, relativa o absoluta, de los datos, y aquella en la abcisa a su clase, unidad de medida o secuencia cronológica.

En la construcción de gráficos hay que tener en cuenta que la graduación de las escalas en las coordenadas es factor importante en la apreciación visual de los fenómenos representados gráficamente. Así, si el aumento de hematíes que sigue a una terapia antianémica cualquiera se representa en dos diagramas, en el primero de los cuales la escala en la ordenada corresponde a miles de hematíes por mm^3., y en el segundo a millones por mm^3, es evidente que en el primer diagrama el aumento de hematíes aparece, visualmente, como un proceso mucho más rápido que en el segundo. De igual manera, si en uno de los diagramas el aumento globular se mide en la abcisa con una escala de tiempo referida a semanas, y en el otro con una escala referida a meses, la impresión visual sería de un proceso regenerativo lento en el primer caso y rápido en el segundo.

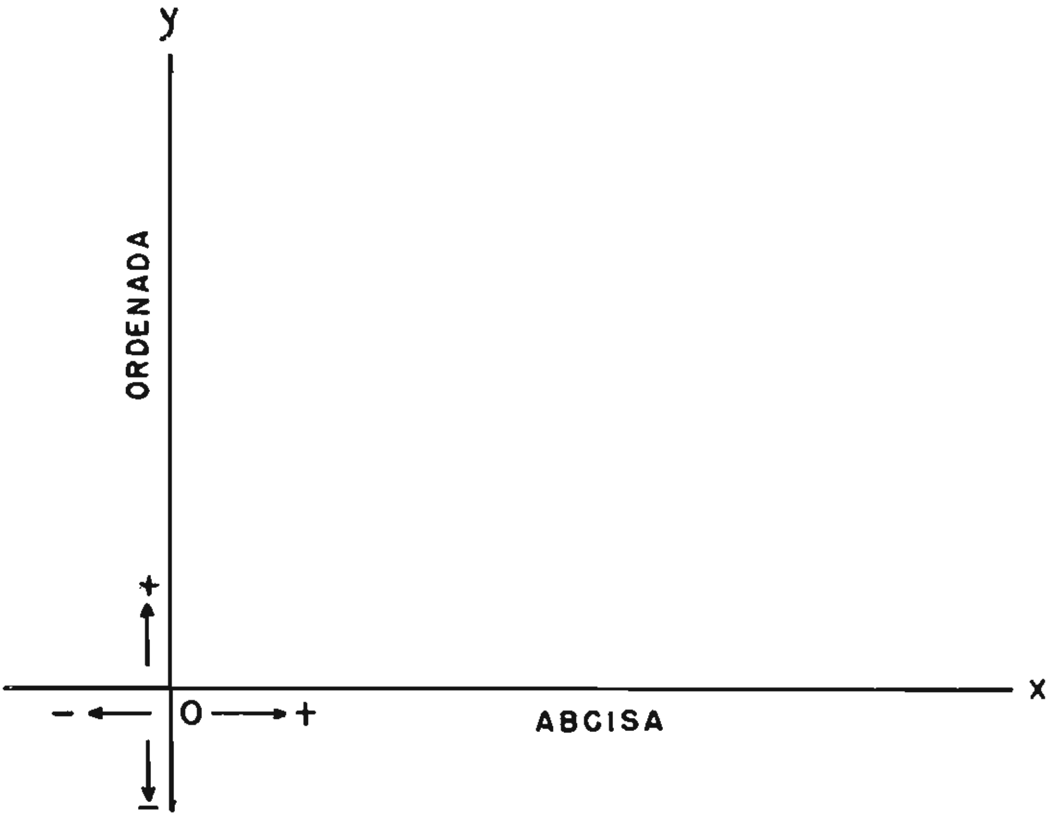

FIGURA 2— Coordenadas rectangulares en un diagrama. Las flechas indican la graduación cuantitativa ascendente a que deben corresponder la escala en ambas coordenadas.

Existen ciertas pautas, adoptadas internacionalmente, * que se refieren a la representación gráfica. Las principales de estas pautas, que tienen importancia en gráficos de carácter médico y biológico, son las siguientes:

* Adoptadas en 1915 por un comité internacional de peritos estadísticos: Joint Committee on Standards for Graphic presentation. Preliminary report. Quart. Publ. Amer. Stat. Ass. 1915, volume 14, page 790.

(1)—Los datos en un diagrama deben ser representados en forma cuantitativa ascendente, de izquierda a derecha en la abcisa, y de abajo a arriba en la ordenada.

(2)—Es más conveniente representar cantidades por magnitudes lineales que por áreas o volúmenes.

(3)—Es conveniente indicar en la ordenada la línea que corresponde a 0. Si esto no es posible, se recomienda romper la escala para señalar la línea 0 en la parte que queda por debajo de la rotura.

(4)—Las líneas correspondientes a 0 y a 100 (en este último caso cuando la escala en la ordenada se refiere a porcentaje) deben ser representadas por líneas gruesas que resalten en el diagrama.

(5)—Cuando se emplea coordenadas con escala logarítmica, las líneas que representan esta escala deben corresponder a potencias de 10.

(6)—Es conveniente no usar más divisiones en las coordenadas que las que son necesarias para la fácil interpretación del diagrama.

(7)—Las líneas que representan las curvas deben ser más gruesas que aquellas que corresponden a la graduación en las coordenadas.

(8)—Cuando una curva representa una serie de observaciones es conveniente señalar en el diagrama la posición de los puntos que corresponden a las observaciones aisladas.

(9)—Las cifras que corresponden a las escalas en ambas coordenadas deben ser escritas claramente a lo largo de las respectivas líneas vertical y horizontal.
La naturaleza de los datos que se representan, así como la unidad de medida empleada, debe también ser indicada en forma concreta; y

(10)—La leyenda debe indicar, en forma clara, lo que representa el diagrama.
Breves descripciones son a veces necesarias para una mejor interpretación del diagrama.

Diversos tipos de diagramas e indicaciones para su elección. — Como ya se ha indicado, en los primeros párrafos de este capítulo, el tipo de diagrama que se elige para la representación gráfica de datos estadísticos depende, principalmente, de la *clase de datos* por representar y del *objeto* del diagrama.

Los principales tipos de diagramas que se usan con mayor frecuencia en la representación gráfica de datos de orden médico, y las indicaciones generales para su empleo, son:

A—Diagramas que representan la variabilidad en frecuencia y la forma general de distribución de los datos:

1—Diagrama de barras horizontales;
2—Histograma;
3—Polígono de frecuencia;
4—Diagrama de frecuencia acumuladas; y
5—Diagrama en coordenadas angulares.

B—Diagramas que representan tendencia y relación:

1—Diagrama de líneas (rectas o curvas) representadas en coordenadas con escala aritmética;
2—Diagrama de líneas representadas en coordenadas con escala logarítmica o semi-logarítmica;
3—Diagrama de dispersión de datos individuales; y
4—Diagrama polar.

Además de los tipos de diagramas mencionados, es frecuente en los laboratorios el uso de diagramas denominados "nomogramas" (desarrollados especialmente por D'Ocagne en 1899—1908), que representan la relación entre dos o más variables en una superficie plana, y que sirven para facilitar la solución numérica de expresiones y relaciones matemáticas complicadas. Considerando las finalidades del nomograma no incluímos en estos apuntes las consideraciones matemáticas que son necesarias para su construcción. Un ejemplo de nomograma está dado en la Figura 3, que corresponde al usado para calcular el área de superficie corporal conociendo la estatura y peso.

Ilustraremos la construcción de los diversos tipos de diagrama mencionados con ejemplos concretos.

A—Diagramas que representan la frecuencia, variabilidad y la forma general de distribución de los datos.

1.—*Diagrama de barras horizontales.*

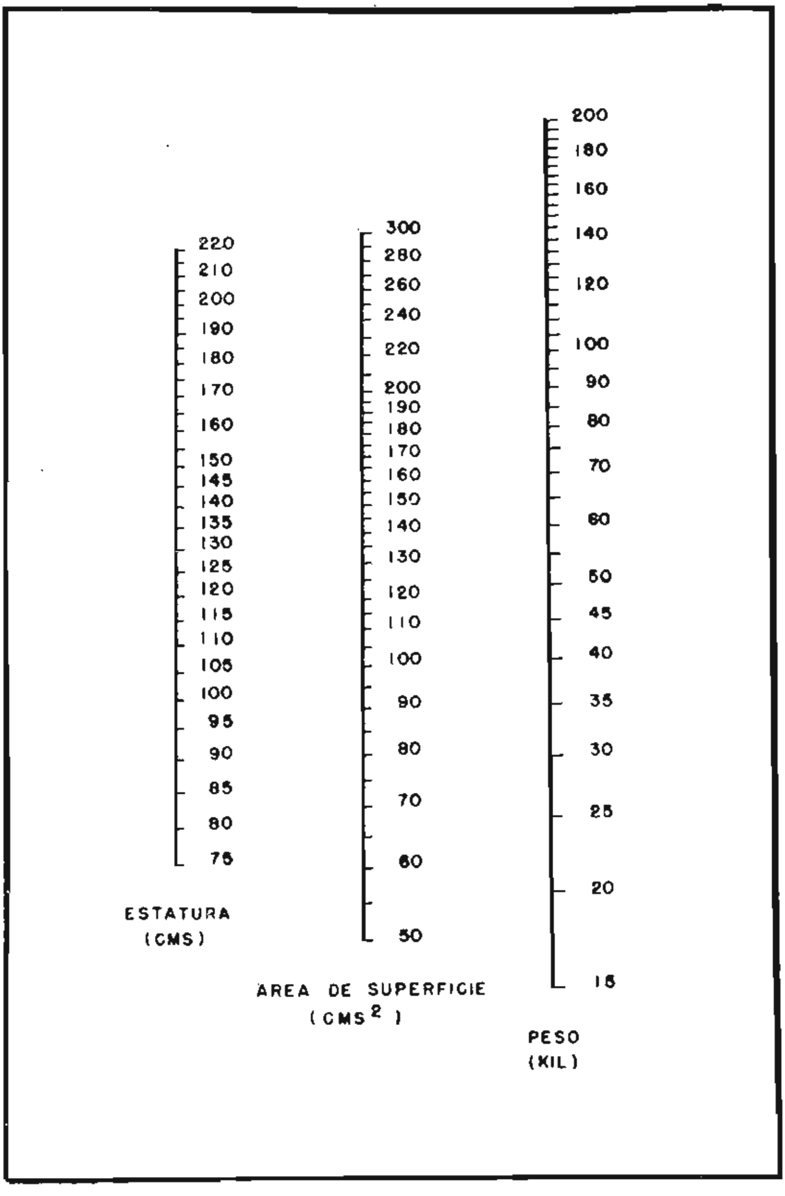

FIGURA 3— Nomograma para calcular el área de superficie corporal conociendo la estatutra y el peso.
Colóquese una regla de tal manera que su borde una los valores correspondientes a la estatura y peso del sujeto; el punto interceptado en la escala central da el área de superficie corporal.

Ejemplo *: Se ha estudiado 94 casos de Icteria Catarral, anotándose en cada caso los síntomas presentes. Se ha calculado, en seguida, la incidencia, en porcentaje, de estos síntomas (así, por ejemplo, dolor abdominal fué un síntoma presente en 60 de los 94 casos; luego $\frac{60}{94} \times 100 = 64\ \%$)

El Cuadro 16 contiene los resultados obtenidos en este estudio.

Se trata de representar gráficamente la incidencia porcentual de los diversos síntomas para indicar, gráficamente, su importancia relativa.

El diagrama de barras horizontales de la Figura 4 corresponde a este propósito.

CUADRO 16

SINTOMATOLOGIA EN ICTERIA CATARRAL.

ESTUDIO DE 94 CASOS

Síntomas	*Frecuencia*	
	Absoluta	*Relativa* %
Icteria	92	98
Orina obscura	92	98
Nauseas	87	93
Acolia	86	91
Anorexia	83	88
Fiebre	77	82
Estreñimiento	75	80
Cefalalgia	71	76
Vómitos	69	73
Postración	66	70
Dolor abdominal	60	64
Escalofríos	48	51
Dolores musculares	39	41
Congestión conjuntivas	35	37
Diarreas	16	17
Hipo	12	13
Epístaxis	6	6
Herpes	3	3

* Tomado, con algunas modificaciones, de Introduction to Medical Biometry and Statistics. R. Pearl — W. B. Saunders Co., Philadelphia, 1930.

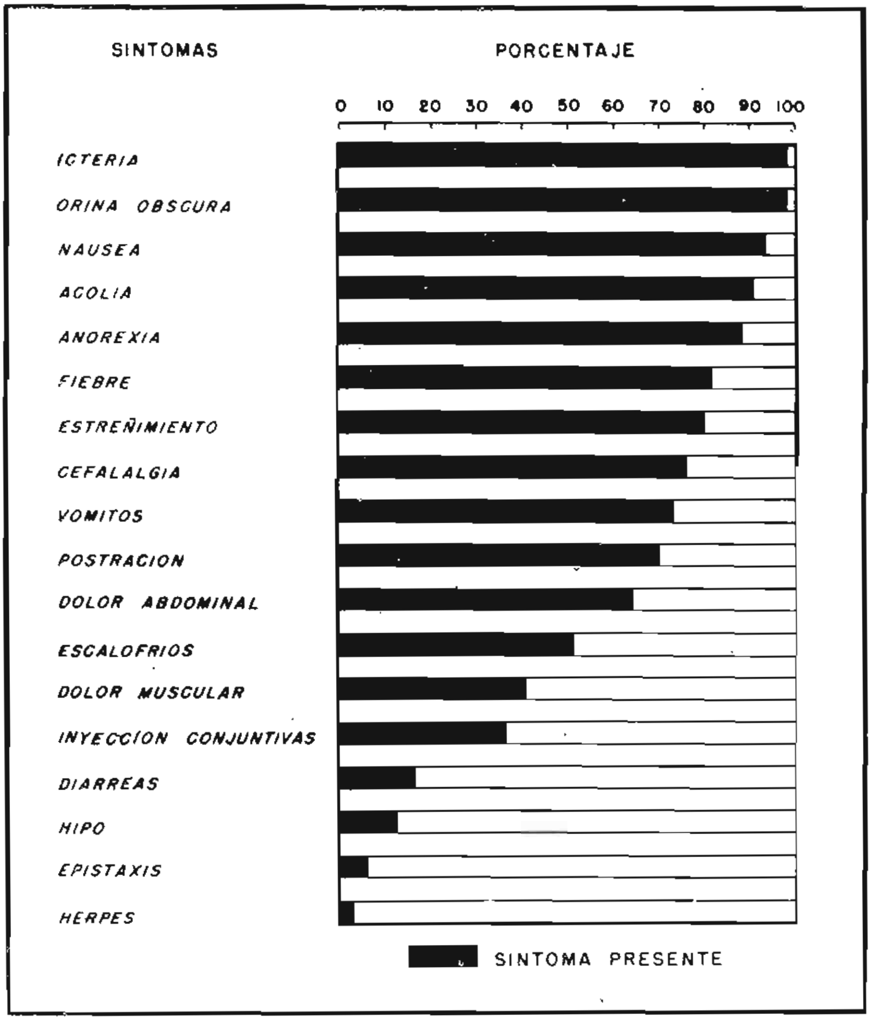

FIGURA 4— Frecuencia relativa de diversos síntomas en Icteria Catarral. Estudio hecho en 94 casos.
(Diagrama de barras horizontales construído con los datos del Cuadro 16).

2—*Histograma.*—El histograma puede representar la distribución de frecuencia absoluta o relativa (en porcentaje) de una serie de datos. Cuando se desea comparar gráficamente, por medio de este diagrama, la distribución de frecuencia de dos series que difieren en el número de datos incluídos en cada una de ellas, es necesario representar la frecuencia relativa. Daremos dos ejemplos, correspondiendo el primero a la representación de la frecuencia absoluta y el segundo de la frecuencia relativa.

Ejemplo: Se ha determinado, en Oroya, la tensión arterial sistólica en 458 casos de Silicosis, cuya edad varía de 20 a 54 años.

Distribuídos los resultados obtenidos en grupos de frecuencia (Cuadro 17), se trata de representar gráficamente su frecuencia absoluta.

El histograma que representa gráficamente los datos del Cuadro 17 puede construírse de diferentes maneras, tal como está ilustrado en Figuras 5, 6 y 7.

CUADRO 17

TENSION ARTERIAL SISTOLICA EN SILICOSIS

OBSERVACIONES HECHAS EN OROYA EN 458 CASOS

(edad: 20 — 45 años)

Tensión sistólica (mmHg)	*Número de observaciones*
80 — 89	15
90 — 99	48
100 — 109	154
110 — 119	123
120 — 129	71
130 — 139	32
140 — 149	7
150 — 159	4
160 — 169	2
170 — 179	1
180 — 189	1

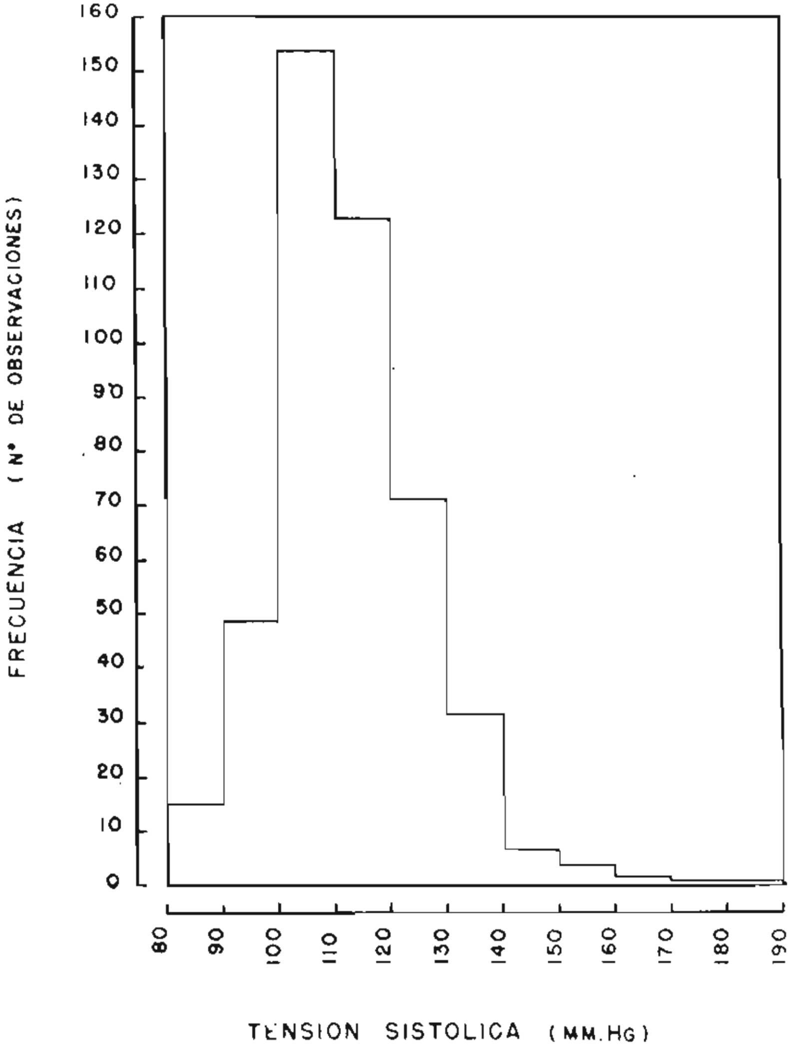

FIGURA 5— Tensión arterial sistólica en 458 casos de Silicosis (Edad: 20—45 años) estudiados en Oroya.
(Histograma construído con los datos del Cuadro 17).

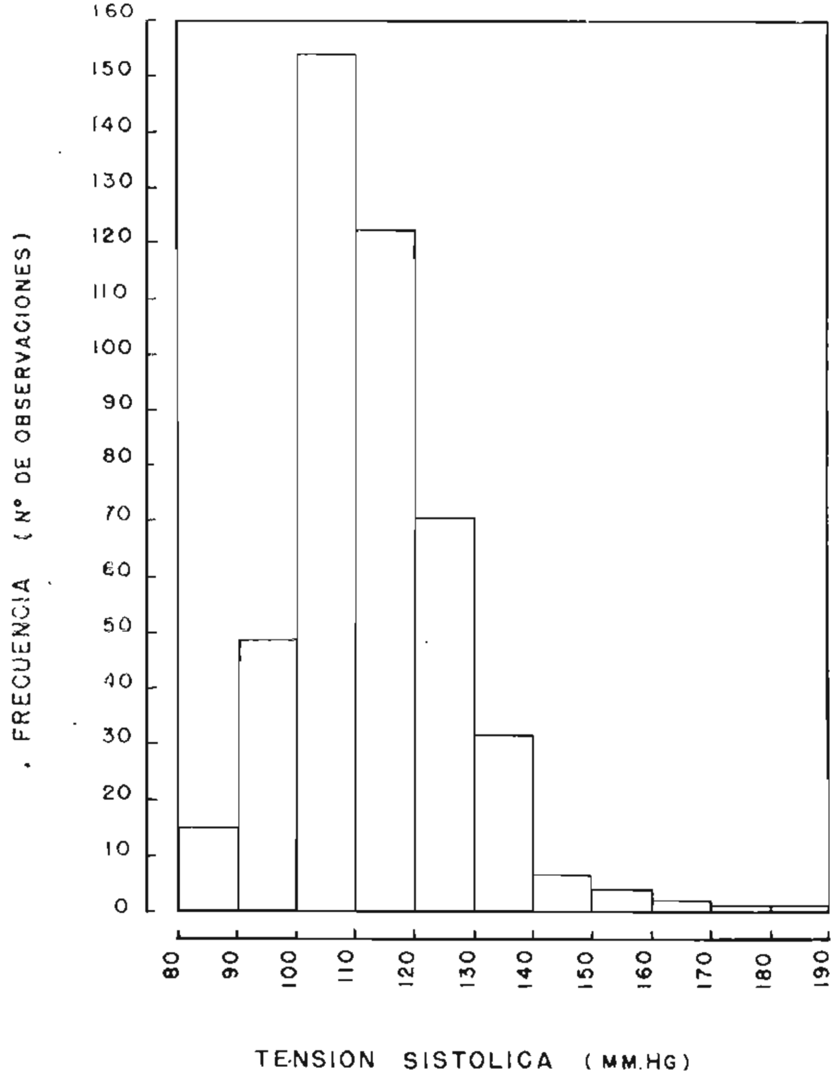

FIGURA 6— Tensión arterial sistólica en 458 casos de Silicosis (Edad: 20—45 años) estudiados en Oroya.
(Histograma construído con los datos del Cuadro 17).

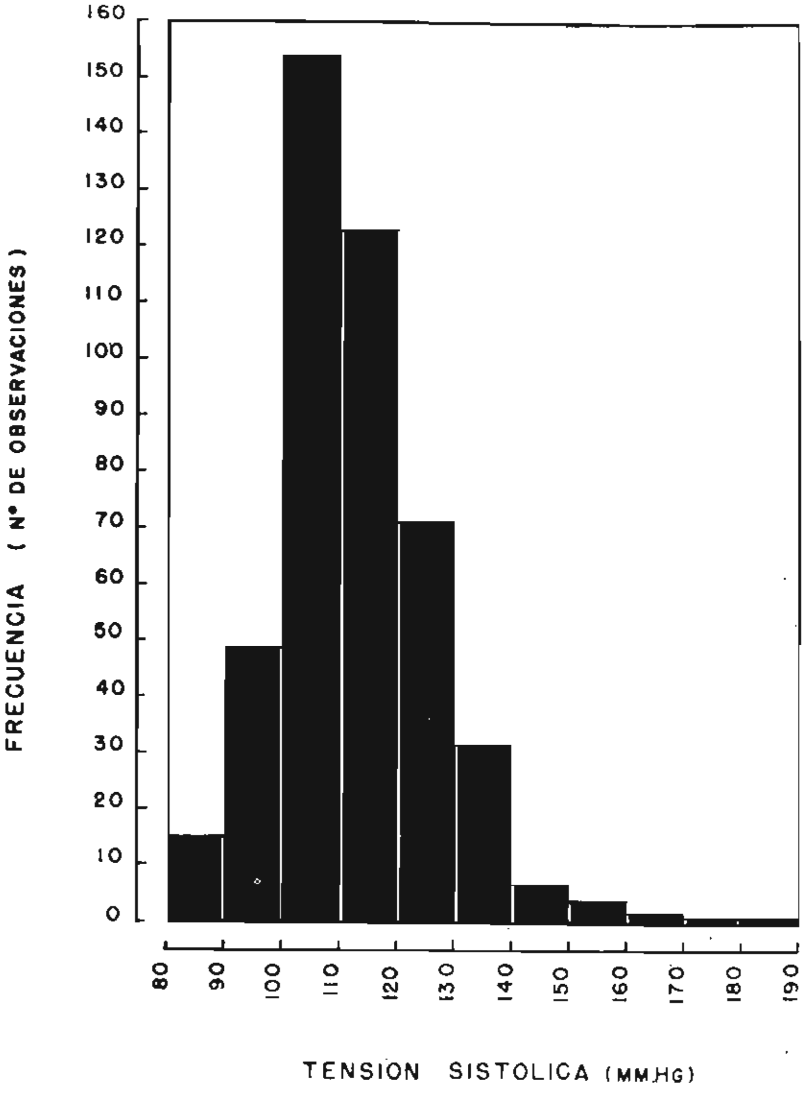

FIGURA 7— Tensión arterial sistólica en 458 casos de Silicosis (Edad: 20—45 años) estudiados en Oroya.
(Histograma construído con los datos del Cuadro 17).

Ejemplo: Se ha determinado, en Lima, la cantidad de hemoglobina en la sangre en dos grupos de sujetos:

Grupo I — 42 sujetos nacidos en lugares elevados y con un tiempo de residencia en Lima de menos de un año;

Grupo II — 84 sujetos nacidos en lugares elevados y con un tiempo de residencia en Lima de más de un año.

Los resultados obtenidos, distribuídos en grupos de frecuencia, están dados en el Cuadro 18.

Los histogramas de la Figura 8, que representan gráficamente, y en forma comparativa, los resultados porcentuales obtenidos en ambos grupos, permiten apreciar la tendencia a un mayor nivel de hemoglobina sanguinea en el grupo de sujetos con un tiempo de residencia más prolongada en Lima.

CUADRO 18

DETERMINACIONES DE HEMOGLOBINA EN SUJETOS ADULTOS SANOS NACIDOS EN LUGARES ELEVADOS.

OBSERVACIONES HECHAS EN LIMA.

Hemoglobina (gms por 100 cc)	GRUPO I *Sujetos con un tiempo de residencia en Lima de menos de 1 año*		GRUPO II *Sujetos con un tiempo de residencia en Lima de más de 1 año*	
Nº de sujetos:	42		84	
	Nº	%	Nº	%
12.0 — 12.9	2	4.8	0	0
13.0 — 13.9	9	21.4	1	1.2
14.0 — 14.9	18	42.8	11	13.1
15.0 — 15.9	8	19.0	36	42.8
16.0 — 16.9	3	7.1	26	30.9
17.0 — 17.9	2	4.8	9	10.7
18.0 — 18.9	0	0	1	1.2

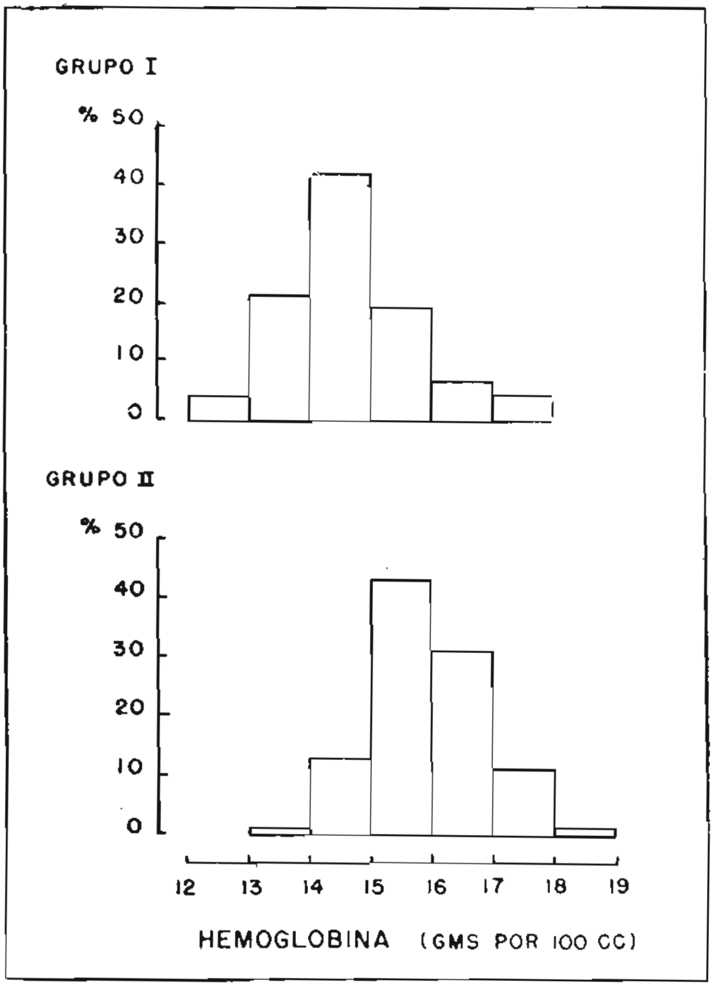

FIGURA 8— Determinaciones de hemoglobina (gramos por 100 cc. de sangre) en hombres adultos sanos, nacidos en lugares elevados y residentes en Lima: Grupo I— Con un tiempo de residencia menor a un año; Grupo II— Con un tiempo de residencia mayor a un año. (Histogramas construídos con los datos del Cuadro 18).

3.—*Polígono de frecuencia.*—Se emplea el polígono de frecuencia para representar, al igual que en el histograma, la frecuencia absoluta o relativa (en porcentaje) de una serie de datos. El segundo procedimiento es necesario cuando se desea comparar gráficamente la distribución de frecuencia en dos series que difieren en el número de datos que incluye cada una.

Ejemplo: Se ha medido la estatura de 606 sujetos adultos indígenas, nacidos en diversos lugares de la Sierra y residentes en Oroya.

Los resultados obtenidos, distribuídos en grupos de frecuencia, están dados en el Cuadro 19.

La Figura 9 corresponde al polígono de frecuencia que representa gráficamente los datos contenidos en el Cuadro 19.

