Hospital Luis Calvo Mackenna.
Cátedra de Pediatría del
Prof. Aníbal Ariztia.

ESTUDIO DEL PESO DEL LACTANTE

II. Comunicación.

Por los Dres. TEODORO ZENTENO y OSCAR CORREA B.

En comunicación aparecida en fecha reciente en esta misma Revista¹, dimos a conocer un estudio comparativo entre el peso del lactante de Hospital con el de la Caja de Seguro. Dicho estudio, como se recordará, comprendió únicamente niños alimentados en forma artificial, los que, en el hospital o policlínicas del Seguro, estuvieron bajo cuidado ininterrumpido de pediatra desde las primeras semanas de la vida hasta el primer año de edad. El presente artículo tiene por objeto exponer los resultados de un trabajo de igual índole en un grupo de lactantes alimentados al pecho.

Plan de estudio.

Nuestro material comprende 200 lactantes de la Caja de Seguro. De igual modo que en el estudio antes mencionado, se incluyeron sólo niños nacidos de término y en buen estado nutritivo a su fecha de inscripción en la policlínica correspondiente y como en aquél, eliminaron casos con anomalías constitucionales, malformaciones congénitas, sífilis y tuberculosis. Respecto a alimentación, todos los niños fueron criados al pecho exclusivo hasta los 6 meses, por lo menos, edad en que se inició el agregado de sopas, purés y frutas, para posteriormente seguir con las adiciones indicadas por las normas pediátricas corrientes. Del total de 200 niños que componen nuestro material, los dos tercios de ellos pertenecen al Consul-

torio N° 15 (Unidad Sanitaria de Quinta Normal) y los restantes al Consultorio N° 16 (Vitacura). En su obtención no se aplicó ningún criterio de selección y las fichas respectivas fueron separadas al azar hasta completar el total mencionado entre aquéllas que cumplían las condiciones antes fijadas.

Antes de dar a conocer los resultados del presente estudio, creemos de interés hacer algunas consideraciones de orden estadístico, que ayudarán a su mejor comprensión.

Consideraciones estadísticas.

Generalmente, el estadístico trabaja con un número considerable de datos u observaciones. De allí, entonces, que uno de los propósitos que tiene en vista al aplicar los métodos de que se vale, es el de simplificar el material numérico que analiza^{2 3 4 5 6 7 8}.

Si se nos mostrara, por ejemplo, en cualquier momento, una Tasa o un registro que comprenda 1,000 números diferentes, en forma tal que cada uno de ellos venga a representar el peso de un niño recién nacido y pretendiéramos obtener del estudio de esas cifras alguna información de valor, es evidente que nos hallaríamos un tanto confundidos frente a la sola exposición del material numérico, lo que, por el contrario, no sucedería si del conjunto total de observaciones lográramos desprender dos o tres expresiones capaces de condensar la información que nos interesa. Así, por ejemplo, podríamos establecer que el promedio aritmético del peso de nacimiento de los niños del sexo masculino fué de 3,450 gramos, mientras que el peso de las niñas, en idénticas condiciones, fué de 3,150 gramos, con lo cual habríamos obtenido dos expresiones que resumen de una manera breve los datos que se ofrecen en una copiosa información.

Sin embargo, es conveniente, desde luego, establecer que los datos así obtenidos no reflejan en manera alguna la variabilidad de los valores de las observaciones que se estudian ni pretenden tampoco señalar que cada uno de los niños varones pesaba exactamente 3,450 gramos, o que cada una de las niñas pesaba 3,150 gramos. Lo que hemos hecho es únicamente consignar uno de los aspectos particulares que se desprenden del estudio de las cifras originales, traduciéndolo en una expresión numérica más susceptible y fácil de manejar.

Si nos remitimos a la Tabla 1 veremos en ella una distribución de los pesos de un grupo de 200 niños pertenecientes a la clase proletaria e inscritos en la Caja de Seguro Obre-

TABLA 1
Distribución de frecuencias del peso de 200 observaciones
pertenecientes a la Caja de Seguro (Serie simple).
Año 1948.