CUADRO 19

ESTATURA DE 606 SUJETOS ADULTOS (EDAD: 20-49 AÑOS) NACIDOS EN DIVERSOS LUGARES DE LA SIERRA Y RESIDENTES EN OROYA

Estatura (centímetros)	Número de observaciones
138 — 142	1
143 — 147	16
148 — 152	76
153 — 157	181
158 — 162	198
163 — 167	105
168 — 172	26
173 — 177	3

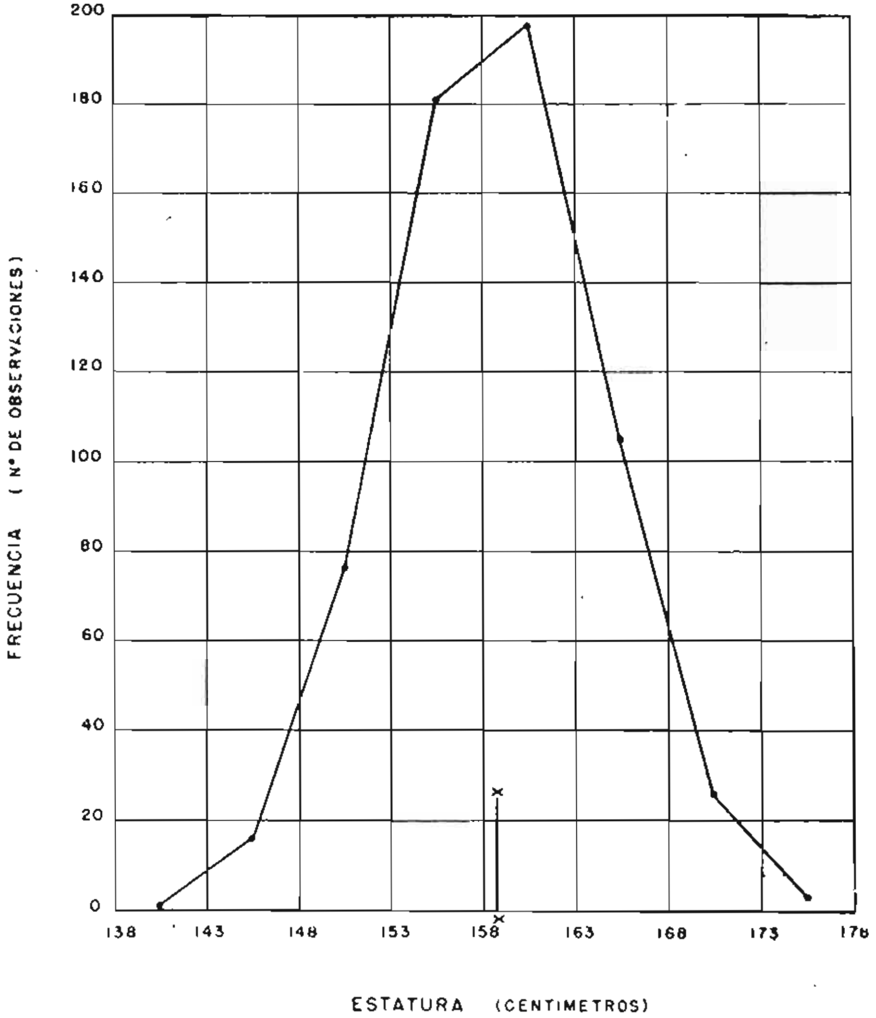

FIGURA 9— Estatutra de 606 hombres adultos (edad: 20—45 años) nacidos en diversos lugares de la Sierra y residentes en Oroya. La línea x————x corresponde al valor medio. (Polígono de frecuencia construído con los datos del Cuadro 19).

Ejemplo: En muestras de sangre obtenidas en 100 sujetos adultos sanos, resi-
po de sujetos las mediciones se hicieron en un total de 30,400 hematíes
dentes en Lima, y en 40 sujetos, de iguales características, residentes
en Oroya, se ha medido el diámetro de los hematíes. En el primer gru-
y en el segundo grupo en un total de 12,000 hematíes.

Los resultados obtenidos, distribuídos en grupos de frecuencia, están contenidos en el Cuadro 20.

La Figura 10 corresponde a los polígonos de frecuencia que indican la frecuencia relativa (en porcentaje) de los diámetros globulares en los dos grupos de sujetos. Esta representación gráfica permite apreciar la tendencia a un mayor diámetro globular en los sujetos residentes en la altura (Oroya).

CUADRO 20

DIAMETRO DE LOS HEMATIES EN SUJETOS ADULTOS SANOS RESIDENTES EN LIMA Y OROYA

Diámetro (micras)	*Residentes en Lima*		*Residentes en Oroya*	
Nº de sujetos:	100		40	
Nº de hematíes medidos:	30,400		12,000	
	Nº	%	Nº	%
5.00 — 5.49	12	0.04	1	0.01
5.50 — 5.99	73	0.2	3	0.03
6.00 — 6.49	813	2.6	80	0.7
6.50 — 6.99	6,276	22.1	1,026	8.5
7.00 — 7.49	5,540	18.2	1,880	15.7
7.50 — 7.99	12,500	41.2	3,576	29.8
8.00 — 8.49	4,267	14.0	4,083	34.0
8.50 — 8.99	317	1.0	888	7.4
9.00 — 9.49	140	0.4	329	2.7
9.50 — 9.99	11	0.03	101	0.8
10.00 — 10.49	1		30	0.2
10.50 — 10.99	0		2	0.02
11.00 — 11.49	0		1	0.01

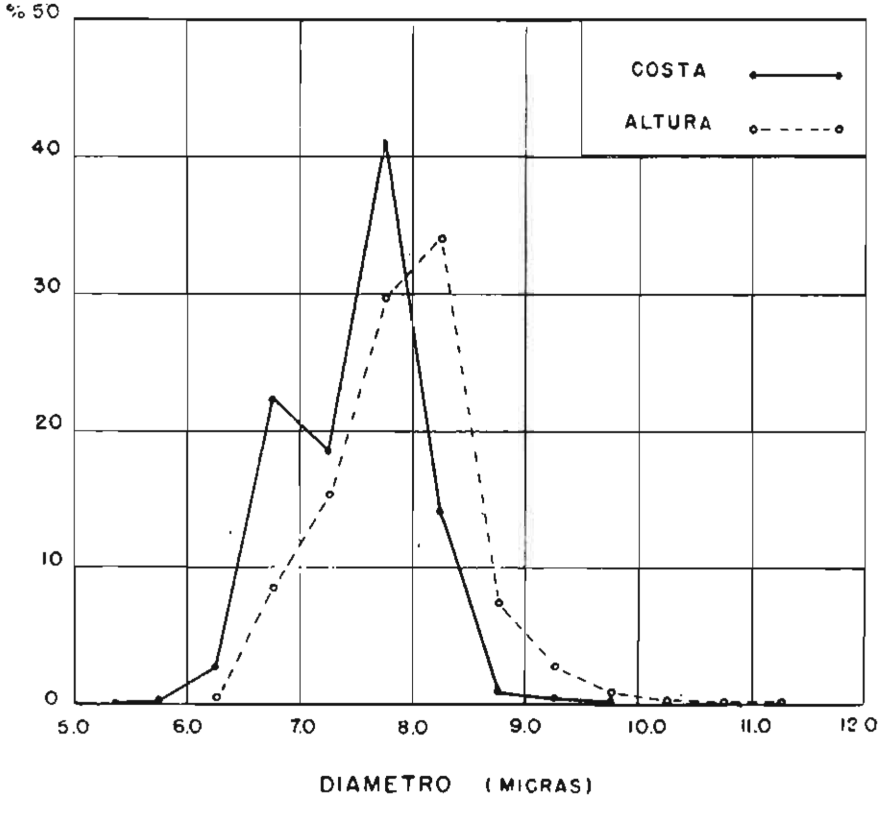

FIGURA 10— Diámetro de los hematíes en dos grupos de hombres adultos sanos: A—nacidos en la costa y residentes en Lima (30,400 hematíes medidos en 100 sujetos) y B— nacidos en lugares elevados y residentes en Oroya (12,000 hematíes medidos en 40 sujetos). (Polígonos de frecuencia construídos con los datos del Cuadro 20).

4.—*Diagrama de frecuencia acumuladas.*—En los gráficos anteriores (histograma y polígono de frecuencia) se ha representado el número o porcentaje de observaciones que corresponden a cada uno de los grupos de frecuencia en los que previamente se han distribuído las series de datos analizados. En el diagrama de frecuencias acumuladas se va sumando las frecuencias que corresponden a cada grupo, y la curva representa gráficamente esta adición progresiva hasta llegar al número total de observaciones.

Este diagrama puede construírse de tal manera que represente acumulativamente, tanto la frecuencia absoluta como la relativa (en porcentaje), de los datos u observaciones analizadas.

Ejemplo: Se ha determinado la viscosidad de la sangre en 48 casos de Silicocosis residentes en Oroya.

Los resultados obtenidos están dados en el Cuadro 21, que contiene, además del número de observaciones o frecuencias que corresponden a cada grupo, la suma progresiva de tales frecuencias.

La curva de la Figura 11, construída con los datos del Cuadro 21, representa la acumulación progresiva de las frecuencias, en forma absoluta y relativa (porcentaje).

CUADRO 21

VISCOSIDAD DE LA SANGRE EN 48 CASOS DE SILICOSIS. OBSERVACIONES HECHAS EN OROYA.

Viscosidad (Unidades)	*Número de observaciones*	*Frecuencia acumulada*
6 — 8	2	2
9 — 11	7	9
12 — 14	8	17
15 — 17	7	24
18 — 20	8	32
21 — 23	7	39
24 — 26	3	42
27 — 29	3	45
30 — 32	3	48

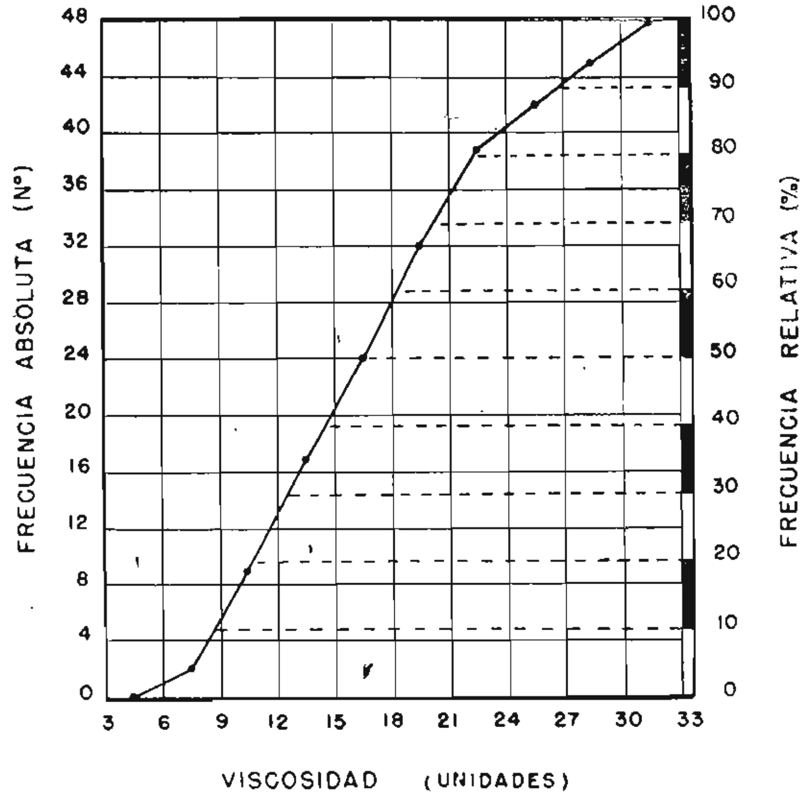

FIGURA 11—Viscosidad de la sangre en 48 casos de Silicosis estudiados en Oroya. (Diagrama de frecuencias acumuladas construído con los datos del Cuadro 21. La escala a la izquierda corresponde a la frecuencia absoluta y su cifra máxima equivale al número total de observaciones: 48 en este ejemplo. La escala a la derecha corresponde a la frecuencia relativa y el 100% equivale al total de observaciones; esta escala que se gradúa dividiendo la distancia entre 0 y 100 en 10 divisiones iguales, permite apreciar, en porcentaje, la distribución de las observaciones. Así, en este diagrama se nota que aproximadamente un 50% de las determinaciones de viscosidad corresponden a valores por encima de 16 unidades).

5.—*Diagrama en coordenadas angulares*—Esta clase de diagrama es poco usado en publicaciones de orden médico. Es difícil de construír y de ser apreciado gráficamente con precisión. Algunos autores recomiendan que su uso sea restringido a trabajos de exhibición y divulgación popular.

Ejemplo *: En determinada ciudad se ha investigado el grupo sanguíneo en (1)—Sujetos de raza blanca y mestiza; y (2)—Sujetos de raza india.

Los resultados obtenidos están dados en el Cuadro 22.

La Figura 12 representa, en coordenadas angulares, la frecuencia relativa de los diversos grupos sanguíneos en los dos grupos de sujetos estudiados, de acuerdo con lo indicado en el Cuadro 22.

CUADRO 22

GRUPOS SANGUINEOS EN (1)—. SUJETOS DE RAZA BLANCA Y MESTIZA, Y (2)—. SUJETOS DE RAZA INDIA.

OBSERVACIONES HECHAS EN 1,979 HOMBRES.

Grupos Sanguineos	*Sujetos de raza blanca y mestiza*		*Sujetos de raza india*	
Nº de observaciones	104		1,875	
	Nº	%	Nº	%
Grupo O	58	55.8	1,567	83.6
Grupo A	25	24.0	109	5.8
Grupo B	21	20.2	79	4.2
Grupo AB	0	0	120	6.4

* Tomado del Boletín Ofic. Sanitaria PanAm., 1944, 23: 1001.

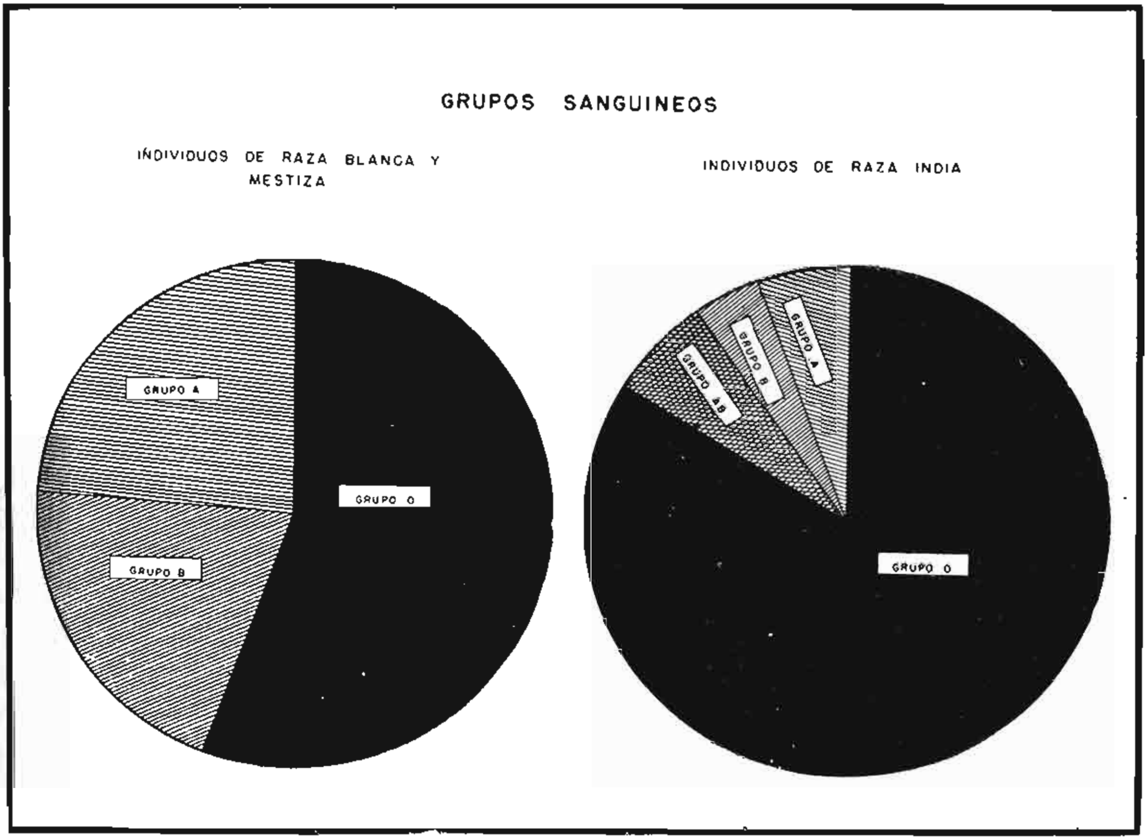

FIGURA 12— Grupos sanguíneos en A— Sujetos de raza blanca y mestiza; y B— Sujetos de raza india.
(Diagrama en coordenadas angulares construído con los datos del Cuadro 22. Teniendo en cuenta que el círculo tiene 360°, el cálculo del número de grados que corresponde a determinada frecuencia en porcentaje se hace multiplicando por 3.6° la cifra respectiva. Así, en nuestro ejemplo, el Grupo sanguíneo 0 que tiene una frecuencia de 55.8% en el Grupo A de sujetos debe estar representado en el círculo por 201° (55.8 × 3.6 = 201°). Para la división del círculo en grados se hace uso de un goniómetro).

B—Diagramas que representan tendencia y relación.

1—*Diagrama de líneas (rectas o curvas) representadas en coordenadas con escala aritmética.*—En este diagrama, las divisio-

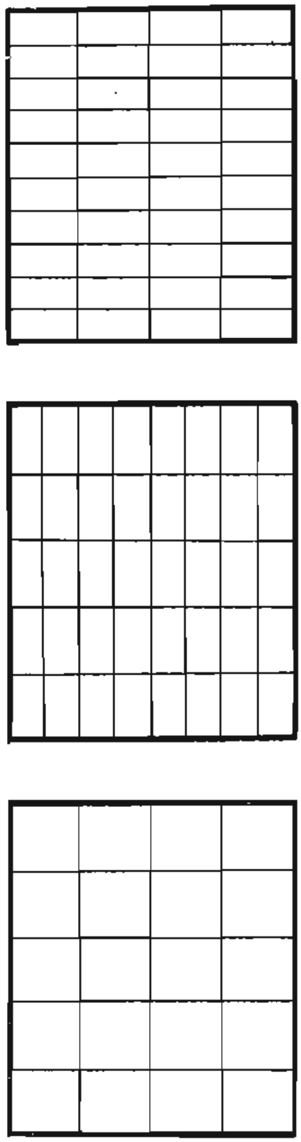

FIGURA 13— Coordenadas rectangulares con graduaciones que corresponden a escala aritmética.

nes que marcan la escala, tanto en la ordenada como en la abcisa, deben ser colocadas a distancias iguales. Esto no significa que la distancia usada en la ordenada debe ser igual a la usada en la abcisa, pero si es necesario la uniformidad en cada una de las coordenadas consideradas separadamente. La Figura 13 demuestra gráficamente este criterio.

En el diagrama pueden estar representadas una o varias líneas, rectas o curvas. Diversos ejemplos ilustrarán algunos de los usos más frecuentes de este diagrama.

Ejemplo: En 338 sujetos indígenas, cuya edad variaba entre 4 y 19 años, se ha medido la estatura. Las observaciones fueron hechas en Morococha y los resultados obtenidos están dados en el Cuadro 23.

La curva de la Figura 14, construída con los datos del Cuadro 23, representa gráficamente la relación entre estatura y edad en el grupo de sujetos estudiados.

CUADRO 23

ESTATURA EN 338 SUJETOS INDIGENAS (EDAD: 4—19 AÑOS) ESTUDIADOS EN MOROCOCHA.

Edad (años)	*Número de sujetos*	*Valor medio de la estatura correspondiente a cada año de edad (centímetros)*
4	2	96.5
5	4	106.7
6	3	110.3
7	16	112.6
8	33	116.4
9	14	118.9
10	21	122.2
11	12	127.9
12	22	132.3
13	8	137.2
14	8	142.7
15	3	147.3
16	13	154.8
17	21	153.1
18	64	156.6
19	94	157.8

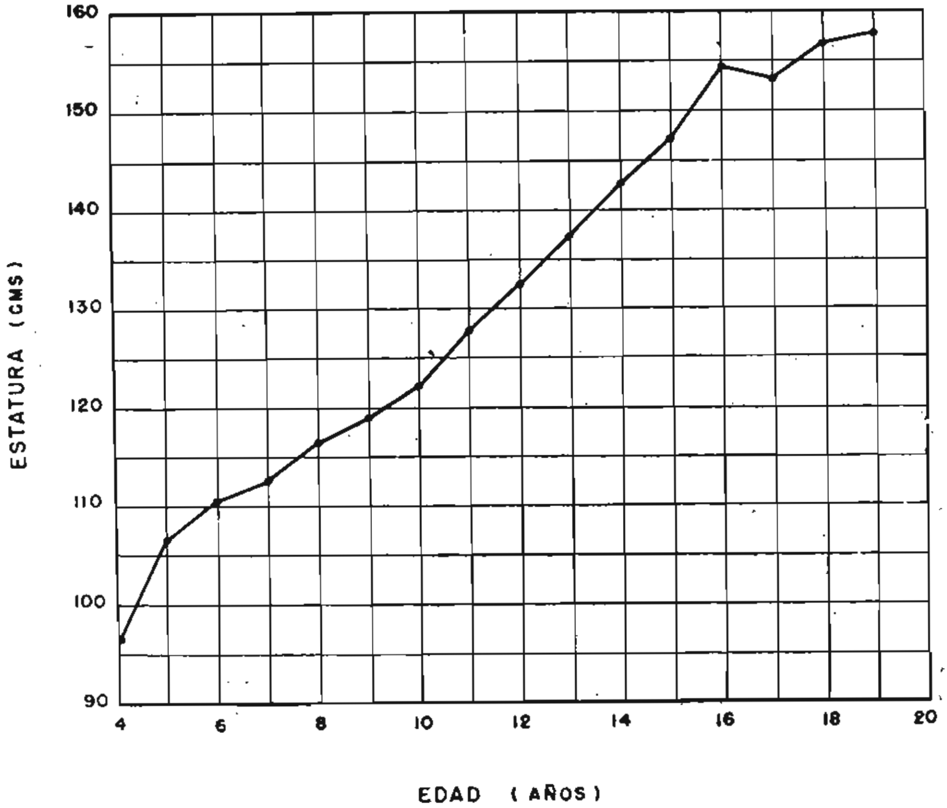

FIGURA 14— Valores medios de estatura correspondientes a diferentes edades (entre 4 19 años). Observaciones hechas en 338 sujetos de raza india, residentes en Morococha.
(Diagrama construído con los datos del Cuadro 23).

Ejemplo: En investigaciones hechas en este Departamento se ha determinado la cantidad de hemoglobina en la sangre de sujetos residentes al nivel del mar y a diversas alturas.

A los resultados obtenidos se ha agregado algunos otros, tomados de la literatura y relacionados con previos estudios. El Cuadro 24 contiene los valores promedios correspondientes a los diversos grupos de sujetos estudiados.

La curva de la Figura 15 representa gráficamente la relación entre el nivel de altura y la concentración de hemoglobina circulante.

CUADRO 24

CANTIDAD DE HEMOGLOBINA EN LA SANGRE DE SUJETOS RESIDENTES AL NIVEL DEL MAR Y A DIVERSAS ALTURAS

Altura (metros sobre el nivel del mar)	*Valores promedios de Hemoglobina (gramos por 100 cc.)*
0	16.00
1,520	16.54
2,390	16.81
3,140	17.90
3,730	18.82
4,540	20.76
4,860	21.66
5,340	22.86

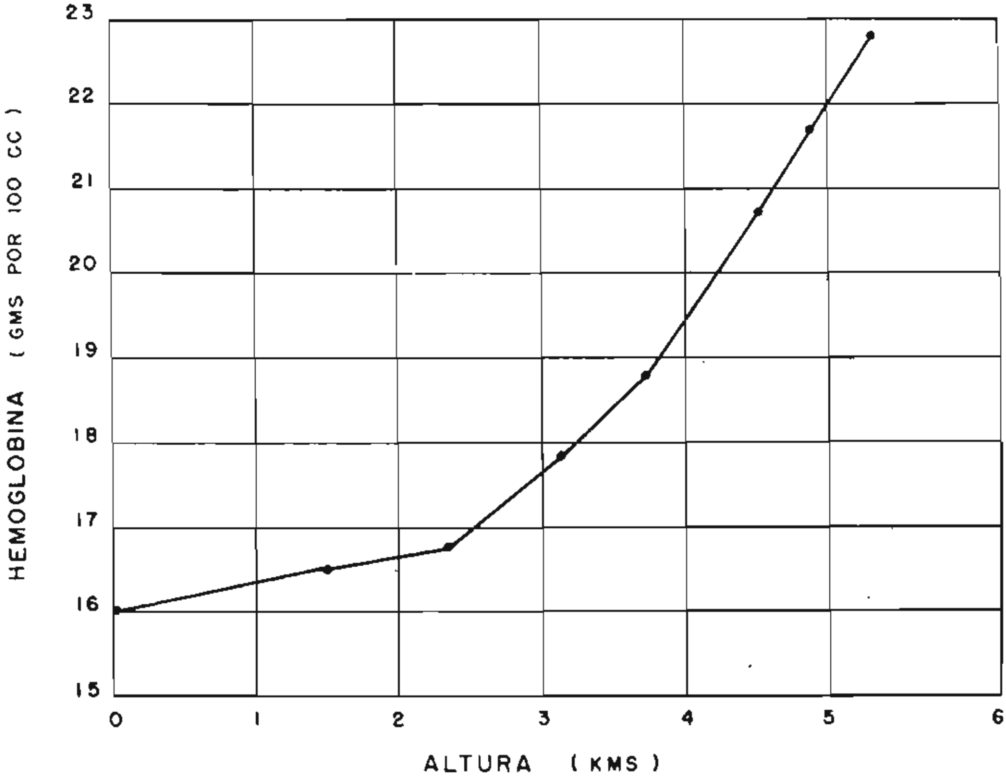

FIGURA 15— Valores medios de Hemoglobina (gramos por 100 cc. de sangre) en hombres adultos sanos, residentes a diversas alturas. (Diagrama construído con los datos del Cuadro 24).

Ejemplo *: En varios grupos de sujetos sanos, de diferente edad, se ha determinado la capacidad total, la capacidad vital y el aire residual pulmonar.

Los valores promedios obtenidos para las diferentes edades están incluídos en el Cuadro 25.

La curva de la Figura 16 representa gráficamente la relación entre las varias capacidades pulmonares y la edad de los sujetos estudiados.

CUADRO 25

DETERMINACIONES DE CAPACIDAD TOTAL, CAPACIDAD VITAL Y AIRE RESIDUAL PULMONAR EN SUJETOS SANOS DE DIFERENTE EDAD

Grupos	*Edad de los grupos (años)*	*Capacidad total (litros)*	*Capacidad vital (litros)*	*Aire residual (litros)*
	Valores promedios			
1	6.0	1.65	1.26	0.39
2	10.5	2.78	2.20	0.57
3	14.1	4.67	3.71	0.95
4	17.4	6.25	4.95	1.30
5	24.5	6.81	5.25	1.66
6	35.1	6.36	4.76	1.60
7	44.6	5.77	4.28	1.48
8	51.8	6.31	4.16	1.81
9	63.0	5.77	4.05	1.72
10	75.0	5.12	3.20	1.92

* Tomado de S. Robinson — Arbeitsphysiol., 1938, 10: 18.

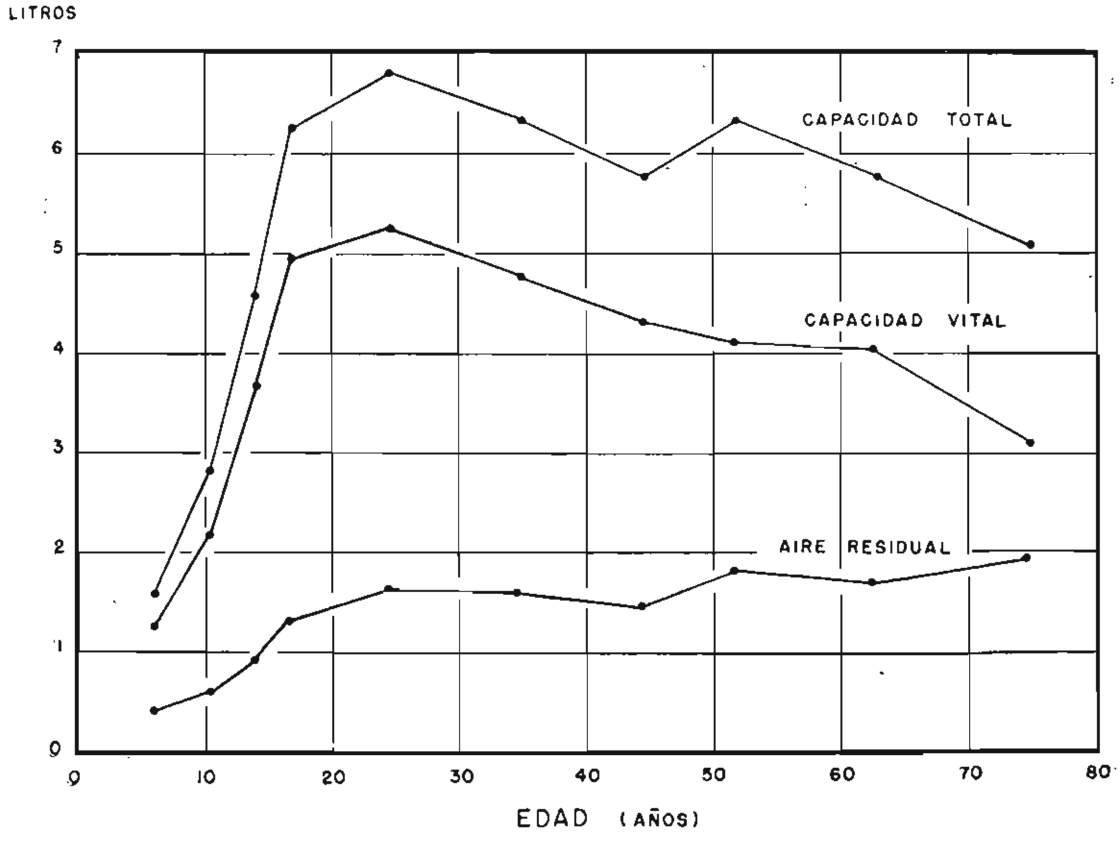

FIGURA 16— Valores medios de capacidad total, capacidad vital y aire residual pulmonar en diferentes edades. Observaciones hechas en hombres. (Diagrama construído con los datos del Cuadro 25).

Ejemplo *: Se ha determinado el consumo de oxígeno, en condiciones básisicas, en varios grupos de sujetos de diferente edad.

Los valores medios y las variaciones extremas obtenidas para las diferentes edades están dados en el Cuadro 26.

La Figura 17, construída con los datos del Cuadro 26, representa gráficamente los valores promedios y los límites de variación del consumo de oxígeno correspondientes a cada uno de los grupos de sujetos estudiados, y expresa, por lo tanto, la relación entre edad y consumo de oxígeno.

CUADRO 26

CONSUMO BASICO DE OXIGENO EN SUJETOS SANOS DE DIFERENTE EDAD.

Grupos	*Edad de los grupos (años) Media*	*Consumo de Oxigeno (en cc. por minuto y por m². de superficie corporal)*	
		Media	*Valores extremos*
1	6.0	184	167 — 199
2	10.5	171	131 — 194
3	14.1	157	140 — 175
4	17.4	151	134 — 170
5	24.9	136	123 — 146
6	35.1	127	115 — 139
7	44.5	129	112 — 141
8	52.1	127	112 — 142
9	63.1	124	103 — 144
10	75.0	114	93 — 136

* Tomado de S. Robinson — Arbeitsphysiol., 1938, 10: 18.

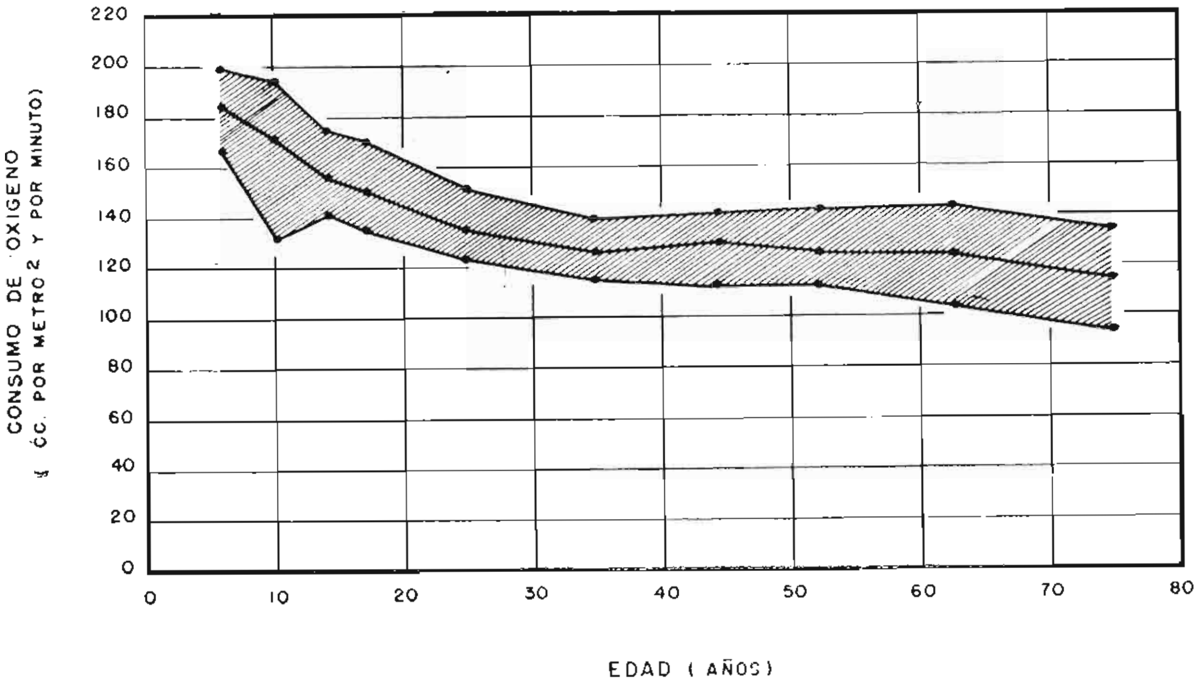

FIGURA 17— Consumo de oxígeno, en condiciones básicas, en diferentes edades. Observaciones hechas en hombres.
Las curvas superior e inferior indican los límites extremos de variación; la curva del centro corresponde a los valores medios.

(Diagrama construído con los datos del Cuadro 26).

El diagrama con escala aritmética en ambas coordenadas es también usado para representar una línea que corresponda a la relación entre cantidades expresadas por sus logaritmos.

Incidentalmente, mencionaremos que curvas de desarrollo complicado pueden transformarse en una línea casi recta cuando son representadas por sus logaritmos en un diagrama con escala aritmética. El siguiente ejemplo ilustra este aspecto.

Ejemplo *: Cuando la sangre humana es expuesta, en recipientes especiales y a una temperatura dada, a diferentes concentraciones de oxígeno, su hemoglobina se satura en relación con la tensión ejercida por dicho gas.

En el Cuadro 27 están dadas, en columnas 1 y 2, las diferentes tensiones de oxígeno (p02 mmHg) que corresponden a los diferentes grados de saturación de la hemoglobina con oxígeno (Hb02%). En el mismo cuadro, en columnas 3 y 4, están dados los logaritmos de cada una de las cifras de las columnas 1 y 2 (estos logaritmos son obtenidos consultando una tabla de logaritmos. Ver Apéndice A).

La curva de la Figura 18 representa la relación entre p02 y Hb02% y está construída con los datos de las columnas 1 y 2 del Cuadro 27.

La línea recta de la Figura 19 representa la misma relación que la expresada por la curva de la Figura 18, y está construída con los datos de las columnas 3 y 4 del Cuadro 27.

Se puede apreciar que la línea curva de la Figura 18 es transformada en una línea recta cuando es representada por medio de los legoritmos de los valores a que corresponde.

CUADRO 27

CURVA DE DISOCIACION DE LA OXIHEMOGLOBINA. SANGRE HUMANA A pHs = 7.40

1	2	3	4
Hb02 %	p02 (mmHg)	log. Hb02 %	log. p02
10	8.0	2.954	0.905
15	10.8	2.754	1.032
20	13.2	2.602	1.120
30	17.6	2.367	1.245
40	21.6	2.176	1.335
50	25.9	2.000	1.412
60	30.5	1.824	1.485
70	33.5	1.632	1.550
80	44.9	1.398	1.652
85	50.8	1.247	1.706
90	60.3	1.046	1.780
94	73.6	0.805	1.867

* Tomado de D. B. Dill, H. T. Edwards, M. Florkin and R. W. Campbell — J. Biol. Chem., 1932, 45: 143.

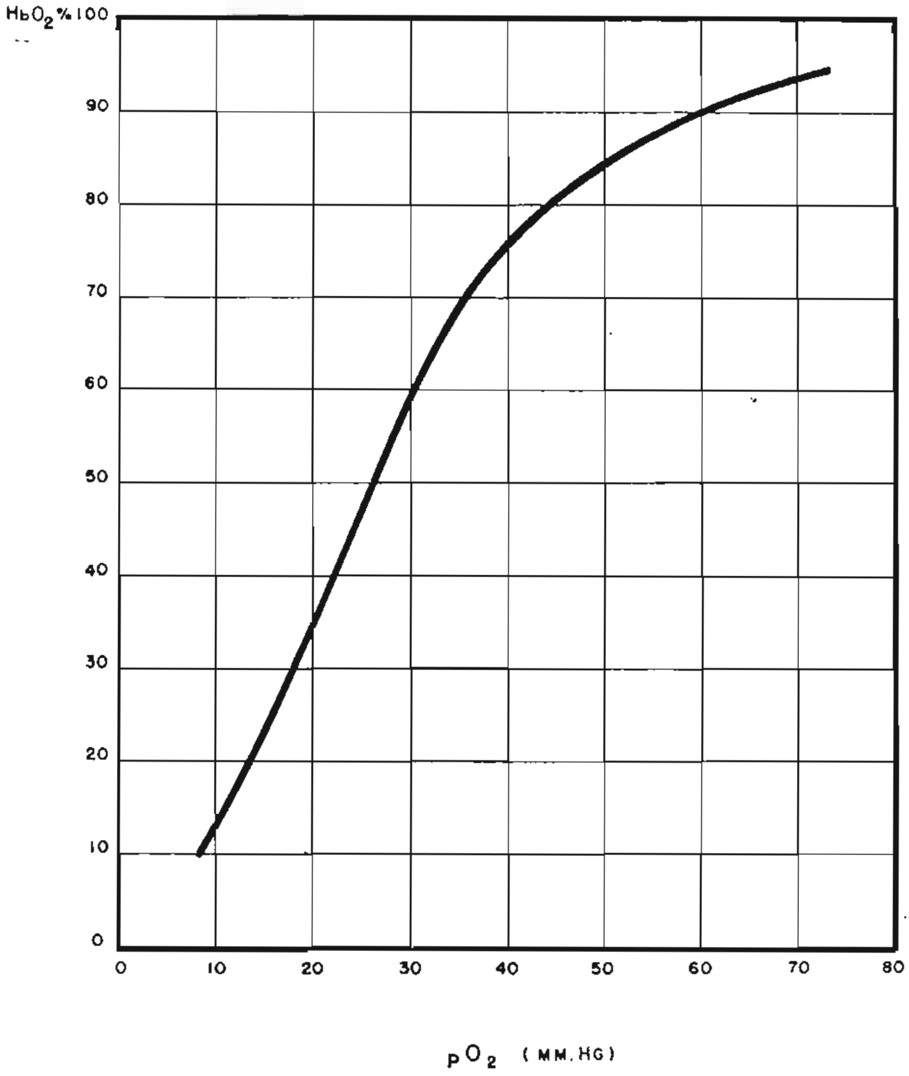

FIGURA 18— Curva de disociación de oxígeno en sangre humana a pHs = 7.40. (Diagrama construído con los datos de las Columnas 1 y 2 del Cuadro 27).

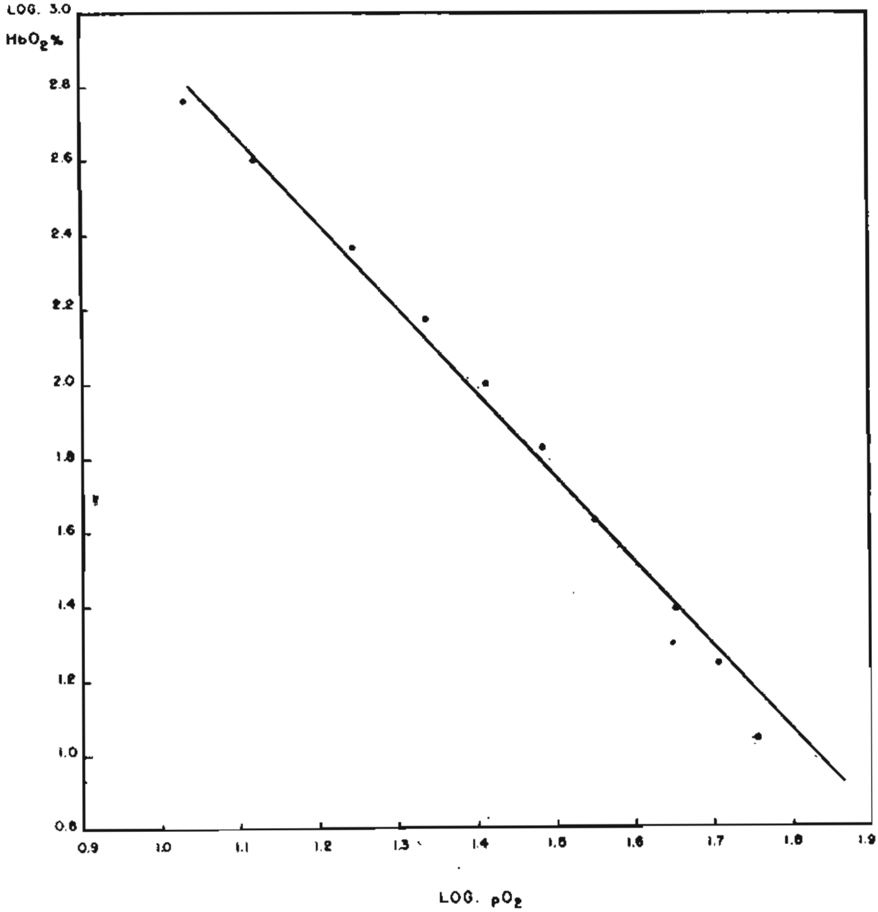

FIGURA 19— Curva de disociación de oxígeno en sangre humana a pHs = 7.40. Esta curva corresponde exactamente a la representada en la Figura 18, pero está construída en relación a los logaritmos de % Hb02 y p0². *La línea recta de este diagrama puede ser convertida en la línea curva de la Figura 18 hallando los antilogaritmos correspondientes a los logaritmos de ambas escalas (ordenada y abcisa).* *(Diagrama construído con los datos de las Columnas 2 y 3 del Cuadro 27).*

Hemos visto, en los diferentes ejemplos utilizados, que en esta clase de diagramas, la graduación aritmética en ambas coordenadas corresponde a cifras o períodos de tiempo cuyo valor aumenta en partes iguales o proporcionales, de abajo a arriba en la ordenada y de izquierda a derecha en la abcisa. Pero, en ciertos casos, puede ser conveniente, para evitar dimensiones exageradas del diagrama y abreviar la escala en lo que se refiere a la representación gráfica de aquellos datos que no varían por un largo período de tiempo, o que no interesan desde el punto de vista del objetivo del diagrama. El siguiente ejemplo demuestra el procedimiento conveniente a seguir en estas circunstancias:

Ejemplo: Se desea representar gráficamente las variaciones en el número de hematíes, reticulocitos y la cantidad de hemoglobina, en un caso de Anemia Perniciosa a quien se le ha administrado extracto hepático. El período de terapia, durante el cual han ocurrido las variaciones hemáticas que se desea representar gráficamente, ha sido de 15 días, pero este período ha sido precedido y seguido de un tiempo de observación de 12 y 15 días, respectivamente.

Para evitar que el diagrama tenga dimensiones exageradas, o incluya observaciones que no tienen interés especial, el gráfico se ha construído —ver la Figura 20— de tal manera que los períodos previo y posterior a aquel correspondiente a la terapia hepática, estén representados por las observaciones correspondientes al primero y último día de ambos períodos. Para esto se rompe la escala en la abcisa, interrumpiendo las graduaciones, tal como está indicado en la Figura 20.

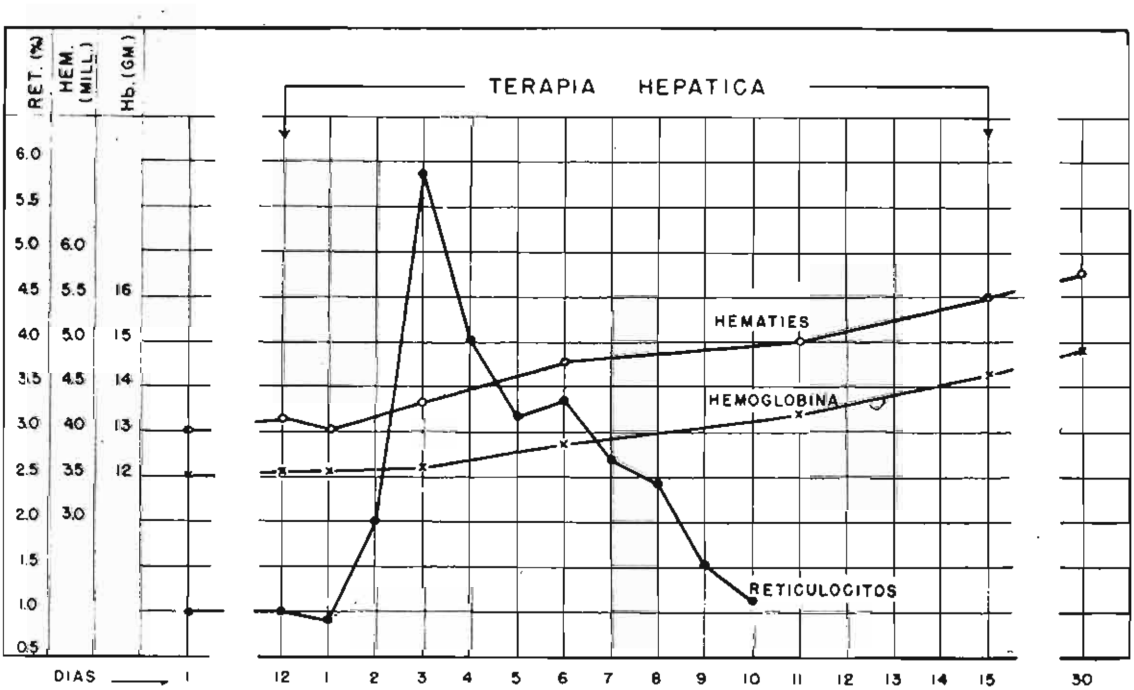

FIGURA 20— Variaciones en el número de hematíes, reticulocitos y la cantidad de hemoglobina en un caso de Anemia Perniciosa tratado con extracto hepático.
Los datos correspondientes a los períodos de observación, antes y después de la terapia hepática han sido representados por lo hallado en el primero y último día de ambos períodos.

En otros casos, en los datos que se desea representar gráficamente, puede haber uno o dos valores que se apartan considerablemente de los correspondientes al resto de las observaciones, por ser muy altos, o, por el contrario, muy bajos. Para evitar que la escala respectiva, en la ordenada o abcisa, se prolongue considerablemente, puede procederse en forma análoga a lo indicado en el ejemplo que acabamos de utilizar (Figura 20), es decir, se rompe la escala, interrumpiendo la graduación, de tal manera que aquella correspondiente al valor exageradamente alto o bajo, se encuentre separada del resto de la escala.

2—*Diagrama de líneas representadas en coordenadas con escala logarítmica o semi-logarítmica.*— En una sección anterior (pág. 218) hemos indicado que las cifras que corresponden a una línea recta o curva pueden ser representadas por sus correspondientes logaritmos en coordenadas con escala aritmética (Figura 19). La operación de encontrar los logaritmos, en las tablas que existen con ese objeto, demanda un tiempo considerable, y por este motivo es, a veces, más conveniente utilizar directamente las cifras (no los logaritmos), en diagramas cuyas coordenadas tengan divisiones que corresponden a una escala logarítmica. Esta clase de diagrama es también más fácilmente interpretado.

Un diagrama que tiene ambas coordenadas (ordenada y abcisa) con escala logarítmica se denomina *diagrama logarítmico;* si solo una de las coordenadas (y esta es la ordenada generalmente) tiene escala logarítmica, y la otra escala aritmética, el diagrama se denomina *diagrama semi-logarítmico.* Este último tipo de gráfico es el que habitualmente se emplea en publicaciones médicas.

La escala logarítmica consta de una o de varias divisiones principales (ver Figura 21), subdivididas en 9 espacios, cuyas dimensiones desiguales coresponden a los logaritmos de las cifras usadas en la escala. Puede adquirirse en el mercado papel que tiene impresa la escala logarítmica; en caso de no ser posible su adquisición es fácil construír una escala logarítmica en papel milimetrado. La manera de proceder es la siguiente:

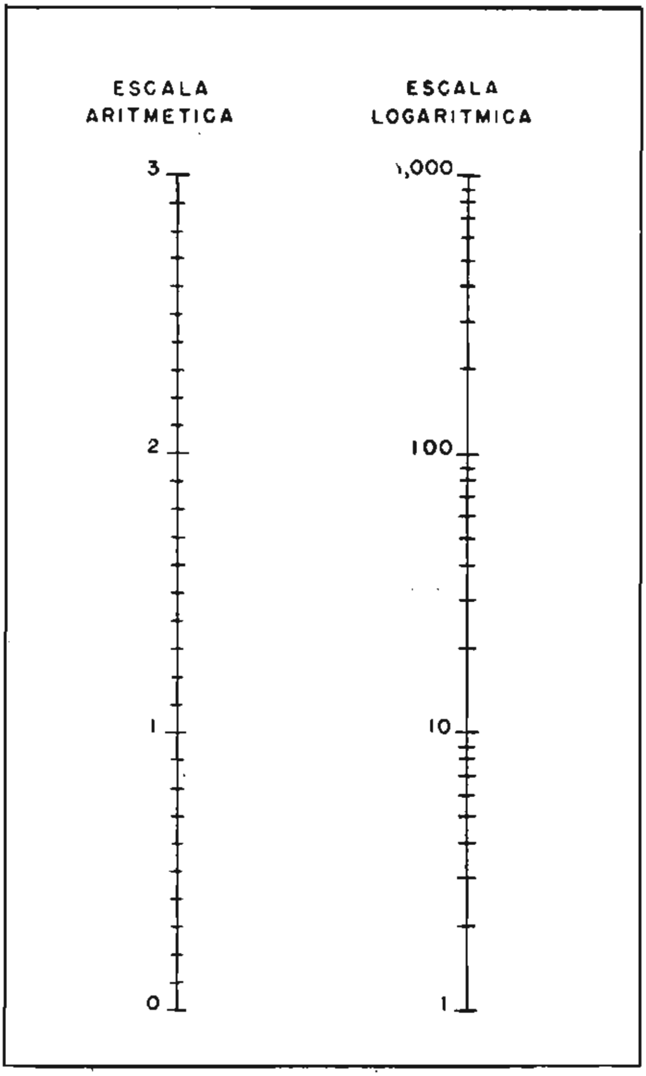

FIGURA 21— Comparación de una escala aritmética con una escala logarítmica. Cada división principal en la escala logarítmica, cuya graduación debe corresponder a una potencia de 10 (véase el texto), está subdividida en 9 espacios desiguales.

El diagrama es construído en papel milimetrado. Se elige, arbitrariamente, el número de milímetros que corresponde a la *división principal* de la escala logarítmica; en seguida, para subdividividir esta división principal en 9 partes, se multiplica el número de milímetros elegido por cada uno de los logaritmos de 1, 2, 3, 4, 5, 6. 7, 8, 9 y 10. Los resultados obtenidos en estas operaciones indican el número de milímetros que debe corresponder a cada una de las subdivisiones.

El siguiente ejemplo ilustrará estos cálculos:

En un papel milimetrado, en el que se va a construir un diagrama semi-logarítmico, se ha elegido, arbitrariamente, 100 milímetros como la altura total de la ordenada que debe llevar una escala logarítmica, compuesta de una división principal subdividida en 9 partes.
El cálculo es el siguiente:

			(1) Logaritmos		(2) Milímetros		(3)
De	1	=	0.0000	X	100	=	0
	2	=	0.3010	X	100	=	30.1
	3	=	0.4771	X	100	=	47.7
	4	=	0.6021	X	100	=	60.2
	5	=	0.6990	X	100	=	69.9
	6	=	0.7782	X	100	=	77.8
	7	=	0.8451	X	100	=	84.5
	8	=	0.9031	X	100	=	90.3
	9	=	0.9542	X	100	=	95.4
	10	=	1.0000	X	100	=	100.0

Las cifras de la columna (3) indican la altura en milímemetros (contando de abajo a arriba si se trata de la ordenada como en este caso, y de izquierda a derecha si se trata de la abcisa) a que se debe colocar las diferentes subdivisiones.
La Figura 22 representa gráficamente esta construcción de escala logarítmica en una ordenada cuya altura es de 100 milímetros.

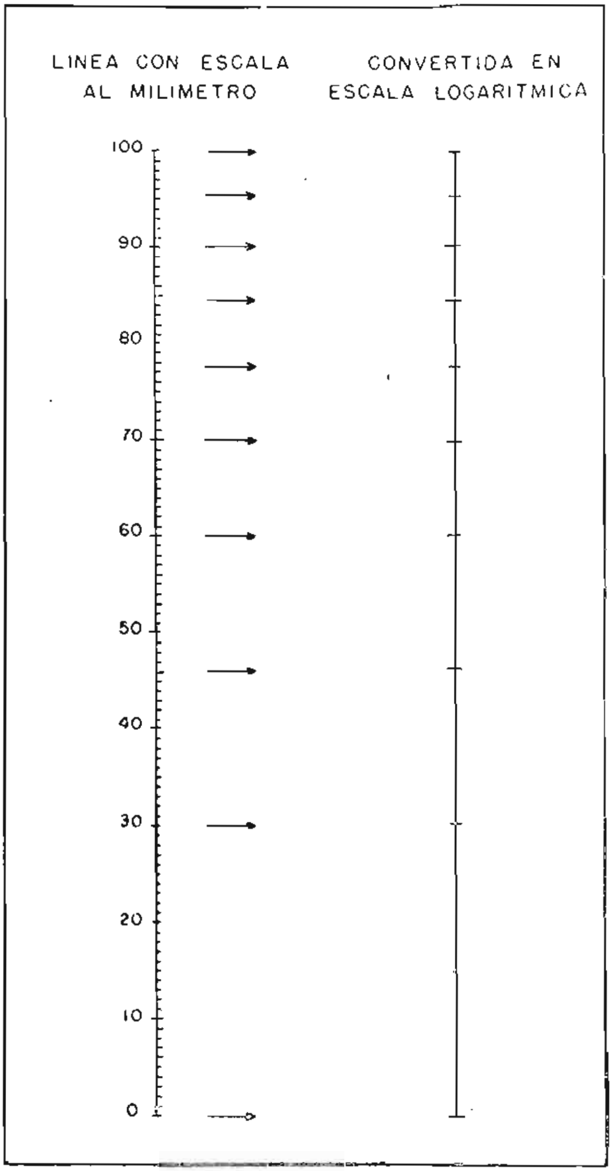

FIGURA 22— Construcción de una escala logarítmica en papel milimetrado. En este ejemplo, la división principal de la escala logarítmica tiene una longitud de 100 milímetros en el diagrama original; las subdivisiones se han hecho de acuerdo con los cálculos descritos en el texto.
(Este diagrama corresponde a la construcción de una escala logarítmica con una división principal. Si la naturaleza de los datos, por representar en el diagrama, requiere una escala logarítmica con varias divisiones principales, se procederá de idéntica manera con respecto al cálculo de cada una de ellas).

La escala logarítmica no tiene un valor 0 y su graduación corresponde a potencias de 10. Principiando de abajo a arriba, si se trata de la ordenada, y de izquierda a derecha en el caso de la abcisa, las *divisiones principales* (no las subdivisiones) deben corresponder a 0.001, 0.01, 0.1, 1, 10, 100, 1,000, 10,000 100,000, 1,000,000, 10,000,000, etc., pudiéndose iniciar la graduación con cualquiera de estas valores, según convenga al objeto del gráfico.

El diagrama logarítmico, o semi-logarítmico, es útil para representar variaciones que difieren grandemente en su valor numérico. Por ejemplo, sería casi imposible representar en un diagrama de coordenadas con escala aritmética una línea curva que en sus valores bajos corresponda a 10 o 100 y en sus valores más altos a 1,000,000 o más.

En esta clase de diagrama, una desviación o inclinación igual en dos curvas, medidas con la misma escala logarítmica, indica una variación proporcional en ambas. Esta característica permite apreciar comparativamente, y en forma gráfica, las variaciones tanto absolutas como relativas, de dos o más curvas representadas en el diagrama. Tal apreciación comparativa es particularmente valiosa en la representación gráfica de coeficientes de mortalidad, morbosidad, etc., calculados en diferentes grupos de población.

Los siguientes ejemplos ilustran el empleo de diagramas semi-logarítmicos:

Ejemplo: En un caso (hipotético) de Leucemia Mieloide ha sido posible determinar el número de leucocitos (por mm^3 de sangre) desde antes del inicio de la enfermedad hasta el fallecimiento, ocurrido meses después.

Los resultados obtenidos están dados en el Cuadro 28.

La curva de la Figura 23, construída con los datos del Cuadro 28, representa gráficamente las fluctuaciones en la numeración de leucocitos durante el período de observación.

CUADRO 28

NUMERO DE LEUCOCITOS EN UN CASO DE LEUCEMIA MIELOIDE OBSERVADO DURANTE UN PERIODO DE 20 MESES

Meses	Leucocitos (por mm^3 de sangre)
0	5,200
1	8,900
3	14,600
5	28,100
6	29,600
7	37,300
9	56,700
10	73,200
11	81,400
12	94,000
14	161,500
15	230,600
17	480,900
18	610,400
19	760,000
20	932,900

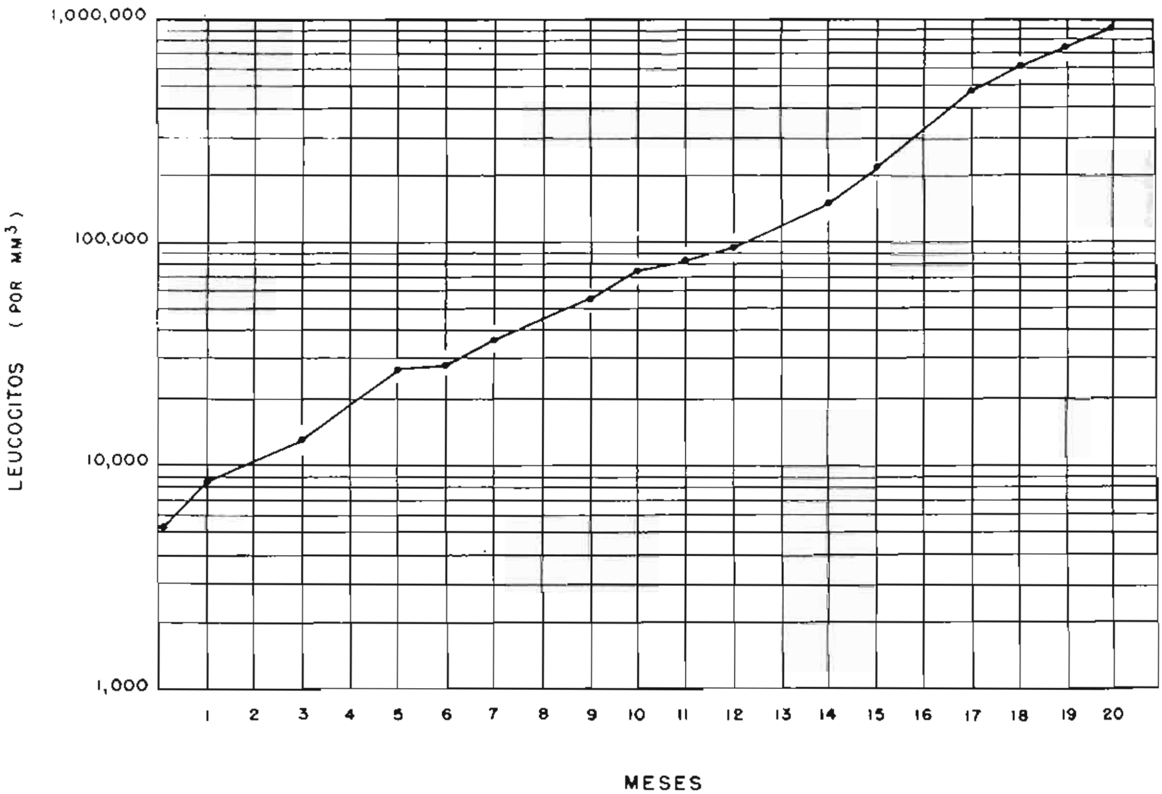

FIGURA 23— Leucocitos (por mm³ de sangre) en un caso de Leucemia Mieloide observado durante un período de varios meses. (Diagrama con escala semi-logarítmica construído con los datos del Cuadro 28).

Ejemplo *: El Cuadro 29 contiene el número de fallecimientos en Estados Unidos, por Tuberculosis y Fiebre Tifoidea, por cada 100,000 habitantes, durante el período 1900 a 1920.

La Figura 24, construída con los datos del Cuadro 29, representa gráficamente, en un diagrama de coordenadas con escala aritméti-

* Tomado de Introduction to Medical Biometry and Statistics. R. Pearl — W. B. Saunders Co., Philadelphia, 1940.

ca. las variaciones en los índices de mortalidad correspondientes a las enfermedades mencionadas.

Este diagrama da la impresión errónea de que la disminución en la mortalidad por Tuberculosis fué, durante aquel período de 20 años, proporcionalmente más marcada que la observada en Fiebre Tifoidea.

En cambio, la Figura 25, también construída con los datos del Cuadro 29 y que representa las mismas curvas de mortalidad en un diagrama semilogarítmico, permite apreciar, correctamente, que la declinación en mortalidad por Fiebre Tifoidea fué, proporcionalmente, más acentuada que la observada en relación a Tuberculosis.

El diagrama semi-logarítmico de la Figura 25 permite comparar variaciones absolutas y relativas en ambos coeficientes de mortalidad.

CUADRO 29

MORTALIDAD POR TUBERCULOSIS Y FIEBRE TIFOIDEA EN LOS ESTADOS UNIDOS.