	8.4	8.9	9.5
6.7	8.4	8.9	9.6
6.9	8.5	8.9	9.6
6.9	8.5	8.9	9.6
7	8.5	8.9	9.6
7	8.5	8.9	9.6
7	8.5	8.9	9.6
7.3	8.5	8.9	9.6
7.4	8.5	9	9.6
7.4	8.5	9	9.6
7.4	8.5	9	9.6
7.4	8.5	9	9.6
7.4	8.5	9	9.6
7.4	8.5	9	9.6
7.5	8.5	9	9.7
7.5	8.5	9	9.7
7.5	8.6	9	9.7
7.6	8.6	9	9.7
7.6	8.6	9	9.7
7.6	8.6	9.1	9.7
7.7	8.6	9.1	9.7
7.7	8.6	9.1	9.7
7.8	8.6	9.1	9.7
7.8	8.6	9.1	9.8
7.8	8.6	9.1	9.8
7.8	8.6	9.1	9.8
7.8	8.6	9.1	9.8
7.8	8.6	9.1	9.9
7.8	8.6	9.2	9.9
7.9	8.7	9.2	9.9
7.9	8.7	9.2	10
8	8.7	9.2	10
8	8.8	9.3	10
8	8.8	9.3	10
8	8.8	9.3	10
8	8.8	9.3	10.1
8	8.8	9.3	10.1
8.1	8.8	9.3	10.2
8.1	8.8	9.3	10.2
8.1	8.8	9.3	10.2
8.2	8.8	9.3	10.5
8.2	8.8	9.3	10.6
8.2	8.8	9.4	10.6
8.2	8.8	9.4	10.8
8.2	8.8	9.4	11
8.2	8.8	9.4	11.1
8.2	8.8	9.5	11.4
8.3	8.8	9.5	11.8
8.4	8.8	9.5	11.9
8.4	8.8	9.5	12
8.4	8.9	9.5	12
8.4	8.9	9.5	12.6

Total: 200 obs.

ro, que fueron alimentados al pecho hasta la edad de 6 meses y cuyo aspecto ponderal fué analizado al cumplir 1 año de edad. Desde luego, observamos en ella una perfecta ordenación de los valores. Para lograr este objeto, el peso registrado en cada uno de los niños, al cumplir un año de edad, se fué inscribiendo en sucesión numérica, desde el valor más bajo hasta el valor más alto presentados, repitiendo en muchos casos todas aquellas observaciones que ofrecían la misma medición. Es esta forma sencilla de presentación la que, en términos estadísticos, se conoce con el nombre de serie simple o no agrupada. En efecto, cada uno de los valores de la variable en estudio corresponde a la medida presentada por un solo individuo, de donde, entonces, su frecuencia para cada uno de los casos u observaciones es igual a 1.

TABLA 2

Distribución de frecuencias del peso de 200 observaciones pertenecientes a la Caja de Seguro (serie agrupada).

Año 1948.

Clase (X)	Frecuencia (F)	Clase (X)	Frecuencia (F)
6.7	1	9.1	8
6.9	2	9.2	4
7	3	9.3	10
7.2	1	9.4	4
7.3	1	9.5	7
7.4	5	9.6	11
7.5	3	9.7	10
7.6	3	9.8	3
7.7	2	9.9	3
7.8	6	10	5
7.9	2	10.1	2
8	6	10.2	3
8.1	3	10.5	1
8.2	7	10.6	2
8.3	1	10.8	1
8.4	6	11	1
8.5	13	11.1	1
8.6	12	11.4	1
8.7	3	11.8	1
8.8	18	11.9	1
8.9	10	12	2
9	10	12.6	1

Total: 200

Ahora bien, el empleo de esta serie resulta conveniente cuando el número de observaciones que se presentan es reducido y, en general, cuando las frecuencias de la variable en estudio no se repiten. Cuando las observaciones individuales tienen una frecuencia mayor que 1, resulta más útil una forma de información más condensada. En la tabla 2 presentamos estos mismos datos de una manera más clara todavía. En la columna 2, que corresponde a las frecuencias totales, anotamos en una sola cifra el número de individuos que presentaron el mismo peso, en el momento de recoger la información.