AÑOS 1900 —— 1920

Año	*Número de fallecimientos por cada 100,000 habitantes*	
	Tuberculosis	*Fiebre Tifoidea*
1900	195.2	31.3
1901	189.8	27.5
1902	174.1	26.2
1903	177.1	24.6
1904	188.5	23.9
1905	180.9	22.4
1906	177.8	22.0
1907	175.6	20.5
1908	169.4	19.6
1909	163.3	17.2
1910	164.7	18.0
1911	159.0	15.3
1912	149.8	13.2
1913	148.7	12.6
1914	148.6	10.8
1915	146.7	9.2
1916	143.8	8.8
1917	147.1	8.1
1918	151.0	7.0
1919	124.9	4.8
1920	112.0	5.0

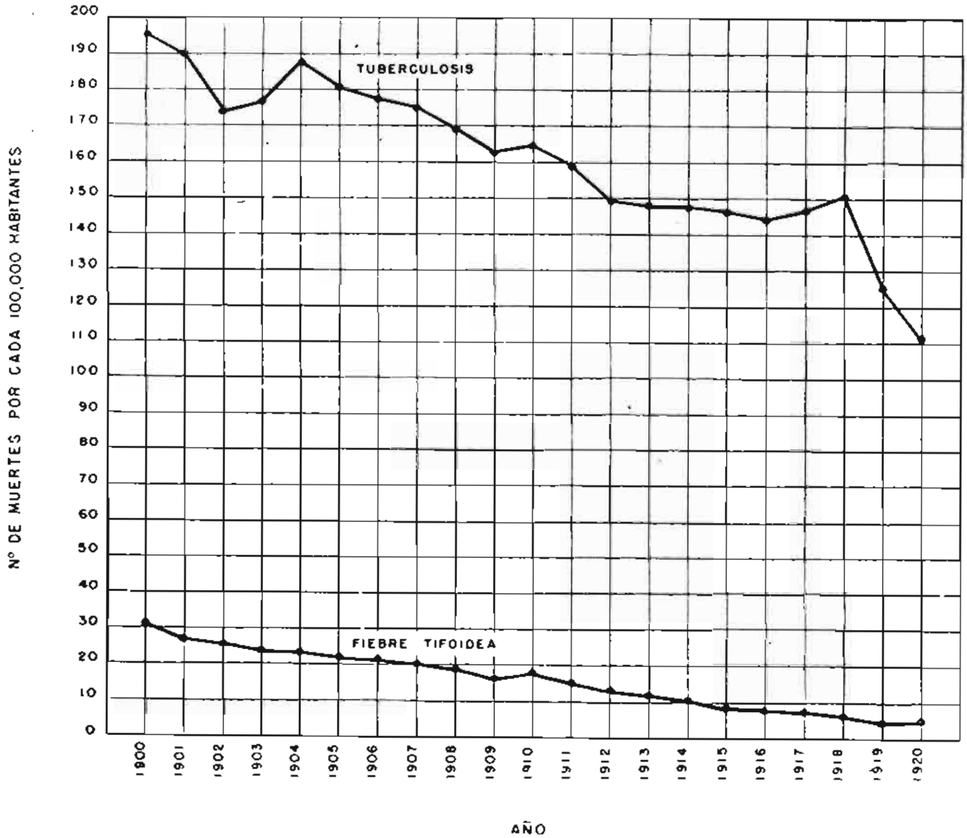

FIGURA 24 —Coeficiente de mortalidad, por cada 100,000 habitantes, correspondiente a Tuberculosis y Fiebre Tifoidea. Estados Unidos: años 1900-1920.
(Diagrama con escala aritmética construído con los datos del Cuadro 29).

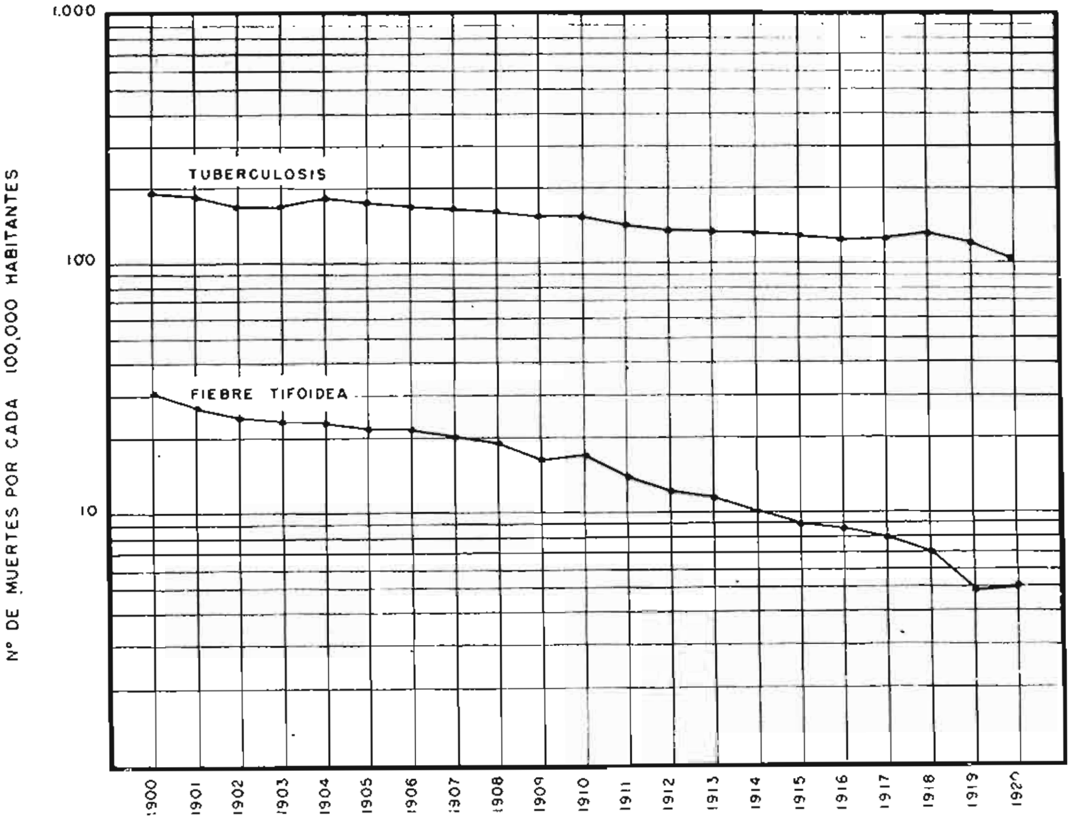

FIGURA 25 —Coeficiente de mortalidad, por cada 100.000 habitantes, correspondiente a Tuberculosis y Fiebre Tifoidea. Estados Unidos; años 1900-1920.
(Diagrama con escala semi-logarítmica construído con los datos del Cuadro 29).

Diagramas con escala semi-logarítmica se emplean en la calibración de colorímetros foto-eléctricos, instrumentos cuyo uso en los laboratorios se está generalizando rápidamente.

Para emplear un colorímetro foto-eléctrico en una determinación dada es preciso establecer, previamente, el factor o curva de calibración que corresponde a dicha determinación.

Ejemplo: Para emplear un colorímetro foto-eléctrico Evelyn en la determinación del hematocrito (hematíes por ciento) en muestras de sangre capilar, procedimiento que es posible efectuar midiendo su opacidad *, se ha obtenido una serie de lecturas en el galvanómetro del colorímetro lecturas correspondientes a muestras de sangre venosa en las que previamente se había determinado el hematocrito utilizando otro procedimiento.

Los resultados obtenidos en 25 de estas determinaciones, y que están dados en el Cuadro 30, han sido representados gráficamente en el diagrama semi-logarítmico de la Figura 26, y se ha construído, en seguida, una línea recta que corresponde a la variabilidad expresada por dichos puntos. Esta es la línea de calibración que debe usarse en la determinación del hematocrito por medio del colorímetro foto-eléctrico.

En posteriores observaciones, cuando se trata de determinar el hematocrito en una muestra de sangre, basta convertir la lectura del galvanómetro en porcentaje de hematíes (hematocrito) utilizando la curva de calibración de la Figura 26.

* Método descrito por A. T. Shohl and L. K. Diamond — Science, 1943, 98: 22.

CUADRO 30

LECTURAS EN EL GALVANOMETRO (COLORIMETRO FOTO-ELECTRICO DE EVELYN) CORRESPONDIENTES A VALORES CONOCIDOS DE HEMATOCRITO.

OBSERVACIONES EN 25 MUESTRAS DE SANGRE.

Lectura en el galvanómetro	Hematocrito (hematíes %)
60 3/4	42.0
83	13.9
83 3/4	12.8
80	17.5
80	17.9
76	21.8
76	21.0
75 2/4	23.0
71 1/4	26.5
70 2/4	27.5
68	31.0
64 2/4	35.0
65	37.0
61	37.7
63 1/4	40.0
60 3/4	40.2
57	45.0
59 2/4	49.4
55 3/4	50.1
54	51.2
51 2/4	59.1
48 2/4	60.7
47	69.0
45 1/4	69.3
41 1/4	78.3

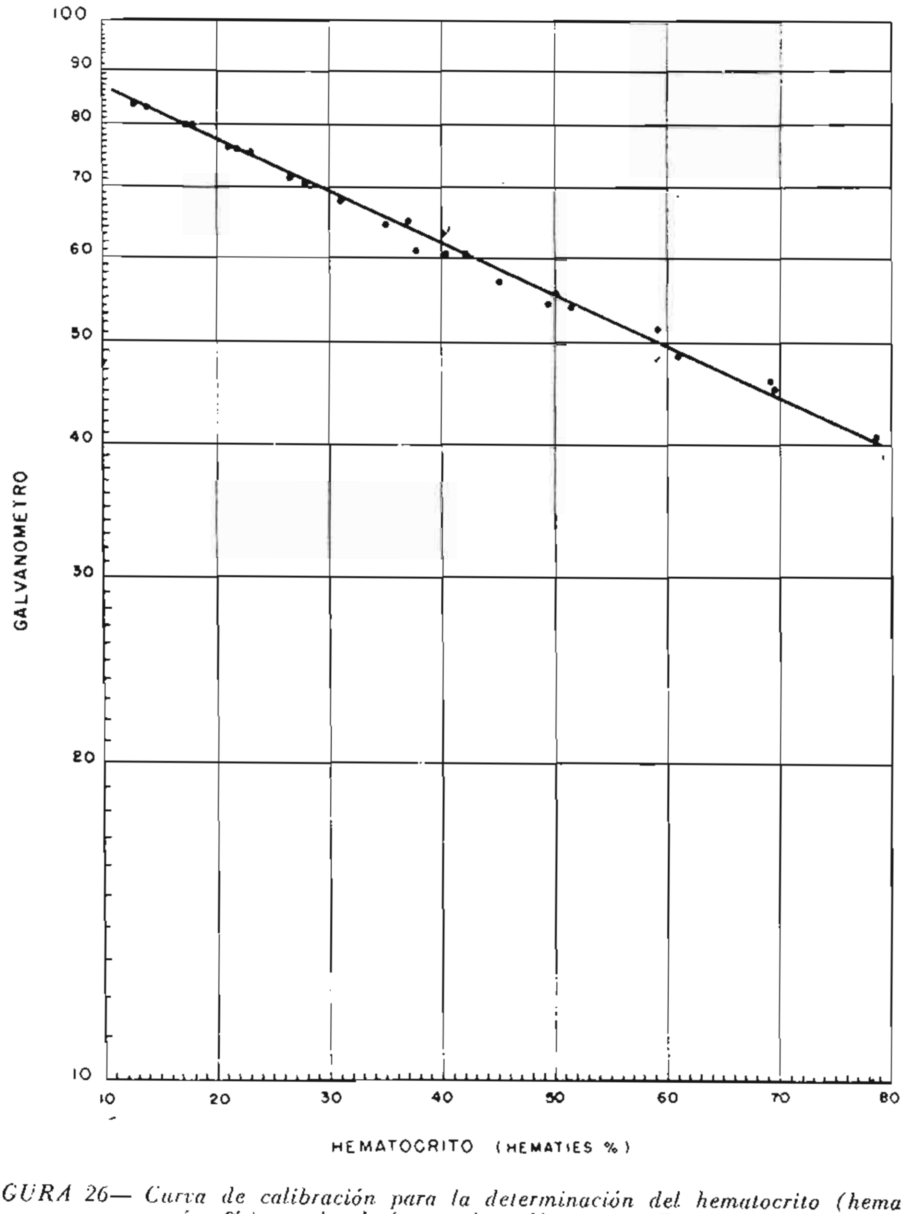

FIGURA 26— Curva de calibración para la determinación del hematocrito (hematíes %) en el colorímetro foto-eléctrico de Evelyn. (Diagrama con escala semi-logarítmica construido con los datos del Cuadro 30).

3—*Diagrama de dispersión de datos individuales.*—La relación existente entre dos variables se indica, en esta clase de diagrama, por medio de la representación gráfica de todas las observaciones individuales hechas. Si se desea, puede agregarse la línea recta o curva que representa la tendencia central o media de las observaciones.

Ejemplo: En 30 sujetos adultos sanos, estudiados en Oroya, se ha determinado el volumen total de sangre circulante. En todos los sujetos se ha determinado también el peso corporal.

Los resultados obtenidos están dados en el Cuadro 31.

En la Figura 27, construída con los datos del Cuadro 31, los puntos representan la relación entre el volumen de sangre y el peso en los 30 sujetos estudiados.

En la Figura 28, también construída con los datos del Cuadro 31, se ha agregado una línea recta que representa, en forma general, la relación entre volumen de sangre y peso corporal, y que ha sido construída tomando en cuenta la posición de todos los puntos individuales en el diagrama.

CUADRO 31

RELACION ENTRE EL VOLUMEN TOTAL DE SANGRE CIRCULANTE Y EL PESO CORPORAL

OBSERVACIONES HECHAS EN 30 SUJETOS ADULTOS SANOS RESIDENTES EN OROYA.

Sujetos	*Peso*	*Volumen de sangre*	*Sujeto*	*Peso*	*Volumen de sangre*
Nº	*(kilos)*	*(litros)*	*Nº*	*(kilos)*	*(litros)*
1	44.5	4.96	16	55.0	6.05
2	52.0	5.16	17	54.0	7.99
3	57.5	6.10	18	52.0	5.71
4	59.0	5.94	19	58.0	5.86
5	44.0	5.21	20	55.5	6.24
6	59.0	6.25	21	57.5	6.00
7	64.5	6.58	22	55.0	6.02
8	58.5	6.33	23	61.5	6.36
9	66.0	6.75	24	56.5	5.18
10	59.5	6.47	25	54.0	5.26
11	49.5	4.68	26	60.5	6.45
12	61.0	6.72	27	60.5	7.27
13	59.0	5.77	28	60.5	7.80
14	48.5	5.31	29	60.2	7.59
15	53.5	4.26	30	55.6	7.06

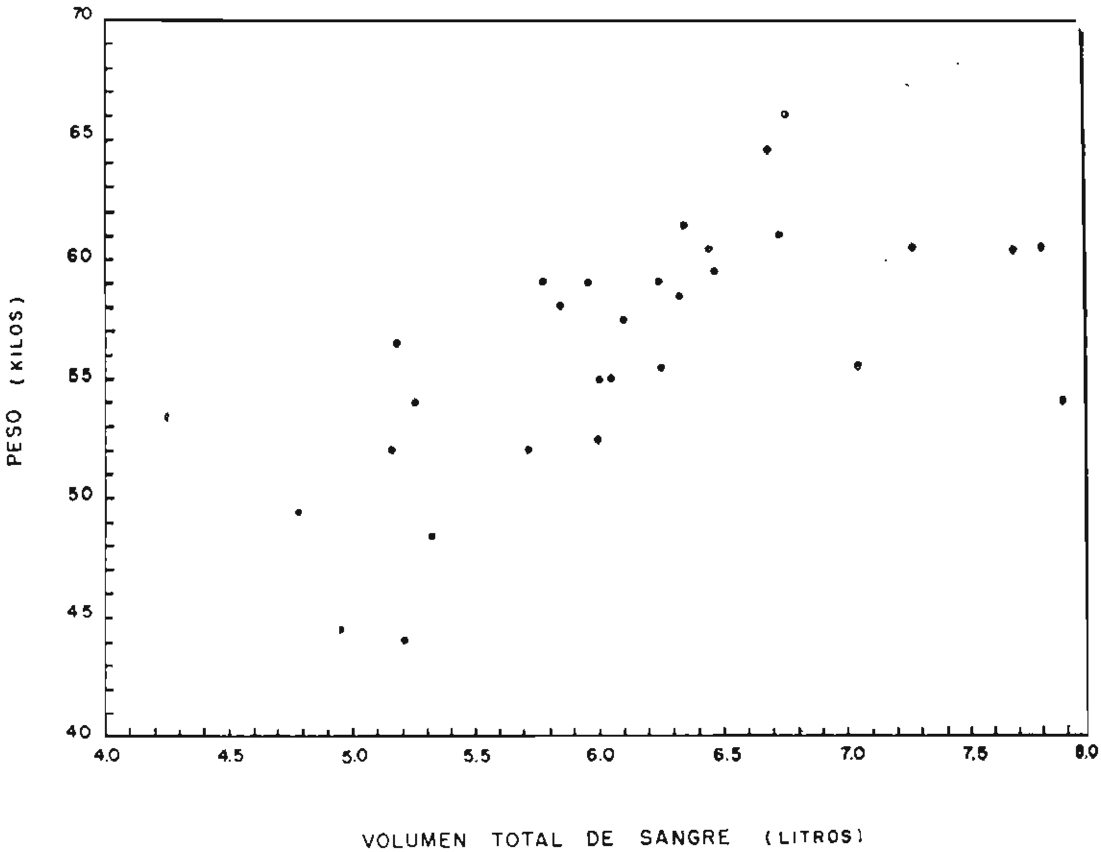

FIGURA 27— Volumen total de sangre circulante en relación al peso corporal. Observaciones hechas en 30 hombres adultos, sanos, residentes en Oroya. (Diagrama construido con los datos del Cuadro 31).

4—*Diagrama polar.*—Diagramas con coordenadas polares son a veces usado para representar gráficamente datos de orden cíclico o la relación entre varios factores, cuya magnitud no está totalmente expresada en forma cuantitativa.

Ejemplo *: En la Figura 29 están representados gráficamente los diversos factores que determinan la tolerancia y sintomatología durante el vuelo. Estos factores son: altura, velocidad de ascenso, ángulo de vuelo, aceleración, vibración, ruido, ventilación, olor, calor y frío. El límite entre las zonas "cómoda" e "incómoda" se le ha denominado: *psicológico*, y entre las zonas "incómoda" e "intolera-

* Tomado de Keeping Fit for flying. Pan American Airways System, New York, 1943.

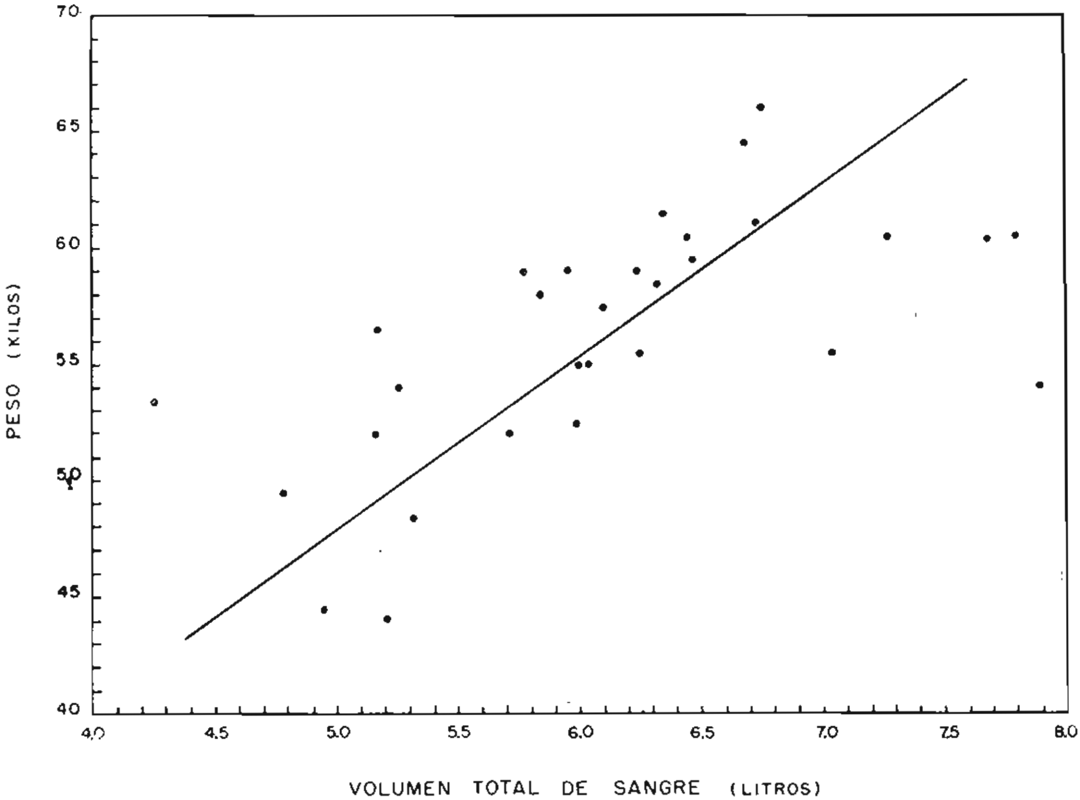

FIGURA 28— Volumen total de sangre circulante en relación al peso corporal. Observaciones hechas en 30 hombres adultos, sanos residentes en Oroya.
La línea recta, que representa la relación general entre volumen de sangre y peso corporal, ha sido construída tomando en cuenta la localización de los 30 puntos, cada uno de los cuales corresponde a una observación.
(Diagrama construído con los datos del Cuadro 31).

ble": *fisiológico*, para diferenciar el grado de severidad de los síntomas que corresponden a cada una de estas zonas.

Se supone que la tolerancia humana a los diversos factores que actúan sobre el organismo durante el vuelo puede ser casi perfecta hasta el límite psicológico, difícil hasta el fisiológico, e imposible más allá de este último límite.

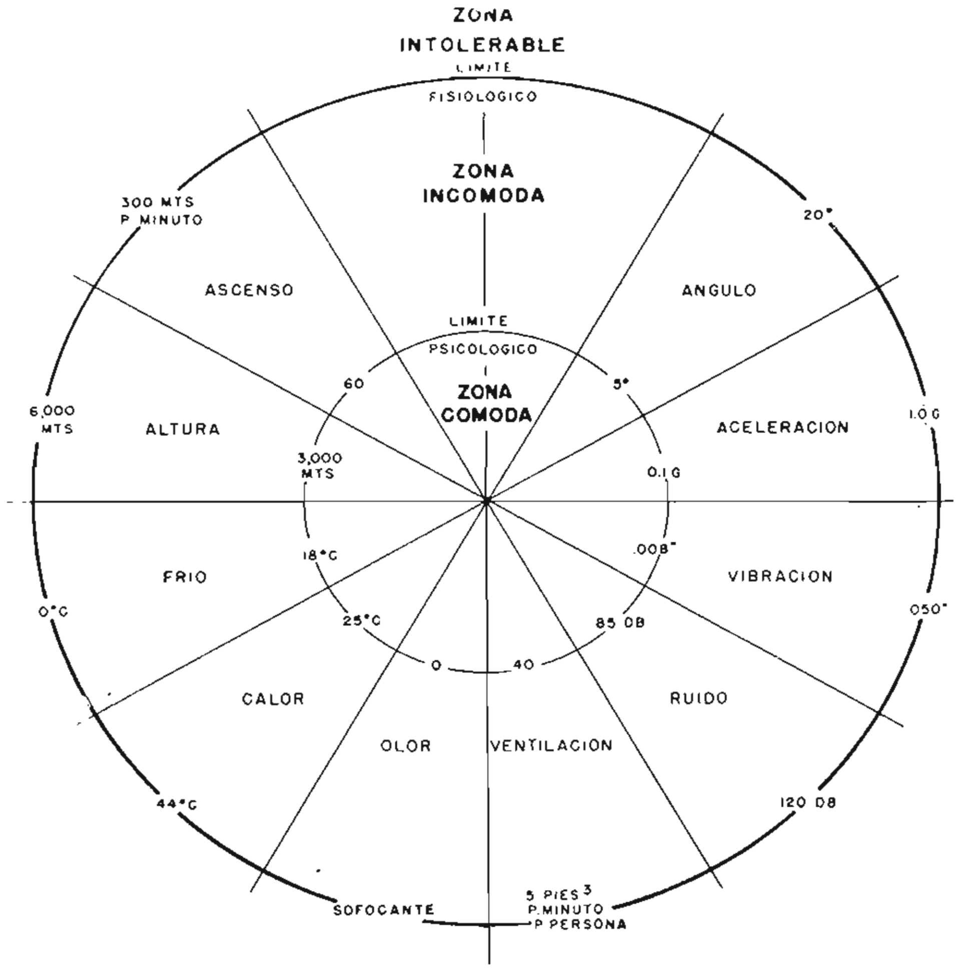

FIGURA 29— Factores que influencian la tolerancia y sintomatología durante el vuelo.

REPRESENTACION GRAFICA DEL COEFICIENTE DE VARIACION.

Como se ha indicado en el Capítulo I, el coeficiente de variación expresa, en porcentaje, el grado de variabilidad de una serie de datos analizados. Es una constante útil para comparar la variación hallada en dos o más series de datos que se refieren a unidades de medida diferentes. Su representación gráfica permite apreciar, visualmente, tal estudio comparativo.

CUADRO 32

OBSERVACIONES SOBRE LA EDAD, PESO Y ESTATURA DE 272 INDIVIDUOS.

Edad			Peso			Estatura		
Anos	Punto medio	Frecuencia	Libras	Punto medio	Frecuencia	Pulgadas	Punto medio	Frecuencia
20 — 21	21	9	99.5 — 109.4	104.5	2	59.5 — 60.4	60	1
22 — 23	23	12	109.5 — 119.4	114.5	12	60.5 — 61.4	61	3
24 — 25	25	34	119.5 — 129.4	124.5	20	61.5 — 62.4	62	0
26 — 27	27	41	129.5 — 139.4	134.5	43	62.5 — 63.4	63	2
28 — 29	29	35	139.5 — 149.4	144.5	54	63.5 — 64.4	64	12
30 — 31	31	44	149.5 — 159.4	154.5	51	64.5 — 65.4	65	19
32 — 33	33	31	159.5 — 169.4	164.5	39	65.5 — 66.4	66	29
34 — 35	35	24	169.5 — 179.4	174.5	26	66.5 — 67.4	67	40
36 — 37	37	15	179.5 — 189.4	184.5	16	67.5 — 68.4	68	50
38 — 39	39	12	189.5 — 199.4	194.5	6	68.5 — 69.4	69	35
40 — 41	41	10	199.5 — 209.4	204.5	3	69.5 — 70.4	70	35
42 — 43	43	3				70.5 — 71.4	71	25
44 — 45	45	1				71.5 — 72.4	72	12
46 — 47	47	0				72.5 — 73.4	73	5
48 — 49	49	1				73.5 — 74.4	74	4
Intervalo de grupo: 2			Intervalo de grupo: 10			Intervalo de grupo: 1		

Los procedimientos que se emplean para construir el diagrama que corresponda a los coeficientes de variación hallados en dos o más series de datos serán ilustrados con el siguiente ejemplo:

Ejemplo *: En 272 individuos se ha obtenido datos referentes a su edad, estatura y peso corporal.
Los resultados hallados divididos en grupos de frecuencia, a cada uno de los cuales se le ha calculado su punto medio, están incluídos en el Cuadro 32.

Se trata de representar gráficamente los coeficientes de variación que van a ser calculados para cada una de estas características, lo que permitirá apreciar, visualmente y en forma comparativa, la variabilidad correspondiente a la edad, peso y estatura de los sujetos estudiados.

(a)—Utilizando los procedimientos indicados en el Capítulo I se calculan las diversas constantes correspondientes a los datos de edad, peso y estatura incluídos en el Cuadro 32.
Los resultados obtenidos están dados en el Cuadro 33.

CUADRO 33

OBSERVACIONES SOBRE LA EDAD, PESO Y ESTATURA DE 272 INDIVIDUOS.

	Media $\pm$ *E.S.*	*Desviación standard* $\pm$ *E.S.*	*Coeficiente de variación (%)*
Edad (años)	30.59 $\pm$ 0.31	5.22 $\pm$ 0.22	17.1
Peso (libras)	151.56 1.22	19.95 0.86	13.2
Estatura (pulgadas)	68.13 0.15	2.45 0.10	3.6

(b)—Para representar gráficamente la variabilidad expresada por los coeficientes de variación: 17.1, 13.2 y 3.6 %, correspondientes a la edad, peso y estatura, respectivamente, de los sujetos estudiados, hay que efectuar algunas operaciones con cada una de las series de datos.
En primer lugar, con los datos referentes a la edad se construye el Cuadro 34.

* Tomado de Introduction to Medical Biometry and Statistics. R. Pearl — W. B. Saunders Co., Philadelphia, 1940.

CUADRO 34

EDAD (AÑOS)

1	2	3	4	5
Grupos	*Punto medio*	*Frecuencia*	*A*	*B*
20 — 21	21	9	68.6	1.4
22 — 23	23	12	75.2	1.8
24 — 25	25	34	81.7	5.2
26 — 27	27	41	88.3	6.3
28 — 29	29	35	94.8	5.4
30 — 31	31	44	101.3	6.7
32 — 33	33	31	107.9	4.7
34 — 35	35	24	114.4	3.7
36 — 37	37	15	121.0	2.3
38 — 39	39	12	127.5	1.8
40 — 41	41	10	134.0	1.5
42 — 43	43	3	140.6	0.5
44 — 45	45	1	147.1	0.2
46 — 47	47	0	153.6	0
48 — 49	49	1	160.2	0.2

Columnas 1, 2, 3—Son exactamente iguales a las del Cuadro 32, referentes a la edad.

Columna 4 (A)—Cada cifra en esta columna se obtiene dividiendo el punto medio de la Columna 2 entre el valor medio de la edad. El resultado es multiplicado por 100.
El valor medio de la edad de todos los sujetos es: 30.59 años (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{21}{30.59} \times 100 = 68.6$$

Para la segunda cifra:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{23}{30.59} \times 100 = 75.2$$

y así sucesivamente.

Columna 5 (B)—Cada cifra en esta columna se calcula aplicando la fórmula:

$$\frac{\text{Frecuencia} \times 1\,\%\ \text{de la Media}}{\text{Intervalo de grupo}}$$

La frecuencia es la cifra correspondiente en Columna 3.
El intervalo de grupo correspondiente a la edad es: 2 (Cuadro 32).
El valor medio de la edad es 30.59 años (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Frecuencia} \times 1\,\%\ \text{media}}{\text{Intervalo de grupo}} = \frac{9 \times 0.3059}{2} = 1.4$$

Para la segunda cifra:

$$\frac{\text{Frecuencia} \times 1\,\%\ \text{media}}{\text{Intervalo de grupo}} = \frac{12 \times 0.3059}{2} = 1.8$$

y así sucesivamente.

(c)—En seguida, con los datos referentes al peso, se construye el Cuadro 35.

CUADRO 35

PESO (LIBRAS)

1	2	3	4	5
Grupos	*Punto medio*	*Frecuencia*	*A*	*B*
99.5 — 109.4	104.5	2	68.9	0.3
109.5 — 119.4	114.5	12	75.5	1.8
119.5 — 129.4	124.5	20	82.1	3.0
129.5 — 139.4	134.5	43	88.7	6.5
139.5 — 149.4	144.5	54	95.3	8.2
149.5 — 159.4	154.5	51	101.9	7.7
159.5 — 169.4	164.5	39	108.5	5.9
169.5 — 179.4	174.5	26	115.1	3.9
179.5 — 189.4	184.5	16	121.7	2.4
189.5 — 199.4	194.5	6	128.3	0.9
199.5 — 209.4	204.5	3	134.9	0.5

Columna 1, 2 y 3—Son exactamente iguales a las del Cuadro 32, referentes al peso.

Columna 4 (A)—Cada cifra en esta columna se obtiene dividiendo el punto medio de la Columna 2 entre el valor medio del peso. El resultado es multiplicado por 100.

El valor medio del peso de todos los sujetos es: 151.56 libras (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{104.5}{151.56} \times 100 = 68.9$$

Para la segunda cifra:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{114.5}{151.56} \times 100 = 75.5$$

y así sucesivamente.

Columna 5 (B)—Cada cifra en esta columna se calcula aplicando la fórmula:

$$\frac{\text{Frecuencia} \times 1\% \text{ de la Media}}{\text{Intervalo de grupo}}$$

La frecuencia es la cifra correspondiente en Columna 3.
El intervalo de grupo correspondiente al peso es: 10 (Cuadro 32).
El valor medio del peso es: 151.56 libras (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Frecuencia} \times 1\% \text{ Media}}{\text{Intervalo de grupo}} = \frac{2 \times 1.5156}{10} = 0.3$$

Para la segunda cifra:

$$\frac{\text{Frecuencia} \times 1\% \text{ Media}}{\text{Intervalo de grupo}} = \frac{12 \times 1.5156}{10} = 1.8$$

y así sucesivamente.

(d)—En seguida, con los datos referentes a la estatura, se construye el Cuadro 36.

CUADRO 36

ESTATURA (PULGADAS)

1	2	3	4	5
Grupos	*Punto medio*	*Frecuencia*	*A*	*B*
59.5 — 60.4	60	1	88.1	0.7
60.5 — 61.4	61	3	89.5	2.0
61.5 — 62.4	62	0	91.0	0
62.5 — 63.4	63	2	92.5	1.4
63.5 — 64.4	64	12	93.9	8.2
64.5 — 65.4	65	19	95.4	12.9
65.5 — 66.4	66	29	96.9	19.8
66.5 — 67.4	67	40	98.3	27.3
67.5 — 68.4	68	50	99.8	34.1
68.5 — 69.4	69	35	101.3	23.8
69.5 — 70.4	70	35	102.7	23.8
70.5 — 71.4	71	25	104.2	17.0
71.5 — 72.4	72	12	105.7	8.2
72.5 — 73.4	73	5	107.1	3.4
73.5 — 74.4	74	4	108.6	2.7

Columnas 1, 2 y 3—Son exactamente iguales a las del Cuadro 32, referentes a la estatura.

Columna 4 (A)—Cada cifra en esta columna se obtiene dividiendo el punto medio de la Columna 2 entre el valor medio de la estatura. El resultado es multiplicado por 100.
El valor medio de la estatura de todos los sujetos es: 68.13 pulgadas (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{60}{68.13} \times 100 = 88.1$$

Para la segunda cifra:

$$\frac{\text{Punto medio}}{\text{Media}} \times 100 = \frac{61}{68.13} \times 100 = 89.5$$

y así sucesivamente.

Columna 5 (B)—Cada cifra en esta columna se calcula aplicando la fórmula:

$$\frac{\text{Frecuencia} \times 1\,\% \text{ de la Media}}{\text{Intervalo de grupo}}$$

La frecuencia es la cifra correspondiente en Columna 3.
El intervalo de grupo correspondiente a la estatura es: 1 (Cuadro 32).
El valor medio de la estatura es: 68.13 pulgadas (Cuadro 33).

Así, para la primera cifra tenemos:

$$\frac{\text{Frecuencia} \times 1\% \text{ Media}}{\text{Intervalo de grupo}} = \frac{1 \times 0.6813}{1} = 0.7$$

Para la segunda cifra:

$$\frac{\text{Frecuencia} \times 1\% \text{ Media}}{\text{Intervalo de grupo}} = \frac{3 \times 0.6813}{1} = 2.0$$

y así sucesivamente.

(e)—Finalmente, con los datos correspondientes a las Columnas 4 (A) y 5 (B), de los Cuadros 34, 35 y 36, se construye la Figura 30 que contiene las curvas correspondientes a la variabilidad hallada en la edad, peso y estatura de los sujetos estudiados. Cada curva es construída separadamente.

Los diferentes puntos, que más tarde se unen para construir las curvas, se anotan en el diagrama teniendo en cuenta que la escala en la abcisa corresponde a los valores de la Columna 4 (A), y la escala en la ordenada a los valores de la Columna 5 (B).

Las escalas en ambas coordenadas (ordenada y abcisa) deben ser lo suficientemente amplias para incluir los valores máximo y mínimo de todas las series de datos. En nuestro ejemplo, los valores de todas las Columnas 4 (A) varían entre 68.6 y 160.2, y, por consiguiente, la escala en la abcisa se extiende de 60 a 170; los valores en todas las Columnas 5 (B) varían entre 0 y 34.1 y, por consiguiente, la escala en la ordenada se extiende de 0 a 35.

En la Figura 30 puede apreciarse que el mayor grado de variabilidad corresponde a la edad de los sujetos y el menor grado a su estatura. La variabilidad en el peso ocupa una situación intermedia, aunque se aproxima más a la observada en la edad.

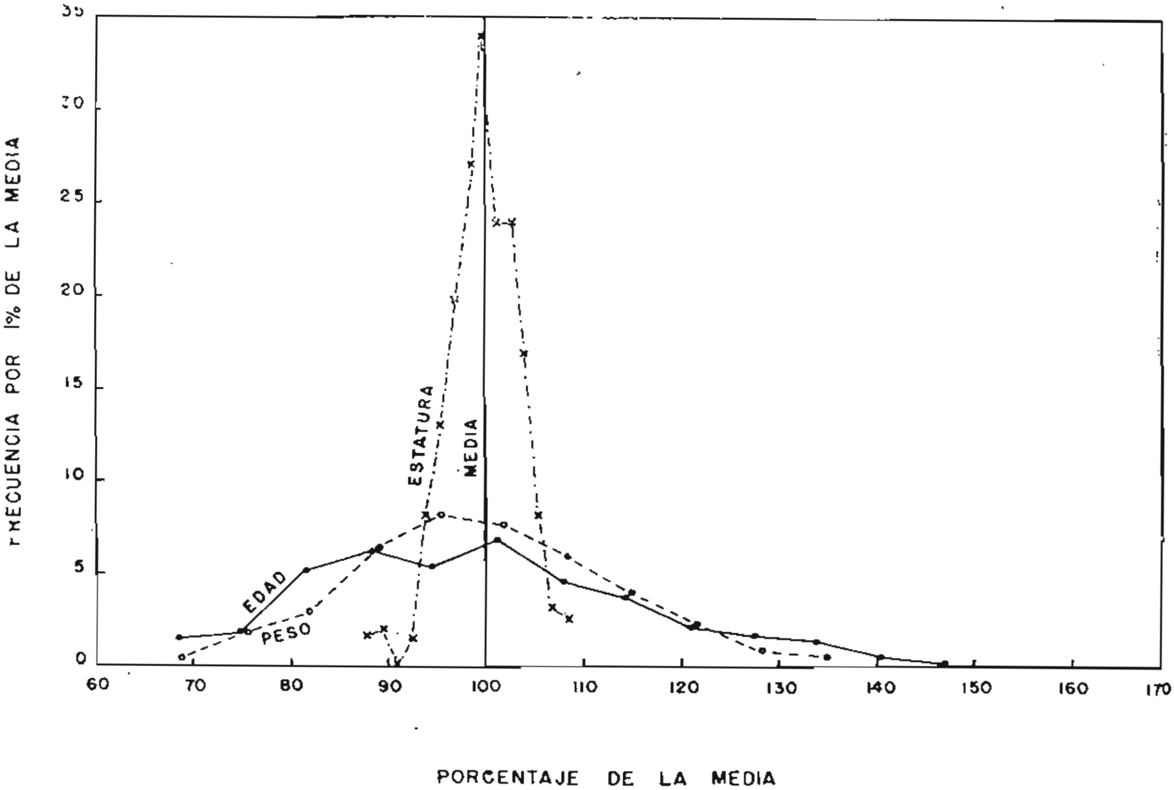

FIGURA 30— Representación gráfica de los coeficientes de variación correspondientes a la edad, estatura y peso, observados en un grupo de 272 hombres adultos.

REPRESENTACION GRAFICA DE LA ECUACION DE REGRESION.

LINEA DE REGRESION

En el Capítulo III hemos indicado que cuando la relación entre dos variables es lineal, y el coeficiente de correlación tiene un valor elevado, es posible derivar una ecuación (ecuación de regresión), por medio de la cual, conocida una de las variables es posible predecir que valor le corresponde en la otra.

Con la ecuación de regresión se puede construir gráficamente una línea recta, denominada *línea de regresión*, que represente la relación entre las dos variables. Ilustraremos con un ejemplo los procedimientos que se emplean con este objeto.

Ejemplo: — Utilizaremos el mismo ejemplo que ha servido para el cálculo del coeficiente de correlación (Capítulo III).

En dicho ejemplo encontramos que el coeficiente de correlación entre el Volumen globular y la Hemoglobina globular (determinaciones hechas en 106 muestras de sangre) fué: $+ 0.8758 \pm 0.0226$.

De este elevado coeficiente de correlación derivamos (página 175) una ecuación de regresión, por medio de la cual es posible calcular el valor de Hemoglobina globular que debe corresponder a un Volumen globular dado. Esta ecuación de regresión es:

Hemoglobina globular = (0.2161 X Volumen globular) + 9.2

En el diagrama de la Figura 31 la escala en la ordenada corresponde a los valores de Volumen globular y la escala en la abcisa a los valores de Hemoglobina globular. La graduación en estas escalas se hace tomando en cuenta los valores mínimo y máximo obtenidos en ambas determinaciones (cuyos valores originales están dados en el Cuadro 11, página ..).

Como la línea de regresión es una línea recta, para su construcción basta determinar dos puntos, uno correspondiente a un valor bajo y el otro a un valor alto, y en seguida, unir ambos puntos por medio de una línea.

Así en nuestro ejemplo, y en referencia a la Figura 31, calcularemos, utilizando la ecuación de regresión, la Hemoglobina globular que corresponde a un Volumen globular de 60 micras^3 y la que corresponde a un Volumen globular de 170 micras^3.

(a)—Para un Volumen globular de 60 micras^3:

Hb globular = (0.2161 X 60) + 9.2 = 22.2

(b)—Para un Volumen globular de 170 micras^3:

Hb globular = (0.2161 X 170) + 9.2 = 45.9

En el diagrama de la Figura 31 se fijan estos dos puntos (marcados: 'X') y se les une por medio de una línea. Esta es la línea de regresión que indica la relación lineal existente entre Volumen globular y Hemoglobina globular.

Desde un punto de vista estadístico no es justificado ni preciso extender esta línea más allá de los límites inferior y superior dado por las observaciones sobre las que se ha basado el cálculo de dicha línea (en nuestro ejemplo: 60 y 170 micras^3 de Volumen globular).

Es conveniente, en esta clase de diagrama, indicar todos los puntos que corresponden a las observaciones aisladas. Así, en la Figura 31 hemos representado por medio de puntos las 106 observaciones (incluídas en el Cuadro 31), que han servido para el cálculo del coeficiente de correlación, la ecuación de regresión y, finalmente, la línea de regresión.

Es también conveniente indicar en el diagrama el valor del coeficiente de correlación entre las variables que se representan. En nuestro ejemplo, el coeficiente de correlación entre Volumen globular y Hemoglobina globular es: + 0.8758 ± 0.0226. Este es anotado en la esquina superior izquierda del diagrama.

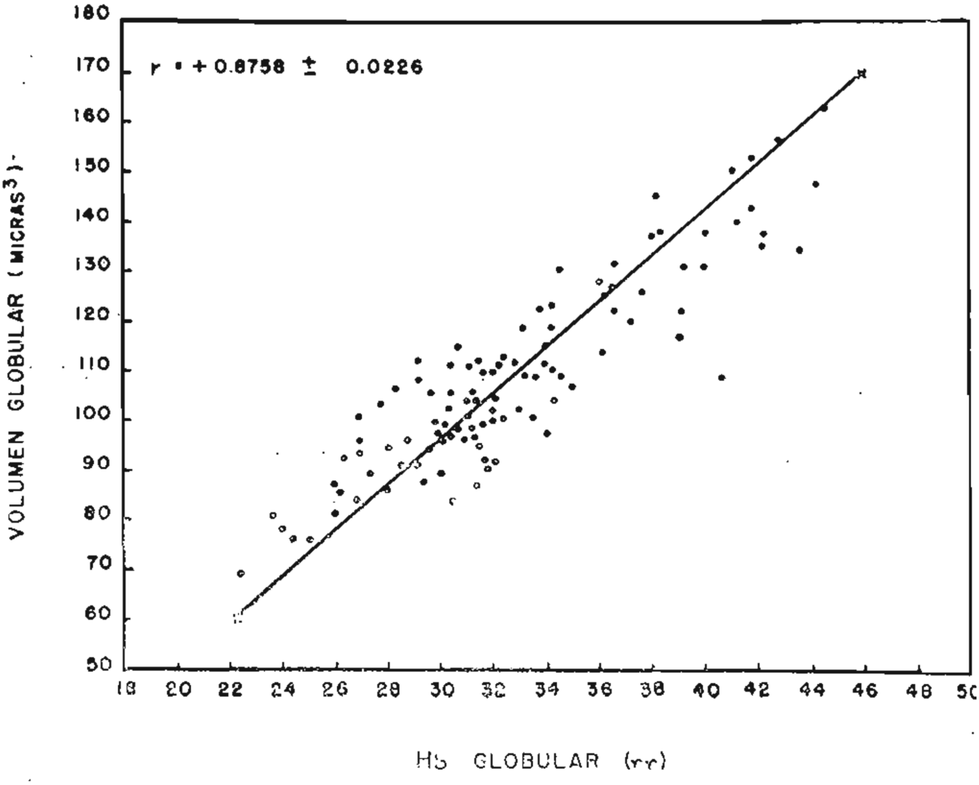

FIGURA 31— Relación entre Volumen y Hemoglobina globular en 106 determinaciones hechas en casos de Enfermedad de Carrión. La línea recta corresponde a la "línea de regresión" derivada del coeficiente de correlación (+ 0.8758 ± 0.0226) hallado. Cada punto corresponde a una observación.

CAPITULO VI

CONSTRUCCION DE UNA LINEA RECTA O CURVA

Cuando se tiene una serie de observaciones aisladas, que se relacionan con dos variables, es conveniente, y a veces importante, construir gráficamente una línea recta o curva que represente la relación general entre dichas variables. Para construir esta línea pueden unirse por medio de líneas, en un diagrama de coordenadas rectangulares, los puntos que corresponden a las observaciones aisladas, pero de esta manera la curva resultante es generalmente irregular, aumentando las fluctuaciones y disminuyendo lo que es realmente importante: la tendencia general.

En la mayoría de los diagramas que se incluyen en publicaciones médicas la curva se construye adaptando, la que, al tanteo, aparezca representar mejor la serie de observaciones aisladas; esta adaptación es visual y sujeta, en gran parte, al criterio subjetivo de quien la hace. Este procedimiento es rutinariamente aceptable, pues el objeto principal de los diagramas es mostrar la tendencia general, y, por consiguiente no son utilizados en derivar datos exactos.

En casos especiales puede, sin embargo, ser necesario calcular la línea recta o curva en la forma más precisa posible, eliminando el factor subjetivo y asegurando que tal línea sea la que fielmente represente la relación indicada por las observaciones aisladas. Tal cálculo ('curve fitting') puede hacerse empleando procedimientos matemáticos.

No es posible, a priori, establecer que tipo de línea va a ser la más apropiada en cada caso; calculada una línea, recta o curva, hay que representarla gráficamente y apreciar si corresponde con exactitud a los datos aislados que han servido para su construcción.

Utilizando un ejemplo, vamos a ilustrar los procedimientos que son necesarios emplear para el cálculo de las siguientes:

A— Línea recta;
B— Curva parabólica; y
C— Curva logarítmica.

Ejemplo: — En varios grupos de sujetos sanos, residentes a diferentes alturas, se ha determinado la saturación al oxígeno (% Hb02) de la sangre arterial.
Los valores medios obtenidos en cada lugar, y las correspondientes alturas, están dados en el Cuadro 37.

Con los datos del Cuadro 37 se ha construído la Figura 32, en la que los valores medios de % Hb02, obtenidos en 9 niveles de altura, están representados por igual número de puntos.

Cuadro 37

Saturación con oxígeno (% Hb02) de la sangre arterial a diversas alturas *.

Observaciones en sujetos residentes

Altura (metros sobre el nivel del mar)	Número de sujetos estudiados	% Hb02 (Media)
150	38	95.9
1,750	10	93.3
2,390	12	91.7
3,140	11	91.0
3,730	15	87.6
4,330	3	83.6
4,540	18	81.4
4,860	12	80.7
5,340	4	76.2

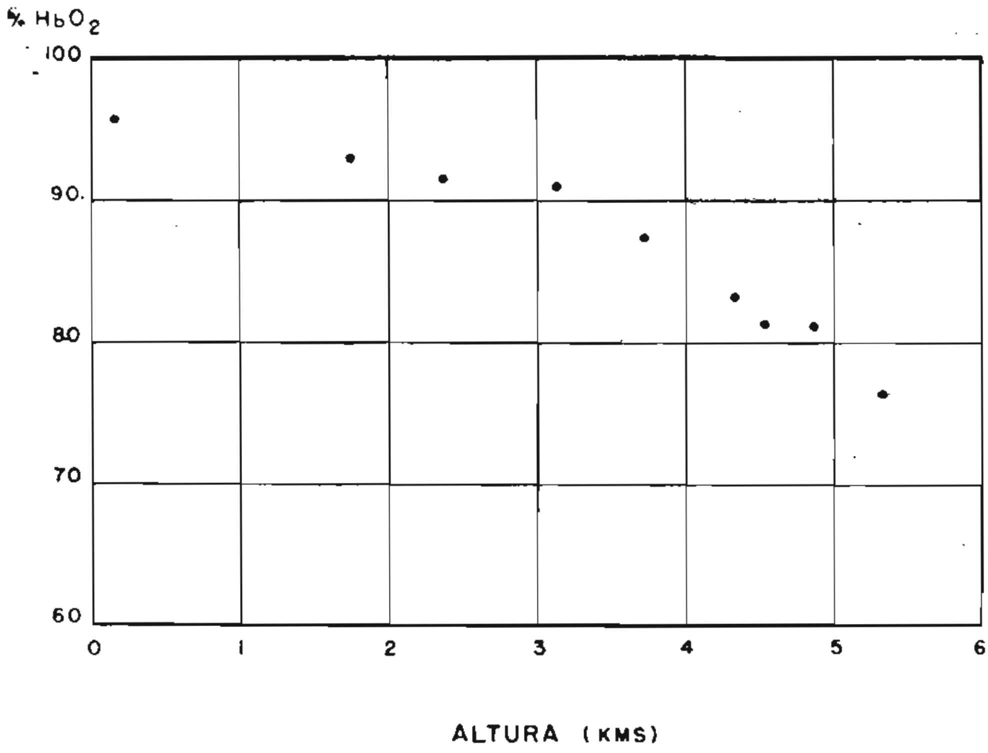

FIGURA 32— Valores medios de la saturación de la sangre arterial con oxígeno (% HbO^2) correspondientes a observaciones hechas en sujetos residentes a diversas alturas.
(Diagrama construído con los datos del Cuadro 37).

* Los valores medios de %Hb0[2] correspondientes a las alturas de 1,750, 4,330 y 5,340 metros han sido tomados de la literatura médica. Los restantes han sido obtenidos en investigaciones realizadas en este Departamento.

Se trata de construir una curva que represente, lo más adecuadamente posible, la relación entre altura y % Hb02, de acuerdo con la posición de los 9 puntos que aparecen en el diagrama de la Figura 32.

El procedimiento más sencillo, con este fin, es unir, por medio de líneas rectas los diversos puntos. Es evidente que tal operación, efectuada en el diagrama de la Figura 33, resulta en una línea no uniforme, que magnifica las fluctuaciones, posiblemente de carácter experimental, y no representa, apropiadamente, la tendencia general de la relación que existe entre altura y % Hb02. Por consiguiente, es necesario calcular matemáticamente una curva que represente, con mayor exactitud, dicha relación.

Aúnque es obvio, por la simple observación visual de la Figura 32, que una línea recta no es la representación adecuada de los 9 puntos, vamos, sin embargo, a efectuar su cálculo para ilustrar los procedimientos que se emplean con este fin.

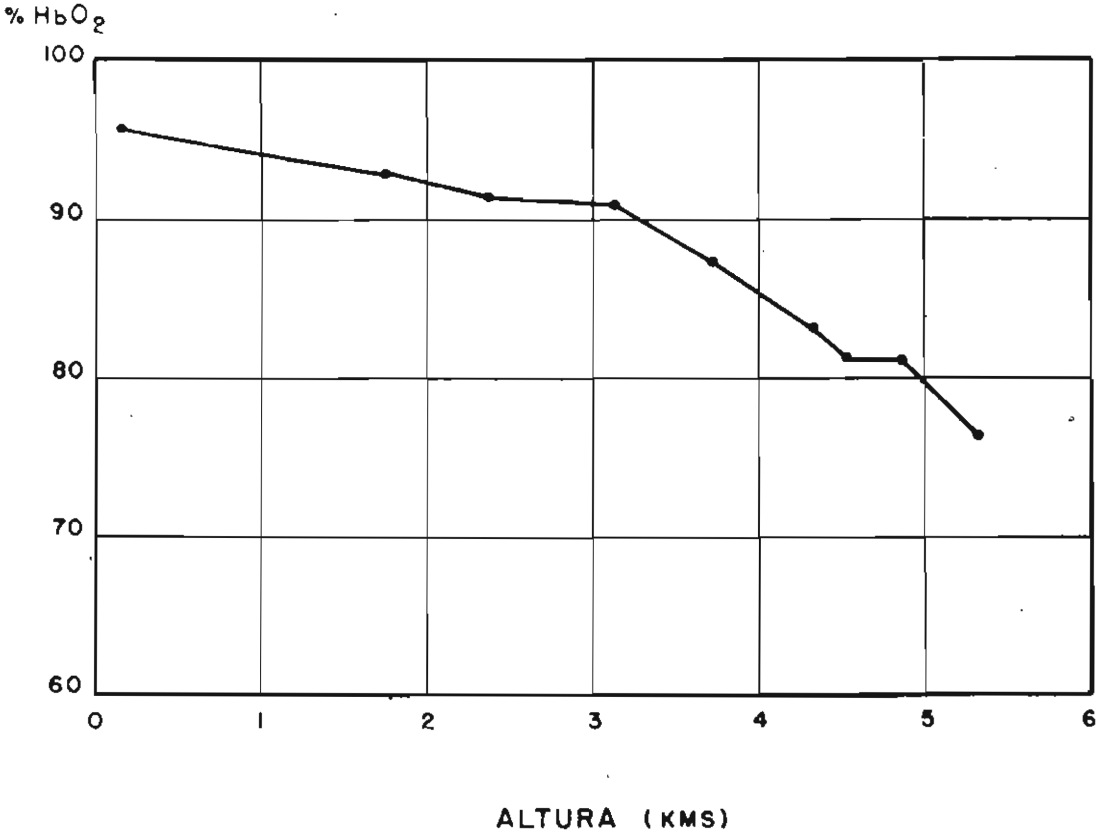

FIGURA 33— Saturación de la sangre arterial con oxigeno (% Hb02) en relación a la altura. Observaciones hechas en sujetos residentes. (En este diagrama se ha unido, por medio de líneas rectas, los puntos representados en la Figura 32, que corresponden a los valores medios hallados en 9 localidades).

A—CALCULO DE UNA LINEA RECTA.

Se construye el Cuadro 38:

Cuadro 38

1	2	3	4
Altura	y	x	xy
150	95.9	1	95.9
1,750	93.3	2	186.6
2,390	91.7	3	275.1
3,140	91.0	4	364.0
3,730	87.6	5	438.0
4,330	83.6	6	501.6
4,540	81.4	7	569.8
4,860	80.7	8	645.6
5,340	76.2	9	685.8
	781.4	45	3762.4

Columna 1 (Altura)—Corresponde a los niveles de altura indicados en el Cuadro 37.

Columna 2 (y)—Corresponde a los valores de % Hb02 indicados en el Cuadro 37.

Columna 3 (x)—Corresponde a la numeración ascendente de cada línea, comenzando por 1.

Columna 4 (xy)—Corresponde a la multiplicación de cada cifra en Columna 2 (y) por su correspondiente en Columna 3 (x).

Así, en nuestro ejemplo:

$$95.9 \times 1 = 95.9$$
$$93.3 \times 2 = 186.6$$
$$91.7 \times 3 = 275.1$$

y así sucesivamente.

En seguida se aplica la fórmula general que corresponde a una línea recta:

$$(1) \; ^{*}y = a + bx$$

Poniendo esta fórmula en ecuación:

$$(2) \quad na + Sx = Sy$$
$$aSx + bSx^2 = Sxy$$

* En esta fórmula, aplicada a nuestro ejemplo, y representa %Hb02.

Los valores correspondientes en nuestro ejemplo son:

n	=	9	Nº total de datos en la Columna 2 (y) del Cuadro 38.
Sx	=	45	Suma de la Columna 3 (x) del Cuadro 38.
Sx^2	=	285	Esta cifra se obtiene directamente en la Columna 3 (Sx^2) del Cuadro 45 (pág. 272) tomando en cuenta el valor más alto de x. En este ejemplo, el valor más alto de x es 9 (Columna 3 (x) del Cuadro 38); a 9 le corresponde 285 en la Columna 3 (Sx^2) del Cuadro 45.
Sy	=	781.4	Suma de la Columna 2 (y) del Cuadro 38.
Sxy	=	3762.4	Suma de la Columna 4 (xy) del Cuadro 38.

Remplazando estos valores en la ecuación (2) tenemos:

$$(3)\quad \begin{aligned} 9a + 45b &= 781.4 \\ 45a + 285b &= 3762.4 \end{aligned}$$

Se despeja 'a' multiplicando las cifras de la línea superior por 45 y las de la inferior por 9:

$$(4)\quad \begin{aligned} 405a + 2025b &= 35163.0 \\ 405a + 2565b &= 33861.6 \end{aligned}$$

Restando el menor del mayor (para considerar cuál es menor se toma en cuenta las cifras después del signo = . En nuestro ejemplo la línea inferior es menor puesto que 33861.6 es menor a 35163.0):

$$(5)\quad \begin{array}{rcrcr} 405a & + & 2025b & = & 35163.0 \\ 405a & + & 2565b & = & 33861.6 \\ \hline 0 & & -540b & = & 1301.4 \end{array}$$

Luego, el valor de 'b' es:

$$(6)\quad b = \frac{1301.4}{-540} = -\ 2.410$$

Remplazando el valor de 'b' en la primera línea de la ecuación (3) y procediendo con las operaciones para hallar el valor de 'a':

$$(7) \quad 9a + (45 \times -2.410) = 781.4$$

$$9a - 108.450 = 781.4$$

$$9a = 781.4 + 108.450$$

$$9a = 889.850$$

$$a = \frac{889.850}{9} = 98.872$$

Remplazando los valores hallados para 'a' y 'b' en la fórmula general (1) tenemos:

$$(8) \quad y = 98.872 + (-2.410x)$$
$$y = 98.872 - 2.410x$$

Para construír la línea recta basta calcular los valores que corresponde a $x = 1$ y a $x = 9$.
Para esto utilizamos la fórmula final de (8):

Cuando $x = 1$

$$y = 98.872 - (2.410 \times 1) = 96.5$$

Cuando $x = 9$

$$y = 98.872 - (2.410 \times 9) = 77.2$$

Finalmente, en el diagrama de la Figura 34, se marcan los puntos (marcados: x) que corresponden a la intersección de los valores 96.5 y 77.2 % HbO_2 con $x = 1$ y $x = 9$, respectivamente, y se unen ambos puntos por medio de una línea recta.

Al marcar estos puntos, en relación con las escalas en la ordenada y en la abcisa, hay que tener presente que $x = 1$ corresponde a una altura de 150 metros, y $x = 9$ corresponde a una altura de 5,340 metros (véuse Cuadro 38).

Se puede apreciar en la Figura 34 que la línea recta, calculada por medio de la fórmula $y = a + bx$, no representa adecuadamente la relación que existe entre altura y % HbO_2, puesto que no coincide con la tendencia expresada por la posición de los 9 puntos que corresponden a las 9 observaciones hechas a diferentes alturas.

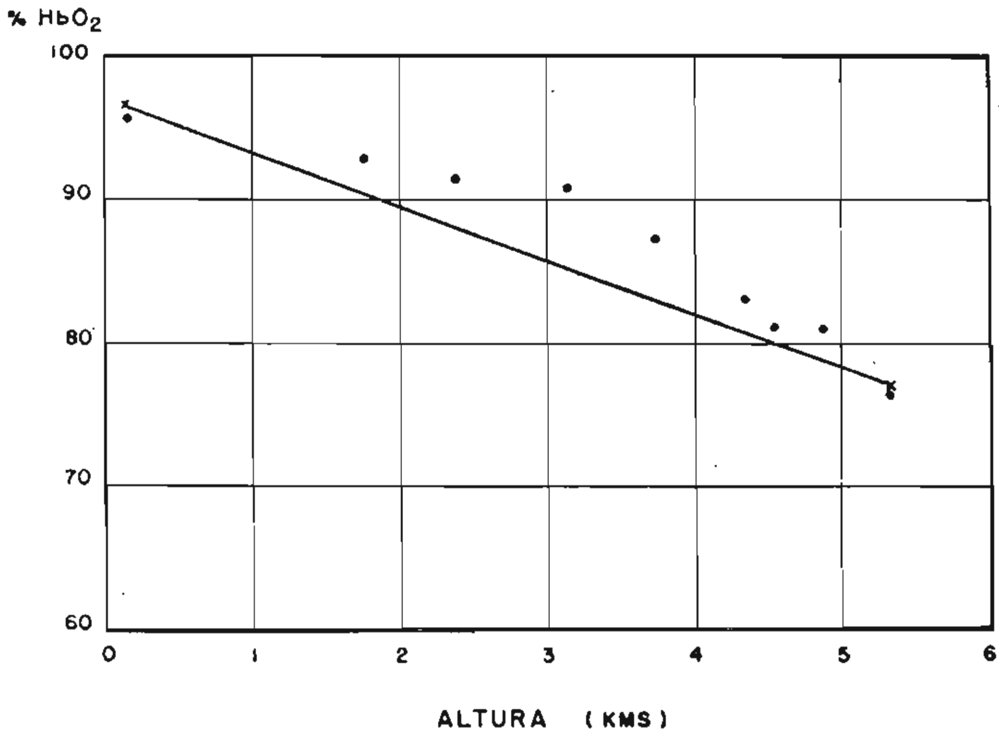

FIGURA 34— Saturación de la sangre arterial con oxígeno (% Hb02) en relación a la altura. Observaciones hechas en sujetos residentes. Los puntos corresponden a los valores medios hallados a diversas alturas. La línea recta ha sido calculada por medio de la fórmula: $y = a + bx$. Puede apreciarse que esta línea recta no se adapta bien a la tendencia expresada por la posición de los 9 puntos, y, por consiguiente, no representa adecuadamente la relación entre altura y % HbO^2.

B—CALCULO DE UNA CURVA PARABOLICA.

Para el cálculo de una curva parabólica que represente la relación entre altura y % Hb02, conforme a la posición de las 9 observaciones incluídas en el diagrama de la Figura 32, es preciso construír, en primer lugar, el Cuadro 39:

CUADRO 39

Altura	y	x	xy	x^2y
1	2	3	4	5
150	95.9	1	95.9	95.9
1,750	93.3	2	186.6	373.2
2,390	91.7	3	275.1	825.3
3,140	91.0	4	364.0	1456.0
3,730	87.6	5	438.0	2190.0
4,330	83.6	6	501.6	3009.6
4,540	81.4	7	569.8	3988.6
4,860	80.7	8	645.6	5164.8
5,340	76.2	9	685.8	6172.2
	781.4	45	3762.4	23275.6

Columnas 1, 2. 3 y 4—Iguales a las del Cuadro 38.

Columnas 5 (x^2y)—Corresponde a la multiplicación de cada cifra en Columna 2 (y) por el cuadrado de su correspondiente cifra en Columna 3 (x).

Así, en nuestro ejemplo:

$$95.9 \times (1)^2 = 95.9$$
$$93.3 \times (2)^2 = 373.2$$
$$91.7 \times (3)^2 = 825.3$$

y así sucesivamente.

En seguida, se aplica la fórmula general que corresponde a una curva parabólica:

$$(1)^* \qquad y = a + bx + cx^2$$

Poniendo esta fórmula en ecuación:

$$(2) \qquad \begin{aligned} na + bSx + cSx^2 &= Sy \\ aSx + bSx^2 + cSx^3 &= Sxy \\ aSx^2 + bSx^3 + cSx^4 &= Sx^2y \end{aligned}$$

Los valores correspondientes en nuestro ejemplo son:

n	=	9	Nº total de datos en la Columna 2 (y) del Cuadro 39.
Sx	=	45	Suma de la Columna 3 (x) del Cuadro 39.
Sx^2	=	285	Esta cifra se obtiene directamente en la Co-

* En esta fórmula, aplicada a nuestro ejemplo, y representa %Hb02.

		lumna 3 (Sx^2) del Cuadro 45 (pág. 272) tomando en cuenta el valor más alto de x. En este ejemplo el valor más alto de x es 9 (Columna 3 (x) del Cuadro 39); a 9 le corresponde 285 en la Columna 3 (Sx^2) del Cuadro 45.
Sx^3	= 2025	Esta cifra se obtiene directamente en la Columna 4 (Sx^3) del Cuadro 45. El procedimiento es el mismo que el que se usa para hallar el valor de Sx^2.
Sx^4	= 15333	Esta cifra se obtiene directamente en la Columna 5 (Sx^4) del Cuadro 45. El procedimiento es el mismo que el se usa para hallar el valor de Sx^2.
Sy	= 781.4	Suma de la Columna 2 (y) del Cuadro 39.
Sxy	= 3762.4	Suma de la Columna 4 (xy) del Cuadro 39.
Sx^2y	= 23275.6	Suma de la Columna 5 (x^2y) del Cuadro 39.

Remplazando estos valores en la ecuación (2) tenemos:

$$
\begin{array}{lrcrcrcr}
(3) & 9a & + & 45b & + & 285c & = & 781.4 \\
 & 45a & + & 285b & + & 2025c & = & 3762.4 \\
 & 285a & + & 2025b & + & 15333c & = & 23275.6
\end{array}
$$

Tomamos las dos primeras líneas de la ecuación (3):

$$
\begin{array}{lrcrcrcr}
(4) & 9a & + & 45b & + & 285c & = & 781.4 \\
 & 45a & + & 285b & + & 2025c & = & 3762.4
\end{array}
$$

Se despeja "a" multiplicando las cifras de la línea superior por 45 y las de la línea inferior por 9:

$$
\begin{array}{lrcrcrcr}
(5) & 405a & + & 2025b & + & 12825c & = & 35163.0 \\
 & 405a & + & 2565b & + & 18225c & = & 33861.6
\end{array}
$$

Restando el menor del mayor (para considerar cuál es el menor se toma en cuenta las cifras después del signo =. En nuestro ejemplo la línea inferior es menor puesto que 33861.6 es menor que 35163.0):

$$
\begin{array}{lrcrcrcr}
(6) & 405a & + & 2025b & + & 12825c & = & 35163.0 \\
 & 405a & + & 2565b & + & 18225c & = & 33861.6 \\
\hline
 & 0 & & -540b & + & -5400c & = & 1301.4
\end{array}
$$

En seguida, se toman la 1ª y 3ª línea de la ecuación (3):

$$
\begin{array}{lrcrcrcr}
(7) & 9a & + & 45b & + & 285c & = & 781.4 \\
 & 285a & + & 2025b & + & 15333c & = & 23275.6
\end{array}
$$

Se despeja "a" multiplicando las cifras de la línea superior por 285 y las de la línea inferior por 9:

$$(8) \quad \begin{array}{l} 2565a + 12825b + 81225c = 222699.0 \\ 2565a + 18225b + 137997c = 209480.4 \end{array}$$

Restando el menor del mayor:

$$(9) \quad \begin{array}{l} 2565a + 12825b + 81225c = 222699.0 \\ 2565a + 18225b + 137997c = 209480.4 \\ \hline 0 \quad -5400b + -56772c = 13218.6 \end{array}$$

Se toma el resultado obtenido en (6) y el obtenido en (9):

$$(10) \quad \begin{array}{l} -540b + -5400c = 1301.4 \\ -5400b + -56772c = 13218.6 \end{array}$$

Se despeja "b" multiplicando las cifras de la línea superior por 5400 y las de la línea inferior por 540:

$$(11) \quad \begin{array}{l} -2916000b + -29160000c = 7027560 \\ -2916000b + -30656880c = 7138044 \end{array}$$

Restando el menor del mayor:

$$(12) \quad \begin{array}{l} -2916000b + -30656880c = 7138044 \\ -2916000b + -29160000c = 7027560 \\ \hline 0 \quad -1496880c = 110484 \end{array}$$

Luego, el valor de "c" es:

$$(13) \quad c = \frac{110484}{-1496880} = -0.074$$

Remplazando el valor de "c" en la primera línea de la ecuación (10) y procediendo con las operaciones para hallar el valor de "b":

$$(14) \quad \begin{array}{l} -540b + (-5400 \times -0.074) = 1301.4 \\ -540b + 339.6 = 1301.4 \\ -540b = 1301.4 - 399.6 \end{array}$$

$$-540b = 901.8$$

$$b = \frac{901.8}{-540} = -1.670$$

Remplazando los valores de "b" y "c" en la primera línea de la ecuación (3), y procediendo con las operaciones para hallar el valor de "a":

$$(15) \quad 9a + (45 \times -1.670) + (285 \times -0.074) = 781.4$$

$$9a - 75.150 - 21.090 = 781.4$$

$$9a - 96.240 = 781.4$$

$$9a = 781.4 + 96.240$$

$$9a = 877.640$$

$$a = \frac{877.640}{9} = 97.516$$

Remplazando los valores hallados para "a", "b" y "c" en la fórmula general (1) tenemos:

$$(16) \quad y = 97.516 - 1.670x - 0.074x^2$$

Para construír la curva parabólica hay que calcular los valores que corresponden a x, desde x = 1 hasta x = 9.
Para esto utilizamos la fórmula final de (16):

Cuando x = 1:

$$y = 97.516 - (1.670 \times 1) - (0.074 \times 1) = 95.8$$

Cuando x = 2:

$$y = 97.516 - (1.670 \times 2) - (0.074 \times 4) = 93.9$$

Cuando x = 3:

$$y = 97.516 - (1.670 \times 3) - (0.074 \times 9) = 91.8$$

Cuando x = 4:

$$y = 97.516 - (1.670 \times 4) - (0.074 \times 16) = 89.7$$

Cuando x = 5:

$$y = 97.516 - (1.670 \times 5) - (0.074 \times 25) = 87.3$$

Cuando x = 6:

$$y = 97.516 - (1.670 \times 6) - (0.074 \times 36) = 84.8$$

Cuando $x = 7$:

$$y = 97.516 - (1.670 \times 7) - (0.074 \times 49) = 82.2$$

Cuando $x = 8$:

$$y = 97.516 - (1.670 \times 8) - (0.074 \times 64) = 79.4$$

Cuando $x = 9$:

$$y = 97.516 - (1.670 \times 9) - (0.074 \times 81) = 76.5$$

Según estos valores hallados para $x = 1$ hasta $x = 9$, tenemos que la curva parabólica debe construirse tomando en cuenta las siguientes correspondencias (Cuadro 40): *

CUADRO 40

VALORES PARA LA CONSTRUCCION DE LA CURVA PARABOLICA

x	y
Altura	%Hb02
150	95.8
1,750	93.9
2,390	91.8
3,140	89.7
3,730	87.3
4,330	84.8
4,540	82.2
4,860	79.4
5,340	76.5

Utilizando los valores dados en el Cuadro 40 se construye la curva parabólica de la Figura 35. Para esta construcción se marcan suavemente (para que no aparezcan en el diagrama final) los puntos correspondientes a la intersección de los 9 valores de altura y %Hb02, y en seguida utilizando una regla flexible, u otra apropiada, se traza una línea que pase por todos los puntos.

Se puede apreciar en la Figura 35 que la curva parabólica, calculada por medio de la fórmula: $y = a + bx + cx^2$ representa, bastante bien, la relación que existe entre altura y %Hb02, puesto que corresponde, aproximadamente, a la tendencia expresada por la posición de las 9 observaciones hechas a diferentes alturas.

* En el Cuadro 40 se ha tomado en cuenta que, según lo expresado en el Cuadro 39, $X = 1$ corresponde a una altura de 150 metros; $X = 2$ corresponde a una altura de 1,750 metros; $X = 3$ corresponde a una altura de 2,390 metros, y así, sucesivamente.

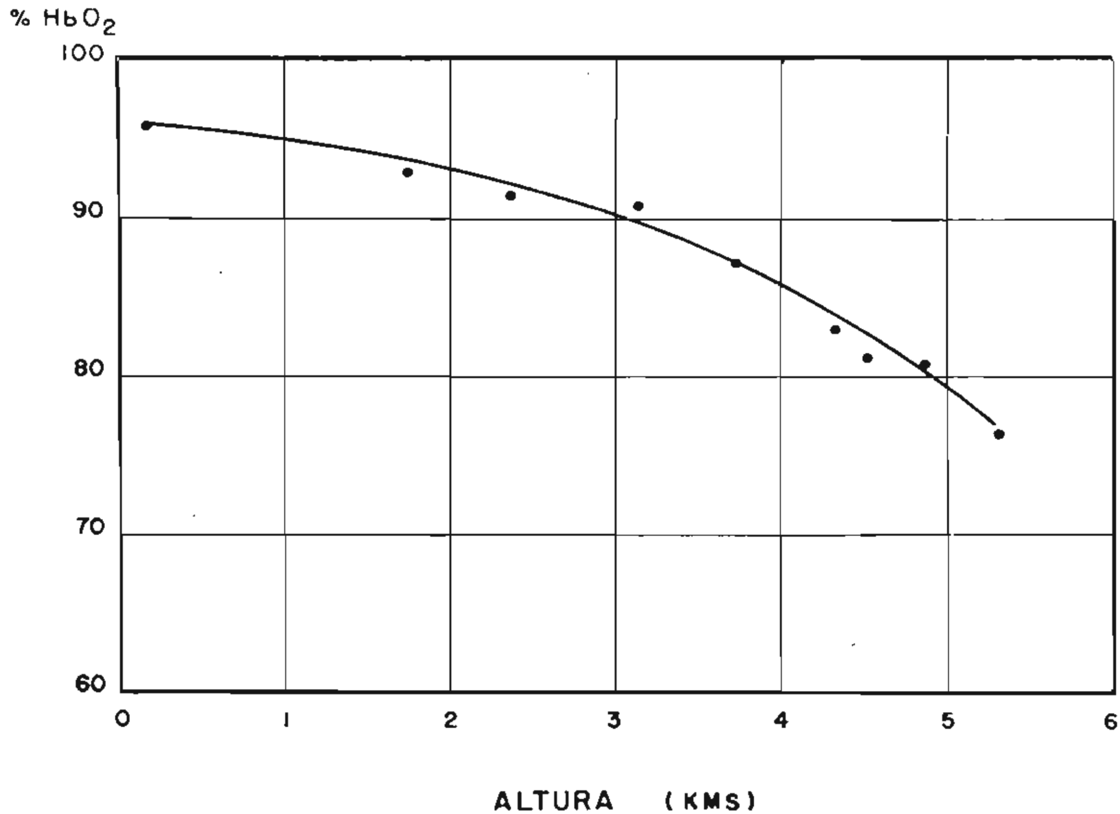

FIGURA 35— Saturación de la sangre arterial con oxígeno (% Hb02) en relación a la altura. Observaciones hechas en sujetos residentes. Los puntos corresponden a los valores medios hallados a diversas alturas. La línea curva, que ha sido calculada por medio de la fórmula: $y = a + bx + cx^2$, expresa, en forma adecuada, la relación entre altura y % Hb02, de acuerdo con la posición de los 9 puntos. Para la construcción de esta curva se han utilizado los datos contenidos en el Cuadro 40.

C—CALCULO DE UNA CURVA LOGARITMICA.

Para el cálculo de una curva logarítmica, que represente la relación entre altura y %Hb02, conforme a la posición de las 9 observaciones incluídas en el diagrama de la Figura 32, es preciso construir, en primer lugar, el Cuadro 41 :

CUADRO 41

1	2	3	4	5
Altura	y	x	xy	ylogx
150	95.9	1	95.9	0
1,750	93.3	2	186.6	28.0833
2,390	91.7	3	275.1	43.7501
3,140	91.0	4	364.0	54.7911
3,730	87.6	5	438.0	61.2324
4,330	83.6	6	501.6	65.0575
4,540	81.4	7	569.8	63.7911
4,860	80.7	8	645.6	72.8802
5,340	76.2	9	685.8	72.7100
	781.4	45	3762.4	467.2957

Columnas 1, 2, 3 y 4—Iguales a las del Cuadro 39.

Columna 5 (ylogx)—Corresponde a la multiplicación de cada cifra en Columna 2 (y) por el logaritmo de la correspondiente cifra en Columna 3 (x).
Los logaritmos de estas últimas cifras pueden obtenerse en el Apéndice A.
Así, en nuestro ejemplo:

log. de 1 = 0 Luego 95.9 $\times$ 0 = 0

log. de 2 = 0.3010 Luego 93.3 $\times$ 0.3010 = 28.0833

y así sucesivamente.

En seguida, se aplica la fórmula general que corresponde a una curva logarítmica:

$$(1)^{*} \qquad y = a + bx + c\log x$$

Poniendo esta fórmula en ecuación:

$$(2) \qquad \begin{aligned} na &+ bSx &+ c\log x &= Sy \\ aSx &+ bSx^2 &+ cSx\log x &= Sxy \\ aS\log x &+ bSx\log x &+ cS\log x^2 &= Sy\log x \end{aligned}$$

* En esta fórmula, aplicada a nuestro ejemplo, y representa %Hb02.

Los valores correspondientes en nuestro ejemplo son:

n	=	9	Nº total de datos en la Columna 2 (y) del Cuadro 41.
Sx	=	45	Suma de la Columna 3 (x) del Cuadro 41.
Sx^2	=	285	Esta cifra se obtiene directamente en la Columna 3 (Sx^2) del Cuadro 45 (pág. 272) tomando en cuenta el valor más alto de x. En este ejemplo el valor más alto de x es 9 (Columna 3 (x) del Cuadro 41); a 9 le corresponde 285 en la Columna 3 (Sx^2) del Cuadro 45.
Sy	=	781.4	Suma de la Columna 2 (y) del Cuadro 41.
Sxy	=	3762.4	Suma de las cifras de la Columna 4 (xy) del Cuadro 41.
Sylogx	=	467.2957	Suma de la Columna 5 (ylogx) del Cuadro 41.
Slogx	=	5.5598	Esta cifra se obtiene directamente en la Columna 2 del Cuadro 46 (pág. 273) tomando en cuenta el valor más alto de x. En este ejemplo el valor más alto de x es 9 (Columna 3 (x) del Cuadro 41); a 9 le corresponde 5.5598 en la Columna 2 (Slogx) del Cuadro 46.
Sxlogx	=	34.3340	Esta cifra se obtiene directamente en la Columna 3 (Sxlogx) del Cuadro 46. El procedimiento es el mismo que el que se usa para obtener el valor de Slogx.
$Slogx^2$	=	4.2152	Esta cifra se obtiene directamente en la Columna 4 ($Slogx^2$) del Cuadro 46. El procedimiento es el mismo que el que se usa para obtener el valor de Slogx.

Reemplazando estos valores en la ecuación (2) tenemos:

$$(3)\qquad \begin{aligned} 9a + 45b + 5.5598c &= 781.4 \\ 45a + 285b + 34.3340c &= 3762.4 \\ 5.5598a + 34.3340b + 4.2152c &= 467.2957 \end{aligned}$$

Tomamos las dos primeras líneas de la ecuación (3):

$$(4)\qquad \begin{aligned} 9a + 45b + 5.5598c &= 781.4 \\ 45a + 285b + 34.3340c &= 3762.4 \end{aligned}$$

Se despeja "a" multiplicando las cifras de la línea superior por 45 y las de la línea inferior por 9:

$$
\begin{array}{lrcrcrcr}
(5) & 405a & + & 2025b & + & 250.191c & = & 35163.0 \\
 & 405a & + & 2565b & + & 309.006c & = & 33861.6
\end{array}
$$

Restando el menor del mayor (para considerar cual es el menor se toma en cuenta las cifras después del signo =. En nuestro ejemplo la línea inferior es menor puesto que 33861.6 es menor que 35163.0):

$$
\begin{array}{lrcrcrcr}
(6) & 405a & + & 2025b & + & 250.191c & = & 35163.0 \\
 & 405a & + & 2565b & + & 309.006c & = & 33861.6 \\
\hline
 & 0 & & -540b & + & -58.815c & = & 1301.4
\end{array}
$$

En seguida se toman la 1ª y 3ª línea de la ecuación (3):

$$
\begin{array}{lrcrcrcr}
(7) & 9a & + & 45b & + & 5.5598c & = & 781.4 \\
 & 5.5598a & + & 34.3340b & + & 4.2152c & = & 467.2957
\end{array}
$$

Se despeja "a" multiplicando las cifras de la línea superior por 5.5598 y las de la línea inferior por 9:

$$
\begin{array}{lrcrcrcr}
(8) & 50.0382a & + & 250.191b & + & 30.9114c & = & 4344.4277 \\
 & 50.0382a & + & 309.006b & + & 37.9368c & = & 4205.6613
\end{array}
$$

Restando el menor del mayor:

$$
\begin{array}{lrcrcrcr}
(9) & 50.0382a & + & 250.191b & + & 30.9114c & = & 4344.4277 \\
 & 50.0382a & + & 309.006b & + & 37.9368c & = & 4205.6616 \\
\hline
 & 0 & & -58.815b & + & -7.0254c & = & 138.7664
\end{array}
$$

Se toma el resultado obtenido en (6) y el obtenido en (9):

$$
\begin{array}{lrcrcr}
(10) & -540b & + & -58.815c & = & 1301.4 \\
 & -58.815b & + & -7.0254c & = & 138.7664
\end{array}
$$

Se despeja "b" multiplicando las cifras de la línea superior por 58.815 y las de la línea inferior por 540:

$$
\begin{array}{lrcrcr}
(11) & -31760.100b & + & -3459.2042c & = & 76541.841 \\
 & -31760.100b & + & -3793.7160c & = & 74933.856
\end{array}
$$

Restando el menor del mayor:

$$
\begin{array}{lrcrcr}
(12) & -31760.100b & + & -3459.2042c & = & 76541.841 \\
 & -31760.100b & + & -3793.7160c & = & 74933.856 \\
\hline
 & 0 & & -334.5118c & = & 1607.985
\end{array}
$$

Luego el valor de "c" es:

$$(13) \qquad c = \frac{1607.985}{-334.5118} = -4.807$$

Reemplazando el valor de "c" en la primera línea de la ecuación (10) y procediendo con las operaciones para hallar el valor de "b":

$$(14) \qquad -540b + (58.815 \times -4.807) = 1301.4$$
$$-540b + 282.7237 = 1301.4$$
$$-540b = 1301.4 - 282.7237$$
$$-540b = 1018.6763$$

$$b = \frac{1018.6763}{-540} = -1.886$$

Reemplazando los valores de "b" y "c" en la primera línea de la ecuación (3) y procediendo con las operaciones para hallar el valor de "a":

$$(15) \qquad 9a + (45 \times -1.886) + (5.5598 \times -4.807) = 781.4$$

$$9a - 84.870 - 26.726 = 781.4$$
$$9a - 111.596 = 781.4$$
$$9a = 781.4 + 111.596$$
$$9a = 892.996$$

$$a = \frac{892.996}{9} = 99.221$$

Reemplazando los valores hallados para "a", "b" y "c" en la fórmula general (1) tenemos:

$$(16) \qquad y = 99.221 - 1.886x - 4.807 \log x$$

Para construir la curva logarítmica hay que calcular los valores que corresponden a x, desde $x = 1$ hasta $x = 9$.

Para esto utilizamos la fórmula final de (16)*:

Cuando $x = 1$:

$$y = 99.221 - (1.886 \times 1) - (4.807 \times 0) = 97.3$$

Cuando $x = 2$:

* En esta fórmula los logarítmicos de 1 hasta 9 (que deben multiplicarse por 4.807) son hallados en el Apéndice A.

$$y = 99.221 - (1.886 \times 2) - (4.807 \times 0.3010) = 94.0$$

Cuando $x = 3$:

$$y = 99.221 - (1.886 \times 3) - (4.807 \times 0.4771) = 91.3$$

Cuando $x = 4$:

$$y = 99.221 - (1.886 \times 4) - (4.807 \times 0.6021) = 88.8$$

Cuando $x = 5$:

$$y = 99.221 - (1.886 \times 5) - (4.807 \times 0.6990) = 86.4$$

Cuando $x = 6$:

$$y = 99.221 - (1.886 \times 6) - (4.807 \times 0.7782) = 84.2$$

Cuando $x = 7$:

$$y = 99.221 - (1.886 \times 7) - (4.807 \times 0.8451) = 82.0$$

Cuando $x = 8$:

$$y = 99.221 - (1.886 \times 8) - (4.807 \times 0.9031) = 79.8$$

Cuando $x = 9$:

$$y = 99.221 - (1.886 \times 9) - (4.807 \times 0.9542) = 77.7$$

Según estos valores hallados para $x = 1$ hasta $x = 9$, tenemos que la curva logarítmica debe construírse tomando en cuenta las siguientes correspondencias (Cuadro 42*):

CUADRO 42

VALORES PARA LA CONSTRUCCION DE LA CURVA LOGARITMICA

x	y
Altura	% Hb02
150	97.3
1,750	94.0
2,390	91.3
3,140	88.8
3,730	86.4
4,330	84.2
4,540	82.0
4,860	79.8
5,340	77.7

* En el Cuadro 42 se ha tomado en cuenta que, según lo expresado en el Cuadro 41, $X = 1$ corresponde a una altura de 150 metros; $X = 2$ corresponde a una altura de 1,750 metros; $X = 3$ corresponde a una altura de 2,390 metros, y así, sucesivamente.

Utilizando los valores dados en el Cuadro 42 se construye la curva logarítmica de la Figura 36. Para esta construcción se marcan suavemente (para que no aparezcan en el diagrama final) los puntos correspondientes a la intersección de los 9 valores de altura y % Hb02, y, en seguida, con una regla flexible, u otra apropiada, se traza una línea que pase por todos los puntos.

Se puede apreciar en la Figura 36 que la curva logarítmica, calculada por medio de la fórmula: $y = a + bx + c\log x$, representa con menos precisión que la curva parabólica (Figura 35) la relación existente entre altura y % Hb02.

Por consiguiente, la conclusión a que se llega en el ejemplo que hemos discutido es que la curva parabólica es la más adecuada para representar la relación que existe entre altura y % Hb02, de acuerdo con la posición de los 9 puntos que corresponden a las observaciones hechas a diferentes alturas.

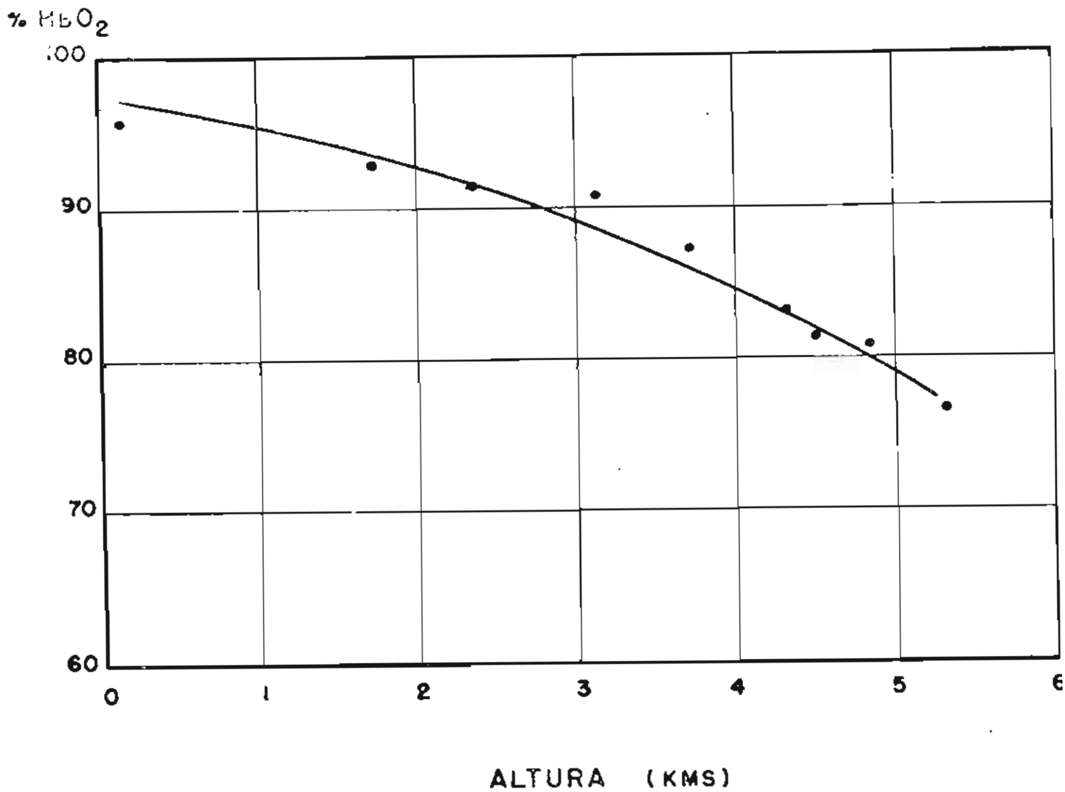

FIGURA 36— Saturación de la sangre arterial con oxígeno (% Hb02) en relación a la altura. Observaciones hechas en sujetos residentes. Los puntos corresponden a los valores medios hallados a diversas alturas. La línea curva, que ha sido calculada por medio de la fórmula: $y = a + bx + c\log x$, expresa, en forma adecuada, aunque con menos precisión que la curva de la Figura 35, la relación entre altura y % Hb02, de acuerdo con la posición de los 9 puntos. Para la construcción de esta curva se han utilizado los datos del Cuadro 42.

Si el cálculo de la línea recta, o de las curvas parabólica o logarítmica, que representan la relación existente entre dos variables, corresponde a un número elevado de observaciones, es conveniente, como procedimiento previo, arreglar los datos de la manera siguiente:

(a)—Dividir a una de las variables en grupos de Frecuencia, indicando el punto medio de cada grupo; y

(b)—Calcular para cada uno de los grupos de frecuencia de (a) el valor promedio que le corresponde en la otra variable.

En seguida, la línea recta, o curva parabólica o logarítmica, se calculan relacionando el punto medio de los grupos de frecuencia de una de las variables con los respectivos valores medios de la otra.

Utilizaremos un ejemplo para ilustrar los procedimientos a seguir.

Ejemplo *: Se ha determinado el peso y la estatura (sentada) de 454 embrios humanos. Con los resultados obtenidos se construye el Cuadro 43.

CUADRO 43

1	2	3
Peso en gramos Grupos de frecuencia	Punto medio	Valores medios de la estatura (mm)
0 — 19	10	58.8
20 — 39	30	76.4
40 — 59	50	91.1
60 — 79	70	99.0
80 — 99	90	108.1
100 — 119	110	115.1
120 — 139	130	122.7
140 — 159	150	129.5
160 — 179	170	135.0
180 — 199	190	141.1
200 — 219	210	144.0
220 — 239	230	150.0
240 — 259	250	152.8
260 — 279	270	155.6
280 — 299	290	158.6
300 — 319	310	161.3
320 — 339	330	160.5
360 — 379	350	171.0
340 — 359	370	169.5
380 — 399	390	173.6

* Tomado de Introduction to Medical Biometry and Statistics. R. Pearl — W. B. Saunders Co., Philadelphia, 1940.

Columna 1— Corresponde a la división de los datos del peso (gramos), obtenidos en las 454 determinaciones, en grupos de frecuencia.

La manera como se procede para calcular los grupos de frecuencia que corresponde a una serie de datos está explicada en el Capítulo I.

En este ejemplo se ha dividido, arbitrariamente, los datos de peso en 20 grupos de frecuencia; el primer grupo incluye el valor lor más bajo (12 gramos) y el último grupo incluye el valor más alto (387 gramos).

El intervalo de grupo equivale a 20 gramos.

Columna 2— Corresponde al punto medio de cada grupo de frecuencia.

El punto medio es igual a la mitad del intervalo de grupo más el valor correspondiente al límite inferior de cada grupo de frecuencia.

Así, en nuestro ejemplo (en el que el intervalo de grupo es igual a 20) tenemos:

Para el 1er. grupo: $\frac{20}{2} + 0 = 10$

Para el 2º. grupo: $\frac{20}{2} + 20 = 30$

Para el 3er. grupo: $\frac{20}{2} + 40 = 50$

y así sucesivamente.

Columna 3— Son los valores medios de la estatura (mm) que corresponden a cada uno de los grupos de frecuencia del peso.

Así, en nuestro ejemplo tenemos que correspondiendo a un peso de 0 a 19 gramos hay 4 datos de estatura:

62.8	60.3
54.8	57.3

El valor promedio equivale a la suma de estos 4 valores (235.2) dividida entre 4:

$$\frac{235.2}{4} = 58.8 \text{ mm}$$

Para un peso que varía entre 20 y 39 gramos tenemos 8 observaciones de estatura:

79.4	77.6
73.4	75.2
78.2	77.5
74.6	75.3

El valor promedio equivale a la suma de estos 8 valores (611.2) dividida entre 8:

$$\frac{611.2}{8} = 76.4 \text{ mm}$$

Y así sucesivamente, se calcula, utilizando los datos originales de las observaciones hechas, los valores promedios de estatura que corresponden a los demás grupos de frecuencia del peso.

En seguida se construye el Cuadro 44.

CUADRO 44

1	2	3
Peso	y	x
10	58.8	1
30	76.4	2
50	91.1	3
70	99.0	4
90	108.1	5
110	115.1	6
130	122.7	7
150	129.5	8
170	135.0	9
190	141.1	10
210	144.0	11
230	150.0	12
250	152.8	13
270	155.6	14
290	158.6	15
310	161.3	16
330	160.5	17
350	171.0	18
370	169.5	19
390	173.6	20

Columna 1—Es igual a la Columna 2 del Cuadro 43.
Columna 2—Es igual a la Columna 3 del Cuadro 43.
Columna 3—Corresponde a la numeración ascendente de cada línea comenzando por 1.

Finalmente, los datos contenidos en el Cuadro 44 son utilizados para el cálculo de la línea recta o de las curvas parabólica y logarítmica, de acuerdo con las instrucciones dadas en las páginas 252 a 267.

En este ejemplo, la curva logarítmica es la que mejor corresponde a la posición de los 20 puntos en el diagrama, y, por consiguiente, es la que mejor representa la relación entre estatura y peso de los embrios (Figura 37).

La curva logarítmica de la Figura 37 ha sido construída calculando, por medio de la fórmula $y = a + bx + c\log x$, los valores de estatura (mm) que corresponde a $x = 1$ (peso: 10 gramos); $x = 2$ (peso: 30 gramos); $x = 3$ (peso: 50 gramos), y así sucesivamente hasta $x = 20$ (peso: 390 gramos).

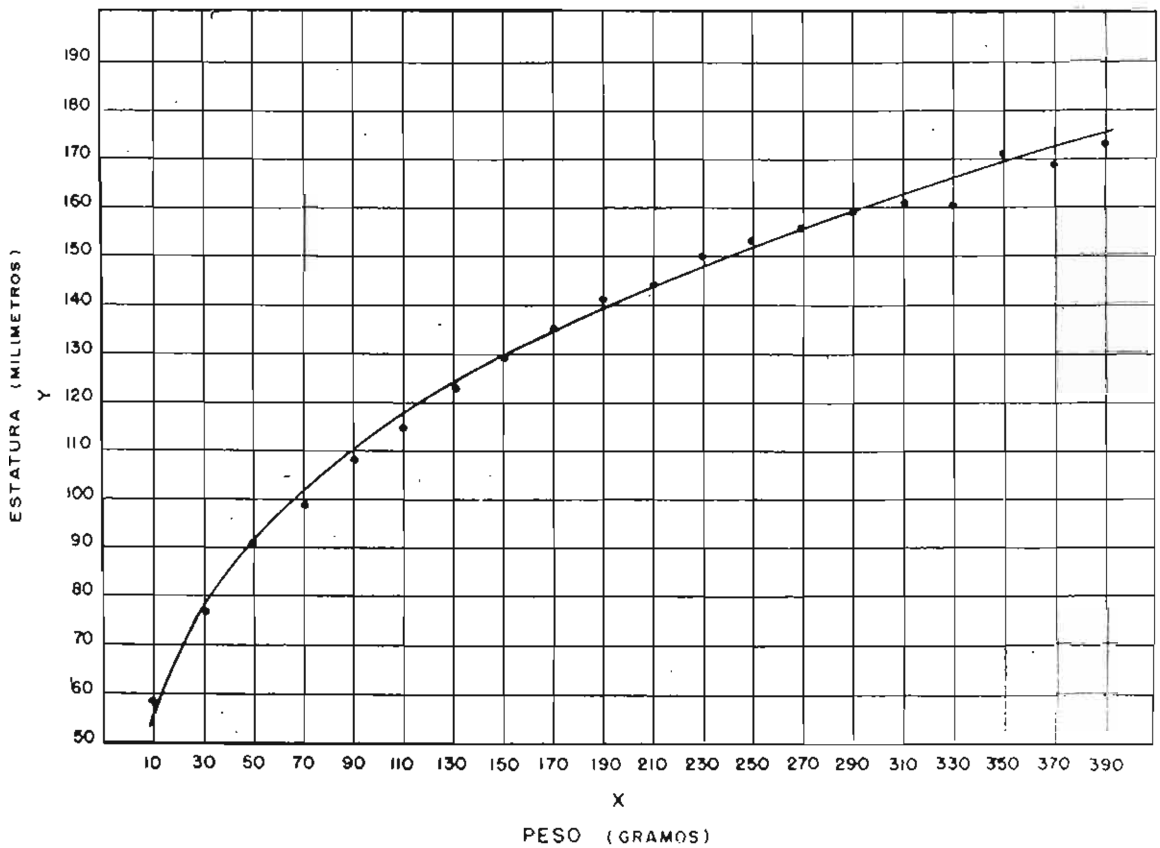

FIGURA 37— Relación entre el peso y la estatura (sentada) de embrios humanos. La curva ha sido construída por medio de la fórmula: $y = a + bx + c\log x$.

CUADRO 45

PARA SER USADO EN LOS CALCULOS DE UNA LINEA RECTA Y DE CURVAS PARABOLICA Y LOGARITMICA

1	2	3	4	5
x	Suma x	Suma x^2	Suma x^3	Suma x^4
1	1	1	1	1
2	3	5	9	17
3	6	14	36	98
4	10	30	100	354
5	15	55	225	979
6	21	91	441	2.275
7	28	140	784	4,676
8	36	204	1.296	8,772
9	45	285	2.025	15.333
10	55	385	3,025	25,333
11	66	506	4,356	39,974
12	78	650	6,084	60,710
13	91	819	8,281	89,271
14	105	1,015	11,025	127.687
15	120	1,240	14,400	178.312
16	136	1,496	18,496	243,848
17	153	1,785	23,409	327,369
18	171	2,109	29,241	432,345
19	190	2,470	36,100	562,666
20	210	2,870	44.100	722,666
21	231	3,311	53,361	917.147
22	253	3,795	64,009	1,151.403
23	276	4,324	76,176	1.431.244
24	300	4.876	89,424	1,749.196
25	325	5.501	105.039	2,139.821
26	351	6,177	122,625	2,596.797
27	378	6.906	142,308	3,128,238
28	406	7,690	164,260	3.742.894
29	435	8,531	188,649	4.450.175
30	465	9,431	215.649	5,260,175
31	496	10,392	245.440	6,183,696
32	528	11.416	278.208	7.232.272
33	561	12,505	314.145	8,418.193
34	595	13,661	353.449	9.754.529
35	630	14.886	396.324	11.255.154
36	666	16.182	442,920	12.934.770
37	703	17.551	493.633	14.808.931
38	741	18,995	548.505	16.894.067
39	780	20,516	607,824	19,207,508
40	820	22.116	671.824	21.767.503
41	861	23.797	740,745	24.593.269
42	903	25,561	814.833	27,704.965
43	946	27.410	894.340	31,123.766
44	990	29,346	979.524	34.871.862
45	1.035	31,371	1,070,649	38.972.487
46	1.081	33,487	1.167,985	43,449,943
47	1,128	35,696	1,271,808	48,329,624
48	1,176	38,000	1,382,400	53,638,040
49	1,225	40,401	1,500,049	59,402,841
50	1,275	42,901	1,625,049	65,652,841

CUADRO 46

PARA SER USADO EN EL CALCULO DE LA CURVA LOGARITMICA

1	2	3	4
x	Slogx	Sxlogx	Slogx²
1	0.000000	0.000000	0.000000
2	0.301030	0.602060	0.090619
3	0.778151	2.033424	0.318264
4	1.380211	4.441664	0.680740
5	2.079181	7.936514	1.169299
6	2.857332	12.605412	1.774818
7	3.702430	18.521107	2.489009
8	4.605520	25.745827	3.304581
9	5.559763	34.334010	4.215159
10	6.559763	44.334010	5.215159
11	7.601156	55.789329	6.299658
12	8.680337	68.739504	7.464290
13	9.794280	83.220768	8.705160
14	10.940408	99.266561	10.018770
15	12.116500	116.907929	11.401960
16	13.320620	136.173849	12.851865
17	14.551068	157.091481	14.365870
18	15.806341	179.686386	15.941579
19	17.085095	203.982704	17.576789
20	18.386125	230.003304	19.269469
21	19.708344	257.769909	21.017732
22	21.050767	287.303208	22.819831
23	22.412494	318.622949	24.674134
24	23.792706	351.748018	26.579117
25	25.190646	386.696519	28.533353
26	26.605619	423.485826	30.535503
27	28.036983	462.132647	32.584305
28	29.484141	502.653072	34.678571
29	30.946539	545.062614	36.817179
30	32.423660	589.376252	38.999066
31	33.915022	635.608464	41.223226
32	35.420172	683.773264	43.488703
33	36.938686	733.884224	45.794587
34	38.470165	785.954507	48.140015
35	40.014233	839.996888	50.524161
36	41.570535	896.023778	52.946238
37	43.138737	954.047242	55.405495
38	44.718520	1,014.079019	57.901211
39	46.309585	1,076.130538	60.432698
40	47.911645	1,140.212938	62.999294
41	49.524429	1,206.337076	65.600366
42	51.147678	1,274.513546	68.235304
43	52.781147	1,344.752690	70.903523
44	54.424599	1,417 064608	73.604460
45	56.077812	1,491.459171	76.337572
46	57.740570	1,567.946031	79.102335
47	59.412668	1,646.534631	81.898246
48	61.093909	1,727.234210	84.724819
49	62.784105	1,810.053818	87.581581
50	64.483075	1,895.002318	90.468080

CAPITULO VII

BIO - ESTADISTICA

COEFICIENTES

En el estudio de ciertos aspectos biológicos y sanitarios de una población, o de un grupo de individuos, es importante apreciar la frecuencia de nacimientos, defunciones y casos de determinada enfermedad, en relación a un período dado de tiempo y al número total de sujetos observados. La expresión de la natalidad, mortalidad y morbosidad, denominaciones aplicadas a los fenómenos mencionados, se hace, generalmente, por medio de coeficientes o tasas, algunos de los cuales se utilizan, además, para inquirir sobre procesos relacionados, tales como el crecimiento de una población.

Es necesario tener presente que la veracidad, y, por lo tanto, el significado, de estos coeficientes, depende, fundamentalmente, de la precisión de los datos censales de población y de la manera como se cumplen los requisitos obligatorios de registro e inscripción.

Algunos de los coeficientes más usados, y cuyo cálculo describiremos en este capítulo; * son los siguientes:

A— MORTALIDAD—

Coeficiente general de mortalidad;
Coeficiente específico de mortalidad;
Coeficiente de mortalidad infantil;
Coeficiente de mortinatalidad;

* La mayoría de los ejemplos utilizados en este capítulo están tomados de: (1)— Censo Nacional de Población y Ocupación de 1940. Vol. I. República del Perú, Dirección Nacional de Estadística, A. Arca Parró, Director. Lima, 1940; (2)— Nociones de Bio-Estadística. Franz Schruefer. Biblioteca de la Caja Nacional de Seguro Social, Lima, 1941; y (3)— Boletín Demográfico Municipal de la Ciudad de Lima. Inspección de Estadística y Demografía del Concejo Provincial de Lima. Lima, 1940. Algunos datos demográficos han sido amablemente proporcionados por el Departamento de Estadística, Ministerio de Salud Pública y Asistencia Social, y por la Dirección Nacional de Estadística, Sección Censos.

Coeficiente de morbimortalidad, y
Coeficiente de letalidad.

Coeficiente general de mortalidad corregido o ajustado.

B— NATALIDAD—

Coeficiente general de natalidad, y
Coeficiente específico de natalidad.

C— MORBOSIDAD—

Coeficiente general de morbosidad; y
Coeficiente específico de morbosidad.

D— CRECIMIENTO DE UNA POBLACION—

Crecimiento vegetativo; e
Indice vital.

A—MORTALIDAD—

Coeficiente general de mortalidad (símbolo: qm).—Se calcula relacionando el número de defunciones, por todas las causas, con la población total. Generalmente se refiere a un año de tiempo y se expresa por cada 1,000 o 100,000 habitantes.

Este coeficiente, que corresponde a un índice general de mortalidad, está influenciado por algunas de las cáracterísticas de la población observada (distribución de edad, sexo, raza, ocupación, etc., de sus habitantes), y por condiciones ambientales (clima, altura, etc.). Por consiguiente su significado es bastante limitado cuando se le usa en estudios comparativos de una localidad con otra. En cambio tiene mayor valor en la investigación de la mortalidad en un mismo lugar, en diferentes períodos, ya que generalmente, excluyendo fluctuaciones anormales, las características mencionadas de una población cambian muy lentamente.

El siguiente ejemplo ilustra el cálculo del coeficiente general de mortalidad:

Ejemplo: En el año 1940 ocurrieron un total de 8,517 defunciones en la ciudad de Lima, cuya población en dicho año fué de 520,528 habitantes.

Luego, el coeficiente general de mortalidad, por cada 1,000 habitantes, fué:

$$\frac{1000 \times 8517}{520528} = 16.36 \ ^{0}/_{00}$$

Es decir, que por cada 1,000 habitantes ocurrieron aproximadamente 16 defunciones en la ciudad de Lima, en el año 1940.

Como en determinado año la mortalidad general puede ser influenciada por fenómenos de carácter accidental (epidemia, guerra, etc.), es a veces conveniente calcular el coeficiente general de mortalidad correspondiente a un período de varios años. El siguiente ejemplo ilustra el procedimiento a seguir:

Ejemplo: Las cifras incluídas en el Cuadro 47 corresponden al número de defunciones, por todas las causas, ocurridas en el Perú durante los años 1931 a 1940, y la población del país durante este período de tiempo.

CUADRO 47

POBLACION Y DEFUNCIONES EN EL PERU

Año	Población	Defunciones
1931	5.277,523	70,132
1932	5.371,489	70,664
1933	5.467,968	70,472
1934	5.567,004	74,275
1935	5.668,641	80,876
1936	5.772,923	89,803
1937	5.879,894	94,374
1938	5.989,598	95,817
1939	6.102,979	90,579
1940	6.217,281	85,996
	57.314,500	822,988

El coeficiente general de mortalidad, por cada 1,000 habitantes, correspondiente a este período de 10 años, es:

$$\frac{1000 \times 822988}{57314500} = 14.36 \ ^{0}/_{00}$$

El coeficiente general de mortalidad puede calcularse en referencia a una enfermedad determinada. En este caso se relaciona el número de

defunciones, causadas por tal enfermedad, con la población observada.
El siguiente ejemplo ilustra su cálculo:

Ejemplo: En el año 1939 ocurrieron en la ciudad de Lima 1.304 defunciones a causa de Tuberculosis. La población de la ciudad en ese año fué de 303.630 habitantes.

Luego, el coeficiente general de mortalidad por Tuberculosis, en el año 1939, fué:

$$\frac{1000 \times 1304}{303630} = 4.29 \text{ ‰}$$

Es decir, que por cada 1,000 habitantes, ocurrieron aproximadamente 4 defunciones por Tuberculosis en la ciudad de Lima, en el año 1939.

En el estudio de la mortalidad referente a determinada enfermedad es a veces deseable establecer la relación o proporción que tiene con la mortalidad general, por todas las causas, es decir, averiguar que número de defunciones se debe a tal enfermedad.

Ejemplo: En el año 1938 ocurrieron en la ciudad de Lima un total de 6,065 defunciones; de este número, 1,340 fueron debidas a Tuberculosis.

Luego:

$$\frac{100 \times 1340}{6065} = 22.1 \%$$

Es decir, que de cada 100 defunciones ocurridas en Lima, en el año 1938, aproximadamente 22 fueron debidas a Tuberculosis.

NOTA—Para un cálculo más exacto del coeficiente general de mortalidad Schruefer (Nociones de Bio-Estadística. Biblioteca de la Caja Nacional de Seguro Social. Lima, 1941) indica que es conveniente tomar en cuenta las fluctuaciones que puede sufrir durante el año la población observada, si los datos censales permiten obtener la información necesaria.

Las diversas fórmulas, indicadas por Schruefer, son las siguientes:

(a)—Coeficiente general de mortalidad $= \dfrac{M}{A + 1/2 \quad (N + I + S)}$

en la que M = nº de defunciones en el año;
A = nº de habitantes al inicio del año;
N = nº de nacimientos en el año;
I = nº de inmigrantes en el año;
S = nº de emigrantes en el año.

Así, por ejemplo, aplicando esta fórmula a la ciudad de Lima, cuyos datos, en el año 1935, fueron:

M = 5,722
A = 281,350
N = 9,199
I = 0 (movimiento migratorio desconocido)
S = 0

Luego el coeficiente general de mortalidad por cada 1,000 habitantes, de la ciudad de Lima, en el año 1935, sería:

$$\frac{1000 \times 5722}{281350 + \frac{9199}{2}} = 20.01 \ ^{0}/_{00}$$

(b)—Cuando no existen datos referentes al movimiento migratorio ni al número de nacimientos ocurridos durante el año, pero en cambio se conoce el número de habitantes al inicio y al fin del año, el cálculo se hace mediante la fórmula:

$$\text{Coeficiente general de mortalidad} = \frac{M}{1/2\,(A + B + M)}$$

en la que M = nº de defunciones en el año;
A = nº de habitantes al inicio del año;
B = nº de habitantes al fin del año.

Así, por ejemplo, aplicando esta fórmula a la ciudad de Lima, cuyos datos, en el año 1935, fueron:

M = 5,722
A = 281,350
B = 284,827

El coeficiente general de mortalidad, por cada 1,000 habitantes, de la ciudad de Lima, en el año 1935, sería:

$$\frac{1000 \times 5722}{\dfrac{281350 + 284827 + 5722}{2}} = 20.01 \text{ ‰}$$

(c)—Cuando se determina el coeficiente general de mortalidad correspondiente a un período de varios años, y se conoce el número de habitantes al inicio y al fin de dicho período, los cálculos se hacen de la manera como está ilustrada en el siguiente ejemplo:

Se trata de determinar el coeficiente general de mortalidad de la ciudad de Lima durante el período de 3 años: 1933, 1934 y 1935.

Los datos de población y defunciones son los siguientes:

A	= nº de habitantes al inicio de 1933 =	276,315
B	= nº de habitantes al fin de 1933 — inicio de 1934 =	278,438
C	= nº de habitantes al fin de 1934 — inicio de 1935 =	281,350
D	= nº de habitantes al fin de 1935 =	284,827

M1 = nº de defunciones en 1933 = 5,833
M2 = nº de defunciones en 1934 — 5,872
M3 = nº de defunciones en 1935 = 5,722

En seguida, se aplica la siguiente fórmula:

$$\text{Coeficiente general de mortalidad} = \frac{M1 + M2 + M3}{\dfrac{A}{2} + B + C + \dfrac{D}{2} + \left(\dfrac{M1 + M2 + M3}{2}\right)}$$

Remplazando, el coeficiente general de mortalidad, por cada 1,000 habitantes, de la ciudad de Lima, en el período 1933, 1934 y 1935, sería:

$$\frac{1000 \times (5833 + 5872 + 5722)}{\dfrac{276315}{2} + 278438 + 281350 + \dfrac{284827}{2} + \left(\dfrac{5833 + 5872 + 5722}{2}\right)} = 20.52 \text{ ‰}$$

(d)—El ceoficiente general de mortalidad puede también calcularse para un período comprendido entre dos censos de la población. En este caso la fórmula que se aplica es la siguiente:

$$\text{Coeficiente general de mortalidad} = \frac{\dfrac{M}{n}}{1/2\left(A + B + \dfrac{M}{n}\right)}$$

en la que M = n⁰ de defunciones en el período entre los dos censos;
n = n⁰ de años entre los dos censos;
A = n⁰ de habitantes según el primer censo;
B = n⁰ de habitantes según el segundo censo;

Así, por ejemplo, en una población X un censo verificado en 1930 dió 281,750 habitantes, y un segundo censo realizado en 1940 dió 343,446 habitantes. En este período de 10 años, transcurridos entre los dos censos, ocurrieron 62,342 defunciones.

Luego, el coeficiente general de mortalidad, por cada 1,000 habitantes, para el período de 10 años entre los dos censos, sería:

$$\frac{1000 \times \dfrac{62342}{10}}{\dfrac{281750 + 343446 + \dfrac{62342}{10}}{2}} = 19.75 \ ^{o}/_{oo}$$

Coeficiente específico de mortalidad.—Es un índice de mortalidad correspondiente a un grupo de la población, seleccionado por edad, sexo o ambas características. Generalmente se expresa por cada 1,000 habitantes del grupo seleccionado y se refiere a un año de tiempo.

Los coeficientes específicos de mortalidad tienen un elevado significado en la estadística demográfica, pues toman en cuenta la influencia decisiva que tiene el factor edad sobre la mortalidad, y proporcionan una medida de las probabilidades de ocurrir, en un tiempo dado, cierto número de defunciones en determinada clase de habitantes.

Cuando se estudia, comparativamente, el grado de mortalidad en dos poblaciones diferentes es conveniente realizar tal estudio por medio de la determinación de los coeficientes específicos para los diversos grupos de habitantes. Estos grupos, con excepción de aquel que corresponde al primer año de vida, se forman con sujetos cuya edad difiere en 5 o 10 años. Aunque es costumbre calcular los coeficientes específicos de mortalidad solamente en relación con la edad o sexo, o ambos factores, la especificidad puede también incluír el concepto de raza, ocupación, clase de vivienda (urbana o rural), etc., pues todos estos factores tienen también influencia sobre el grado de mortalidad. Puede, por ejemplo, en el estudio de una población, compararse el grado de mortalidad en habitantes de 10 a 19 años de edad, separados en dos grupos: (a)—

sujetos de raza blanca, y (b)—sujetos de raza india; o en grupos de habitantes de 20 a 25 años de edad. clasificados según su ocupación.

El siguiente ejemplo ilustra el cálculo del coeficiente específico de mortalidad correspondiente a un grupo de habitantes de determinada edad:

Ejemplo: En el año 1940 ocurrieron en la ciudad de Lima 804 defunciones en habitantes de 20 a 29 años de edad. Según los datos censales, el número total de habitantes de esa edad alcanzó en Lima, en el mismo año, la cifra de 113,501.

Luego, el coeficiente específico de mortalidad, por cada 1,000 habitantes de 20 a 29 años de edad, fué, en la ciudad de Lima y en el año 1940:

$$\frac{1000 \times 804}{113501} = 7.08\ {}^{o}/oo$$

Es decir, que de cada 1,000 habitantes de 20 a 29 años de edad, aproximadamente 7 fallecieron en la ciudad de Lima, en el año 1940.

El coeficiente específico de mortalidad alcanza su valor máximo en el primer año de vida; en seguida desciende, bruscamente, presentando sus valores más bajos entre los 10 y 19 años de edad, ascendiendo después, progresivamente, durante el período adulto y de ancianidad. El siguiente ejemplo ilustra estas fluctuaciones, e indica la marcada influencia que tiene el factor edad sobre el grado de mortalidad.

Ejemplo: En el cuadro 48 están consignados los datos referentes a la distribución de la población de Lima por grupos de edades en el año 1940. Igualmente está incluído el número de defunciones ocurridas durante ese año en cada uno de los grupos.
Los respectivos coeficientes específicos de mortalidad han sido calculados utilizando los datos anteriores, y están dados en la última columna.

La mortalidad, elevada para el primer año de vida, desciende, hasta alcanzar su nivel más bajo en la edad 10 — 14 años; sube, en seguida alcanzando, nuevamente, un valor elevado sobre los 60 años de edad.

CUADRO 48

POBLACION Y DEFUNCIONES DISTRIBUIDAS SEGUN EDAD

CIUDAD DE LIMA — AÑO 1940

Edad (años)	Nº de habitantes	Nº de defunciones	Coeficiente específico de mortalidad **por** cada 1,000 habitantes
Menos de 1	13,605	2,083	153.11
1 — 9	97,828	1,749	17.88
10 — 14	58,274	249	4.27
15 — 19	63,295	411	6.49
20 — 29	113,501	804	7.08
30 — 39	75,296	544	7.22
40 — 49	47,779	559	11.70
50 — 59	27,779	606	21.82
60 y más	23,171	1,464	63.18

El siguiente ejemplo se refiere al cálculo del coeficiente específico de mortalidad en dos grupos de habitantes, de igual edad pero de diferente sexo:

Ejemplo: En el año 1940 ocurrieron en la ciudad de Lima 373 defunciones en hombres de 20 a 29 años de edad, y 431 defunciones en mujeres de la misma edad.

El número de habitantes correspondiente a estos grupos de edad fué: Hombres: 58,819; Mujeres: 54,682.

Luego, el coeficiente específico de mortalidad por cada 1,000 habitantes de 20 a 29 años de edad fué:

$$\text{En hombres: } \frac{1000 \times 373}{58819} = 6.34 \ {}^{0}/_{00}$$

$$\text{En mujeres: } \frac{1000 \times 431}{54682} = 7.88 \ {}^{0}/_{00}$$

El coeficiente específico de mortalidad puede referirse a determinada enfermedad, tal como está ilustrado en el siguiente ejemplo:

Ejemplo: En el año 1940 ocurrieron en la ciudad de Lima 435 defunciones por Tuberculosis en habitantes de 20 a 29 años de edad.

En dicho año, el número total de habitantes de esa edad, fué 113,501 en la ciudad de Lima.

Luego, el coeficiente específico de mortalidad por Tuberculosis, en la ciudad de Lima, por cada 1,000 habitantes de 20 a 29 años de edad, fué en el año 1940:

$$\frac{1000 \times 435}{113501} = 3.83 \ ^{0}/_{00}$$

El cálculo de la relación entre el número de defunciones ocurridas en una edad dada y el número total de defunciones, en todas las edades, proporciona información importante en el estudio de las características de la mortalidad en una población, especialmente cuando se comparan las observaciones hechas cada cierto tiempo. La proporción se expresa generalmente en porcentaje y en referencia a un año de tiempo.

Ejemplo: En el año 1940 ocurrieron en la ciudad de Lima 1,749 defunciones en habitantes de 1 a 9 años de edad. El número total de defunciones, en todas las edades, alcanzó en ese año, en la ciudad de Lima, la cifra de 8,517.

Luego:

$$\frac{100 \times 1749}{8517} = 20.5 \ \%$$

Es decir, que de cada 100 defunciones aproximadamente 21 ocurrieron en habitantes de 1 a 9 años de edad.

Coeficiente de mortalidad infantil.—Se calcula relacionando el número de defunciones ocurridas en niños menores de 1 año de edad con el número de nacimientos. Los nacidos muertos no son considerados en el cálculo. Generalmente se refiere a un año de tiempo y se expresa por cada 1,000 nacimientos (vivos).

El cálculo del coeficiente de mortalidad infantil se hace en referencia al número de nacimientos, y no al número de habitantes menores de 1 año, por ser difícil la estimación de la población de esta edad. En cambio, el número de nacimientos indica, con aproximación, el número de niños expuestos a morir durante el próximo año de vida. Esta última suposición no es estrictamente cierta, pues si se calcula el coeficiente de mortalidad infantil en un año dado, por ejemplo en 1940, resulta que los niños nacidos en Noviembre o Diciembre sólo están expuestos a morir durante unas pocas semanas de ese año. Sin embargo, como este es un error constante, que se repite de año en año, no influencia significativamente el estudio comparativo del coeficiente obtenido en años sucesivos o separados.

El coeficiente de mortalidad infantil es un índice sanitario importante.

El adelanto en la salubridad de una población está generalmente acompañado por una disminución en este coeficiente.

El cálculo del coeficiente de mortalidad infantil está ilustrado con el siguiente ejemplo:

Ejemplo: En el año 1940 ocurrieron en la ciudad de Lima, 2,083 defunciones en niños menores a 1 año de edad. En el mismo año tuvieron lugar 15,328 nacimientos (vivos).

Luego, el coeficiente de mortalidad infantil, por cada 1,000 nacimientos, fué:

$$\frac{1000 \times 2083}{15328} = 13.59 \ ^{0}/_{00}$$

Es decir, que de cada 1,000 niños nacidos, aproximadamente 14 murieron durante el primer año de vida.

En el Cuadro 49 están consignados el número de nacimientos y el número de defunciones de niños menores a 1 año de edad, ocurridos en la ciudad de Lima en los años comprendidos en el período 1936—1940. Los respectivos coeficientes de mortalidad infantil están dados en la última columna.

CUADRO 49

MORTALIDAD INFANTIL. CIUDAD DE LIMA 1936 - 1940

Año	Nº de nacimientos	Nº de defunciones (niños menores de 1 año)	Coeficiente de mortalidad infantil (por cada 1,000 nacimientos)
1936	9,653	1,550	16.01
1937	10,414	1,462	14.04
1938	10,858	1,298	11.87
1939	11,642	1,246	10.70
1940	15,328	2,083	13.59

La mortalidad infantil puede también ser apreciada en relación al número total de defunciones ocurridas en todas las edades durante el año de observación. Esta proporción es generalmente expresada en porcentaje.

Ejemplo: En el año 1940 ocurrieron en Lima 8,517 defunciones en toda la población. De este número 2,083 correspondieron a niños menores de 1 año de edad.

Luego:

$$\frac{100 \times 2083}{8517} = 24.4 \ \%$$

Es decir, que de cada 100 defunciones ocurridas en la ciudad de Lima en 1940 aproximadamente 24 correspondieron a niños menores a 1 año de edad.

Coeficiente de mortinatalidad (o natimortalidad).—Los nacidos muertos no son incluídos en el cálculo de los coeficientes de mortalidad y natalidad. Son considerados aparte, en el llamado coeficiente de mortinatalidad (o natimortalidad), que, expresado en porcentaje, relaciona el número de nacidos muertos con el número total de nacimientos en un período dado de tiempo, el que generalmente es de un año.

Ejemplo: En el año 1940 se registraron en la ciudad de Lima 15,328 nacimientos (vivos) y 628 nacidos muertos.

Relacionando el número de nacidos muertos con el número total de nacimientos (15,328 + 628 = 15,956) tenemos:

$$\frac{100 \times 628}{15956} = 3.9 \ \%$$

Es decir, que de cada 100 nacimientos ocurridos en la ciudad de Lima en el año 1940 aproximadamente 4 correspondieron a nacidos muertos.

Coeficiente de morbimortalidad.—Este coeficiente indica la relación entre el número de defunciones y el número total de enfermos. Se refiere habitualmente a un año de tiempo y se expresa por cada 100 o 1,000 enfermos. Es difícil, si no imposible, conocer el número total de enfermos en una población, ya que la mayoría de las enfermedades no se registran obligatoriamente; por consiguiente este coeficiente es generalmente calculado en determinados grupos de sujetos enfermos, tales como los hospitalizados, en quienes tal apreciación es posible. El coeficiente de morbimortalidad hospitalario es más alto que el corresponde a la población total, pues los enfermos hospitalizados adolescen por lo general, de enfermedad de cierta gravedad.

Ejemplo: En el año 1935 se asistieron en el Hospital Dos de Mayo, de la ciudad de Lima, un total de 10,078 enfermos. En el mismo año ocurrieron en ese Hospital un total de 1,263 defunciones.

Luego, el coeficiente de morbimortalidad, por cada 1,000 enfermos, fué:

$$\frac{1000 \times 1263}{10078} = 125.32 \ ^{o}/oo.$$

Es decir, que de cada 1,000 enfermos hospitalizados en el Hospital Dos de Mayo, en 1935, fallecieron 125, aproximadamente.

El coeficiente de morbimortalidad puede calcularse en referencia a una enfermedad determinada, relacionando el número de defunciones, a consecuencia de esta enfermedad con el número total de enfermos (por todas las causas).

Ejemplo: En el año 1935 ocurrieron 513 defunciones por Tuberculosis en el Hospital Dos de Mayo de la ciudad de Lima. En el mismo año se asistieron en este Hospital un total de 10,078 enfermos.

Luego, el coeficiente de morbimortalidad hospitalario por Tuberculosis, por cada 1,000 enfermos, fué:

$$\frac{1000 \times 513}{10078} = 50.90 \ ^{o}/oo$$

Es decir, que de cada 1,000 enfermos asistidos en el Hospital Dos de Mayo, en el año 1935, aproximadamente 51 fallecieron a consecuencia de Tuberculosis.

Coeficiente de letalidad—Este coeficiente relaciona el número de defunciones a causa de una enfermedad determinada con el número total de enfermos a consecuencia de la misma enfermedad. Se refiere, por lo general, a un año de tiempo y se expresa en porcentaje.

Si se desea, este coeficiente puede hacerse específico, agrupando los casos según sus características de edad,, sexo raza, ocupación, localidad de la vivienda (urbana o rural), etc.

El coeficiente de letalidad, teóricamente de importante significado, y que interesa particularmente al clínico, no puede calcularse satisfactoriamente en una población total, por carecerse de registros que indiquen el número de casos que ocurren por tal o cual enfermedad. Aún, tratándose de algunas enfermedades infecto-contagiosas, cuyo registro es obligatorio, no se cumple adecuadamente esta prescripción. Por esta razón el coeficiente de letalidad es generalmente aplicado a estudios estadísticos hospitalarios, o en relación a determinadas enfermedades que pueden ocurrir en grupos de individuos sujetos a un control especial (cuarteles, colegios, etc.).

Los coeficientes de letalidad hospitalarios no reflejan satisfactoriamente la situación en la población total. Los enfermos hospitalizados son sujetos en quienes la enfermedad ha adquirido por lo general, cierta gravedad. Además, el tratamiento hospitalario puede influenciar, significativamente, la evolución de la enfermedad, en contraste con lo que ocurre en el enfermo tratado en su hogar o en condiciones desfavorables.

El siguiente ejemplo ilustra el cálculo del coeficiente de letalidad:

Ejemplo: En el año 1935 se asistieron en el Hospital Dos de Mayo de la ciudad de Lima un total de 1,455 enfermos de Tuberculosis; de este número, 513 fallecieron.

Luego, el coeficiente de letalidad (hospitalario) para esta enfermedad, fué:

$$\frac{100 \times 513}{1455} = 35.3 \ \%$$

Es decir, que de cada 100 enfermos hospitalizados por Tuberculosis en el año 1935, en el Hospital Dos de Mayo, fallecieron 35, aproximadamente.

Coeficiente general de mortalidad corregido o ajustado

Hemos visto, en párrafos anteriores, que el factor edad influencia, notablemente, el índice de mortalidad. Este hecho hace conveniente que en la comparación de coeficientes generales de mortalidad, correspondientes a dos o más localidades, se tome en cuenta la distribución de las respectivas poblaciones por edades. Puede, por ejemplo, una localidad A tener un coeficiente general de mortalidad más elevado que otra localidad B, revelando, aparentemente, condiciones sanitarias más favorables en B. Sin embargo, un estudio de la distribución en grupos de edades de las respectivas poblaciones muestra que la población A tiene una proporción mayor de habitantes de edad avanzada, lo que naturalmente influencia en elevar su coeficiente general de mortalidad. Ambas localidades no son, pues, estrictamente comparables desde este punto de vista.

Por estas razones, es conveniente, en el estudio comparativo de coeficientes generales de mortalidad correspondients a dos o más localidades, corregir o ajustar, estos coeficientes en relación a la distribución por edad de las respectivas poblaciones. Una de las maneras de efectuar esta corrección, o ajustamiento, es relacionar los coeficientes específicos de mortalidad (coeficientes correspondientes a diferentes grupos de habitantes separados según su edad) de las localidades comparadas a una población tipo o standard, la que puede ser la correspondiente al país entero, y en seguida calcular, con los datos obtenidos, los respectivos coeficientes generales de mortalidad. Estos coeficientes expresan, entonces, el índice de mortalidad que correspondería a dichas localidades en el supuesto que ambas poblaciones tengan una idéntica distribución en edad. Se elimina, así, la influencia que pueda tener en los coeficientes generales de mortalidad una distribución desigual en edad de las poblaciones.

Ilustraremos con un ejemplo los procedimientos que se emplean en la corrección, o ajustamiento, de coeficientes generales de mortalidad.

Ejemplo: Se trata de comparar los coeficientes generales de mortalidad obtenidos en las ciudades de Lima y Arequipa en el año 1940.

En Lima, con una población de 520,528 habitantes, ocurrieron 8,469 defunciones en 1940.
Luego, el coeficiente general de mortalidad, por cada 1,000 habitantes, fué:

$$\frac{1000 \times 8469}{520528} = 16.27 \ ^{0}/_{00}$$

En Arequipa, con una población de 79,185 habitantes, ocurrieron 1,063 defunciones en el mismo año.
Luego, el coeficiente general de mortalidad, por cada 1,000 habitantes, fué:

$$\frac{1000 \times 1063}{79185} = 13.42\ ^{0}/_{00}$$

Aparentemente, la mortalidad general es pues menor en Arequipa que en Lima (16.27 — 13.42 = 2.85 $^{0}/_{00}$). Sin embargo, para una comparación más precisa, vamos a corregir, o ajustar, dichos coeficientes en relación a la distribución por edades de las respectivas poblaciones.

Para esta corrección, o ajustamiento, se procede de la manera siguiente:

(a)—En primer lugar, es necesario tomar en cuenta, la distribución de las dos poblaciones por grupos de edades, y el número de defunciones ocurridas en cada grupo, en el año 1940.

El Cuadro 50 consigna dichos datos, obtenidos en los estudios censales y estadísticas municipales correspondientes a dicho año.

CUADRO 50

POBLACION Y DEFUNCIONES DISTRIBUIDAS SEGUN EDAD.

CIUDADES DE LIMA Y AREQUIPA — AÑO 1940

	Lima		Arequipa	
Edad (años)	Número de habitantes	Número de defunciones	Número de habitantes	Número de defunciones
Menos de 1	13,605	2,083	2,550	236
1 — 9	97,828	1,749	18,235	265
10 — 14	58,274	249	10,128	29
15 — 19	63,295	411	9,463	38
20 — 29	113,501	804	14,673	99
30 — 39	75,296	544	8.924	50
40 — 49	47.779	559	6,351	50
50 — 59	27,779	606	3,928	47
60 y más	23,171	1,464	4,933	249

(b)—En seguida, con los datos del Cuadro 50, y de acuerdo con los procedimientos ya descritos (Coeficiente específico de mortalidad) se calculan los coeficientes específicos de mortalidad correspondientes a cada uno de los grupos de habitantes clasificados según edad.

Los resultados obtenidos están dados en el Cuadro 51.

CUADRO 51

CIUDADES DE LIMA Y AREQUIPA — AÑO 1940.

	Coeficiente específico de mortalidad por cada 1,000 habitantes	
Edad (años)	Lima	Arequipa
Menos de 1	153.11	92.55
1 — 9	17.88	14.53
10 — 14	4.27	2.86
15 — 19	6.49	4.02
20 — 29	7.08	6.75
30 — 39	7.22	5.60
40 — 49	11.70	7.67
50 — 59	21.82	11.97
60 y más	63.18	50.48

(c)—Se toma, en seguida, como referencia, la distribución en grupos de edades de una población tipo o standard de 1,000,000 de habitantes. Esta población puede ser la que corresponde al país entero.

En el Cuadro 52 está consignada la distribución por edades de 1.000.000 de habitantes del Perú. *

CUADRO 52

DISTRIBUCION POR EDAD DE 1,000,000 DE HABITANTES DEL PERU

Edad (años)	Número de habitantes
Menos de 1	35,366
1 — 9	268,225
10 — 14	117,318
15 — 19	94,885
20 — 29	163,133
30 — 39	122,149
40 — 49	83,768
50 — 59	51,277
60 y más	63,879

* Para este cálculo se ha utilizado los datos contenidos en el Cuadro Nº 7 del Censo Nacional de Población y Ocupación de 1940. Vol. I (página 67). República del Perú, Dirección Nacional de Estadística, A. Arca Parró, Director. Lima, 1940, que corresponde a la población total del Perú distribuída por grupos de edades.

Así, eliminando el número de habitantes con edad no declarada, tenemos que en una población total de 6.205.997 hay 219,483 habitantes menores de 1 año de edad. Luego en 1.000,000 hay:

$$\frac{1000000 \ \text{X} \ 219483}{6205997} = 35666$$

De la misma manera se ha calculado el número de habitantes por cada 1.000,000 que corresponde a los otros grupos de edad.

(d)—En seguida, los coeficientes específicos de mortalidad, obtenidos en las ciudades de Lima y Arequipa (Cuadro 51) se aplican a los respectivos grupos de edades de la población tipo o standard (Cuadro 52), para obtener el número de fallecimientos que ocurriría en esta población si sus coeficientes específicos de mortalidad fueran: (1)— iguales a los hallados en la ciudad de Lima y (2) iguales a los hallados en la ciudad de Arequipa.

Los cálculos correspondientes están dados en el Cuadro 53.

CUADRO 53

Edad (años)	Lima	Arequipa
Menos de 1	153.11 X 35.366 = 5414.9	92.55 X 35.366 = 3273.1
1 — 9	17.88 X 268.225 = 4795.9	14.53 X 268.225 = 3897.3
10 — 14	4.27 X 117.318 = 500.9	2.86 X 117.318 = 325.5
15 — 19	6.49 X 94.885 = 615.8	4.02 X 94.885 = 381.4
20 — 29	7.08 X 163.133 = 1155.0	6.75 X 161.133 = 1101.1
30 — 39	7.22 X 122.149 = 881.9	5.60 X 122.149 = 684.0
40 — 49	11.70 X 83.768 = 980.1	7.87 X 83.768 = 659.3
50 — 59	21.82 X 51.277 = 1118.9	11.97 X 51.277 = 613.8
60 y más	63.18 X 63.879 = 4035.9	50.48 X 63.879 = 3224.6
	19499.3	14170.1

NOTA—El número de habitantes en cada grupo de edad (que ha sido tomado del Cuadro 52) ha sido previamente dividido entre 1,000, pues los coeficientes específicos de mortalidad de las ciudades de Lima y Arequipa corresponden a 1,000 habitantes.

(e)—Finalmente, los coeficientes generales de mortalidad, corregidos, o ajustados, por cada 1,0000 habitantes, para las ciudades de Lima y Arequipa, son:

$$\text{Lima:} \quad \frac{1000 \times 19499}{1000000} = 19.50 \ ^0/_{00}.$$

$$\text{Arequipa:} \quad \frac{1000 \times 14170}{1000000} = 14.17 \ ^0/_{00}$$

Interpretación: El estudio comparativo de los coeficientes generales de mortalidad, por cada 1,000 habitantes, correspondientes a las ciudades de Lima y Arequipa, en el año 1940 ,proporciona los siguientes datos:

	Coeficiente general de mortalidad por cada 1,000 habitantes
Lima:	16.27 $^{o}/_{oo}$
Arequipa:	13.42 $^{o}/_{oo}$
Diferencia:	2.85 $^{o}/_{oo}$

Proporción: $\dfrac{16.27}{13.42} = 1.21$

Utilizando los procedimientos que acaban de ser descritos dichos coeficientes han sido corregidos, o ajustados, obteniéndose los siguientes datos:

	Coeficiente general de mortalidad, corregido o ajustado, por cada 1,000 habitantes
Lima	19.50 $^{o}/_{oo}$
Arequipa:	14.17 $^{o}/_{oo}$
Diferencia:	5.33 $^{o}/_{oo}$

Proporción: $\dfrac{19.50}{14.17} = 1.38$

En resumen, la comparación directa de los coeficientes generales de mortalidad correspondientes a las ciudades de Lima y Arequipa, en el año 1940, revela un índice de mortalidad más bajo en Arequipa.
Si ambas poblaciones tuvieran una idéntica distribución por edad de sus habitantes, se acentuaría la diferencia, que a favor de Arequipa, tiene el índice de mortalidad.

B— NATALIDAD—

Coeficiente general de natalidad.—Indica la relación entre el número de nacimientos (excluyendo los nacidos muertos) y el número total de habitantes. Generalmente se refiere a un año de tiempo y se expresa por cada 1,000 habitantes.

Este coeficiente no proporciona una medida exacta de la capacidad reproductiva de una población, puesto que en su cálculo están incluídos habitantes en quienes no existe tal capacidad. Debido a la influencia que sobre su valor tiene la distribución por edad y sexo de la población observada, este coeficiente debe usarse con cierta reserva en estudios comparativos de dos o más poblaciones, que pueden ser diferentes en lo que respecta a tales características de sus habitantes.

En cambio, como una población determinada cambia muy lentamente su composición, en referencia a la edad y sexo, el coeficiente general de natalidad puede indicar con bastante fidelidad, cuando es calculado en años sucesivos, la evolución de la natalidad en una misma población. El cálculo de este coeficiente está ilustrado con el siguiente ejemplo:

Ejemplo: Según el censo realizado en el año 1940, la población del Perú fué calculada en ese año en 6.217,381 habitantes. En ese mismo año ocurrieron en el país un total de 167,290 nacimientos.

Luego, el coeficiente general de natalidad, por cada 1,000 habitantes, fué:

$$\frac{1000 \times 167290}{6217381} = 26.91\ ^{0}/_{00}$$

Es decir, que por cada 1,000 habitantes en el Perú, en el año 1940, ocurrieron, aproximadamente, 27 nacimientos.

El coeficiente general de natalidad, por cada 1,000 habitantes, puede también calcularse para un período de varios años, como se puede apreciar en el siguiente ejemplo:

Ejemplo: Los datos consignados en el Cuadro 54 corresponden a la población total y al número de nacimientos ocurridos en el Perú, durante los años comprendidos en el período 1931 - 1940.

CUADRO 54

NUMERO DE HABITANTES Y NACIMIENTOS EN EL PERU. PERIODO 1931 - 1940

Año	Número de habitantes	Número de nacimientos
1931	5.277,523	165,074
1932	5.371,489	166,636
1933	5.467,968	168,962
1934	5.567,004	174,731
1935	5.668,641	179,863
1936	5.772,923	182,981
1937	5.879,894	187,134
1938	5.989,598	187,996
1939	6.102,079	177,664
1940	6.217,381	167,290
	57.314,500	1.758,331

El coeficiente general de natalidad, por cada 1,000 habitantes, correspondiente a este período de 10 años, es:

$$\frac{1000 \times 1758331}{57314500} = 30.68 \ ^{o}/_{oo}$$

Coeficiente específico (o natural) de natalidad.—Se refiere al número de nacimientos (excluyendo nacidos muertos) ocurridos durante un período dado, que es generalmente de un año, en relación con el número de habitantes mujeres entre 15 y 44 años de edad, período en el que se les considera aptas para ser madres. * Este coeficiente se expresa por cada 100 o 1,000 mujeres.

Ejemplo: En el año 1940 ocurrieron en el Perú 167,290 nacimientos (excluyendo nacidos muertos). El número de habitantes mujeres, entre 15 y 44 años de edad, fué, en ese año: 1.334,065.

Luego, el coeficiente específico o natural de natalidad, por cada 1,000 mujeres, fué:

$$\frac{1000 \times 167290}{1334065} = 125.40 \ ^{o}/_{oo}$$

Es decir, por cada 1,000 mujeres, aptas para ser madres, ocurrieron en el Perú, en el año 1940, aproximadamente 125 nacimientos.

Con frecuencia se interpreta el coeficiente específico (o natural) de natalidad como un índice de fertilidad o fecundidad de una población. Sin embargo, este criterio es más veraz si se deriva de la relación entre el número de nacimientos legítimos y el número de mujeres casadas, cuya edad las hace aptas para ser madres (entre 15 y 44 años).

Ejemplo: En el año 1935 ocurrieron en Lima 4,702 nacimientos legítimos. En dicho año el número de habitantes mujeres, casadas de 15 a 44 años de edad, fué de 28,756.

Luego:

$$\frac{1000 \times 4702}{28756} = 163.51 \ ^{o}/_{oo}$$

* Algunos autores consideran 15 y 50 años de edad como límites de este período.

Es decir, que de cada 1,000 mujeres casadas, en edad de poder ser madres, aproximadamente 164 contribuyeron a la natalidad en 1935, en la ciudad de Lima, representando esta cifra un índice de fertilidad.

Este índice de fertilidad puede calcularse en diferentes edades, si se conoce el número de mujeres casadas en los diferentes períodos de edad, y la edad de las madres al tiempo del nacimiento.

Ejemplo*: El Cuadro 55 contiene el número de mujeres casadas, divididas en diferentes grupos de edad, y el número de estas mujeres que tuvieron un hijo, en el año 1911, en Australia.

En la última columna se ha calculado el coeficiente específico de natalidad (o fertilidad en este caso), por cada 1,000 mujeres, en los diferentes grupos de edad.

El Cuadro 55 muestra que el índice de fertilidad en mujeres casadas desciende a medida que su edad se hace mayor.

CUADRO 55

DATOS CORRESPONDIENTES A AUSTRALIA. AÑO 1911

Edad (años)	*Número de mujeres casadas*	*Número de mujeres casadas que tuvieron un hijo durante el año*	*Coeficiente específico de natalidad (o fertilidad) por cada 1,000 mujeres*
19 o menos	8,716	4,146	476
20 — 24	65,956	25,957	394
25 — 29	110,591	33,817	306
30 — 34	113,310	25,682	227
35 — 39	105,550	16,839	160
40 — 44	95,573	6,763	71
45 y más	82,933	713	9

C— MORBOSIDAD—

Coeficiente general de morbosidad.—Relaciona el número de enfermos (por todas las causas) con el número total de habitantes, en un tiempo dado, el que es generalmente de un año. Se expresa por cada 1,000 habitantes, aunque a veces es también expresado por cada 10,000 o

* Tomado de Introduction to Medical Biometry and Statistics. R. Pearl — W. B. Saunders Co., Philadelphia, 1940.

100,000 habitantes. Es difícil, sino imposible, calcular este coeficiente en referencia a la población de un país o ciudad, por la falta de datos demográficos que registren la incidencia de todas las enfermedades. Por consiguiente, los coeficientes de morbosidad se refieren, casi siempre, a pequeños núcleos de habitantes, tales como obreros de un centro industrial, escolares, miembros de institutos armados, etc., que son individuos que pueden estar sujetos a un control médico más o menos ríg.do. Basándose en estudios estadísticos relacionados con el Seguro Social, establecido en distintos países, Schruefer * señala que de la población asegurada enferman anualmente un 35 a 40%; de estos una quinta parte, o sea más o menos un 7 % de la población asegurada, requiere tratamiento hospitalario, lo que significa la necesidad de 4 a 6 camas por cada 1,000 habitantes en regiones urbanas y 3 camas por cada 1,000 habitantes en zonas rurales.

Los coeficientes de morbosidad tienen una destacada importancia desde un punto de vista sanitario, pues los coeficientes de mortalidad, que cuentan con un registro demográfico que permite su cálculo más o menos preciso, no siempre reflejan adecuadamente los resultados obtenidos en actividades sanitarias o la necesidad de su desarrollo. Muchas enfermedades, de evidente significado sanitario, tienen un índice bajo de mortalidad.

El estudio comparativo de coeficientes generales de morbosidad, determinados en diversos núcleos de sujetos, debe tomar en cuenta la edad, sexo, raza, ocupación, vivienda, características ambientales, etc., en cada núcleo, pues todos estos factores influyen marcadamente en tal coeficiente.

El ejemplo que sigue se refiere al cálculo del coeficiente general de morbosidad en un grupo seleccionado de sujetos.

Ejemplo: En el año 1943 se registraron 406 enfermos, por diversas causas, en un centro industrial constituído por 1,125 obreros.

Luego, el coeficiente general de morbosidad, por cada 1,000 individuos obreros de este centro, fué:

$$\frac{1000 \times 406}{1125} = 360.89 \text{ ‰}$$

Es decir, que de cada 1.000 obreros de este centro industrial, enfermaron aproximadamente 361 en dicho año.

* Nociones de Bio-Estadística. Franz Schruefer. Biblioteca de la Caja Nacional del Seguro Social, Lima, 1941.

El coeficiente general de morbosidad puede referirse al número de enfermos por una causa determinada. Su cálculo se hace, casi siempre, en grupos pequeños de sujetos, por la falta de datos demográficos que registren la incidencia de enfermedades en la población de un país o ciudad. Aún en el caso de ciertas enfermedades infecto-contagiosas, cuyo registro es obligatorio, las disposiciones legales no se cumplen satisfactoriamente.

Ejemplo: En el año 1939 se asistieron en el Hospital Dos de Mayo de la ciudad de Lima un total de 9,375 sujetos. De este número, 122 correspondieron a Fiebre Tifoidea.

Luego, el coeficiente general de morbosidad por Fiebre Tifoidea, por 1,000 asistidos en el Hospital Dos de Mayo, en el año 1939, fué:

$$\frac{1000 \times 122}{9375} = 13.01 \text{ }^{o}/_{oo}$$

Es decir, que de cada 1,000 asistidos en el Hospital Dos de Mayo, en 1935, aproximadamente 13 lo fueron debido a Fiebre Tifoidea.

Coeficiente específico de morbosidad.—Indica la relación entre el número de enfermos, por todas las causas, o por una causa determinada, en un grupo de habitantes seleccionados por edad, sexo, o ambas características. Se refiere generalmente a un año de tiempo y se expresa por cada 1,000, 10,000 o 100,000 habitantes.

*Ejemplo**: En el año 1928 ocurrieron en California 73 casos de Poliomielitis en niños de 1 a 4 años de edad. El número de habitantes de esa edad era, en dicho estado y en el mismo año: 292,000.

Luego, el coeficiente específico de morbosidad por Poliomielitis, por cada 100,0000 niños de 1 a 4 años de edad, fué:

$$\frac{100000 \times 73}{292000} = 25.00 \text{ }^{o}/_{oooo}$$

Es decir, que por cada 100,000 niños de 1 a 4 años de edad, ocurrieron 25 casos de Poliomielitis en California, en el año 1928.

* Tomado de Introduction to Medical Biometry and Statistic. R. Pearl—W. B. Saunders Co., Philadelphia, 1940.

Puede ser interesante, en determinado caso, aumentar aún más la especificidad del coeficiente de morbosidad, calculándolo en grupos de sujetos divididos no solamente según edad y sexo, sino también en referencia a su raza, ocupación, lugar de residencia (urbana o rural), etc.

D— CRECIMIENTO DE UNA POBLACION—

Crecimiento vegetativo.—El crecimiento vegetativo de una población se aprecia relacionando el número de nacimientos con el número de defunciones ocurridos durante un período dado, que puede ser uno o varios años.

Si se resta el número de defunciones del número de nacimientos, en el período considerado, el resultado expresa el crecimiento vegetativo en cifras absolutas. Si se resta el coeficiente general de mortalidad del coeficiente general de natalidad, ambos expresados por cada 1,000 habitantes, el resultado indica el crecimiento vegetativo en cifras relativas, o sea por cada 1,000 habitantes.

Los siguientes ejemplos ilustran los cálculos correspondientes.

Ejemplo: En el año 1938 ocurrieron en el Perú 187,996 nacimientos; en el mismo año se registraron 95,817 defunciones.

Luego, el crecimiento vegetativo, en cifras absolutas, fué en ese año:

187996 — 95817 = 92179 habitantes

En el año 1938, en el Perú, el coeficiente general de natalidad y el coeficiente general de mortalidad, ambos expresados por cada 1,000 habitantes, fueron, respectivamente: 31.39 ‰ y 16.00 ‰.

Luego, el crecimiento vegetativo, por cada 1,000 habitantes, fué:

31.39 — 16.00 = 15.39 ‰

Es decir, que por cada 1,000 habitantes del Perú, en el año 1938, la población aumentó aproximadamente 15 habitantes.

En el Cuadro 56 están consignadas las cifras de crecimiento vegetativo, absoluto y relativo, de la población del Perú durante el período 1931—1940.

CUADRO 56

REPUBLICA DEL PERU. PERIODO 1931—1940.

Año	Nacimientos		Defunciones		Crecimiento vegetativo	
	Total	Por 1,000 hab.	Total	Por 1,000 hab.	Absoluto	Por 1,000 hab.
1931	165,074	31.28	70,132	13.29	94,942	17.99
1932	166,636	31.02	70,664	13.15	95,972	17.87
1933	168,962	30.90	70,472	12.89	98,490	18.01
1934	174,731	31.39	74,275	13.34	100,456	18.05
1935	179,363	31.73	80,876	14.27	98,937	17.46
1936	182,981	31.70	89,803	15.56	93,178	16.14
1937	187,134	31.83	94,374	16.05	92,760	15.78
1938	187,996	31.39	95,817	16.00	92,179	15.39
1939	177,664	29.11	90,579	14.84	87,085	14.27
1940	167,290	26.91	85,996	13.83	81,294	13.08

Indice vital.—Se obtiene dividiendo el número de nacimientos entre el número de defunciones (N/D), ocurridos durante un período determinado, que puede ser uno o varios años.

Si el índice vital es igual a 1 la población es *estacionaria* desde el punto de vista de crecimiento; si es menor a 1 la población *decrece*, y, finalmente, si el índice vital es mayor a 1 la población *crece*. *

Ejemplos En el año 1941 ocurrieron en el Perú 157,492 nacimientos y 91,367 defunciones.

Luego, el índice vital, correspondiente a dicho año, fué:

$$\frac{157492}{91367} = 1.72$$

En el Cuadro 57 está incluído el índice vital que corresponde a la ciudad de Lima durante los años comprendidos en el período 1936—1940. Puede apreciarse, en este cuadro, que en el año 1939 por cada defunción ocurrieron, aproximadamente, dos nacimientos.

* El índice vital no toma en cuenta las fluctuaciones de la población debidas a los movimientos inmagratorio y emigratorio.

CUADRO 57

CIUDAD DE LIMA.

Año	Número de nacimientos	Número de defunciones	Indice vital
1936	9,653	6,073	1.59
1937	10,414	6,222	1.67
1938	10,858	6,065	1.79
1939	11,642	5,610	2.08
1940	15,328	8,517	1.80

El índice vital puede adquirir cierto carácter de especificidad si se le calcula separadamente para cada núcleo racial de habitantes de una población. Los resultados, así obtenidos, pueden ser interesantes desde un punto de vista comparativo.

Por ejemplo, si se quisiera establecer el índice vital correspondiente a los habitantes de raza blanca y a los habitantes de raza negra de una población, se aplicarían las siguientes fórmulas:

$$\text{Indice vital en la raza blanca} = \frac{\text{N}^{\circ}\text{ de nacidos de raza blanca}}{\text{N}^{\circ}\text{ de muertos de raza blanca}}$$

$$\text{Indice vital en la raza negra} = \frac{\text{N}^{\circ}\text{ de nacidos de raza negra}}{\text{N}^{\circ}\text{ de muertos de raza negra}}$$

APENDICE A

LOGARITMOS

Nos.	Logaritmos	Nos.	Logaritmos
1	0.0000	51	1.7076
2	0.3010	52	1.7160
3	0.4771	53	1.7243
4	0.6021	54	1.7324
5	0.6990	55	1.7404
6	0.7782	56	1.7482
7	0.8451	57	1.7559
8	0.9031	58	1.7634
9	0.9542	59	1.7709
10	1.0000	60	1.7782
11	1.0414	61	1.7853
12	1.0792	62	1.7924
13	1.1139	63	1.7993
14	1.1461	64	1.8062
15	1.1761	65	1.8129
16	1.2041	66	1.8195
17	1.2304	67	1.8261
18	1.2553	68	1.8325
19	1.2788	69	1.8388
20	1.3010	70	1.8451
21	1.3222	71	1.8513
22	1.3424	72	1.8573
23	1.3617	73	1.8633
24	1.3802	74	1.8692
25	1.3979	75	1.8751
26	1.4150	76	1.8808
27	1.4314	77	1.8865
28	1.4472	78	1.8921
29	1.4624	79	1.8976
30	1.4771	80	1.9031
31	1.4914	81	1.9085
32	1.5051	82	1.9138
33	1.5185	83	1.9191
34	1.5315	84	1.9243
35	1.5441	85	1.9294
36	1.5563	86	1.9345
37	1.5682	87	1.9395
38	1.5798	88	1.9445
39	1.5911	89	1.9494
40	1.6021	90	1.9542
41	1.6128	91	1.9590
42	1.6232	92	1.9638
43	1.6335	93	1.9685
44	1.6435	94	1.9731
45	1.6532	95	1.9777
46	1.6628	96	1.9823
47	1.6721	97	1.9868
48	1.6812	98	1.9912
49	1.6902	99	1.9956
50	1.6990	100	2.0000

APENDICE B

Ejemplos para efectuar operaciones aritméticas de cantidades con signo diferente.

Suma—

$$(-2) + (-4) = -6$$
$$(-2) + (+4) = +2$$
$$(+2) + (-4) = -2$$
$$(+2) + (+4) = +6$$

Resta—

$$(-2) - (-4) = +2$$
$$(-2) - (+4) = -6$$
$$(+2) - (-4) = +6$$
$$(+2) - (+4) = -2$$

Multiplicación—

$$(-2) \times (-4) = +3$$
$$(-2) \times (+4) = -3$$
$$(+2) \times (-4) = -3$$
$$(+2) \times (+4) = +3$$

División—

$$\frac{-4}{+2} = -2$$

$$\frac{-4}{-2} = +2$$

$$\frac{+4}{-2} = -2$$

$$\frac{+4}{+2} = +2$$

Resta y multiplicación—

$$-10 - (-2 \times -3) =$$
$$-10 - (+6) = -16$$

$$-10 - (+2 \times -3) =$$
$$-10 - (-6) = -4$$

$$+10 - (-3 \times -2) =$$
$$+10 - (+6) = +4$$

$$+10 - (+3 \times -2) =$$
$$+10 - (-6) = +16$$

$$-10 - (+3 \times +2) =$$
$$-10 - (+6) = -16$$

$$+10 - (+3 \times +2) =$$
$$+10 - (+6) = +4$$

INDICE ALFABETICO