Series agrupadas. Series de frecuencia. Distribución de frecuencia. — La Tabla 3 presenta los datos de una manera más condensada todavía y habitualmente es ésta la forma más general de rendir una información. En la columna 1 hemos agrupado el peso de los niños en 20 filas, cada una de las cuales constituye un grupo, una clase o una categoría. Cada una de estas clases viene a constituir un aspecto particular de la variable, peso en este caso, en estudio. Así, por ejemplo, en la primera de las clases, 1ª fila, columna 1, leemos 6.700-6.999 gramos. Dentro de ella debemos incluir todos aquellos niños cuyos pesos se hallen comprendidos dentro de esos límites. Como se desprende del estudio de la tabla hemos observado un intervalo regular de 300 gramos para cada una de las diferentes clases y esta constancia regular en el mantenimiento de este intervalo, en ésta como en cualquiera otra tabla de frecuencia, es de importancia fundamental, puesto que facilita enormemente el cálculo que al final hacemos con los datos allí expresados para la obtención de las constantes que interesan.

¿Es arbitraria la elección de este intervalo? Es evidente que si nosotros, en vez de agrupar los pesos en clases de 300 grs. cada uno, lo hubiéramos hecho en intervalos de 600 grs., habríamos obtenido al final 10 filas, en lugar de 20 grupos o clases diferentes. En rigor, la elección del intervalo es arbitraria, pero conviene tener presente dos hechos fundamentales: 1º una amplitud demasiado grande en el intervalo que se emplee, puede traer consigo un cierto grado de inexactitud en el resultado de los cálculos, y 2º una estrechez forzada de los datos puede determinar una menor sencillez de presentación.

Notamos, finalmente, de la Tabla 3, ciertos aspectos que vale la pena recalcar. Al disponer los datos en esa forma

de manera tal que podamos conocer el número de veces que un acontecimiento, peso en este caso, se registra, decimos que hemos construido una Tabla o hemos realizado una Distribución de las frecuencias. Las Tablas de frecuencia señalan, entonces, la distinta gradación de los valores de la variable que se analiza y el número de veces que dichos valores se presentan.

Clases. Límites de Clase. — Volvamos nuestra atención para considerar nuevamente los medios de que nos valemos para señalar y distinguir cada una de las diferentes clases, grupos o categorías. En la Tabla 3 las encontramos descritas como sigue:

6.700	—	6.999
7.000	—	7.299
7.300	—	7.599
7.600	—	7.899
etc.		

Cada una de estas agrupaciones se halla limitada por dos números, que constituyen los límites de cada clase. Así, por ejemplo, los límites de clase del primero de los grupos de la Tabla 3 son los números 6.700 y 6.999. El primero de ellos, 6.700, es el límite inferior y el que le sigue y el mayor de ellos, 6.999, el límite superior. La diferencia existente entre el límite inferior y verdadero de una clase cualquiera y el límite inferior y verdadero de la clase inmediatamente superior es lo que se llama Intervalo de clase. Así, por ejemplo, el intervalo de clase de la serie en estudio, Tabla 3, es de 300 grs. cada uno ($7.000 - 6.700 = 300$ grs.).

No debemos confundir este término o expresión con el denominado Rango o Amplitud de la serie. En efecto, si clasificamos los términos de una serie en orden creciente o decreciente se obtendrán necesariamente los valores máximos y mínimos de las clases consideradas. Amplitud de la serie o Rango es la diferencia que existe entre estos valores extremos. En la serie citada anteriormente (Tabla 3) los valores extremos son: 6.700 y 12.700 grs.; por lo tanto, la amplitud es $12.700 - 6.700 = 6.000$ grs. Si dividimos esta cifra por 20, que es el número de grupos o clases que hemos elegido, tendríamos por cociente 300 grs., que es el intervalo de clase que hemos adoptado.

¿Qué es lo que entendemos por límites superior e inferior de las cifras que constituyen los extremos de una clase cualquiera? Analicemos la serie siguiente;

6.7 — 6.9
 7 — 7.2
 7.3 — 7.6
 etc.

Si dicha serie representa peso en kilogramos no podríamos, en un momento dado, establecer dentro de qué categoría quedan incluidos aquellos niños cuyos pesos sean superiores a 6.9 e inferiores a 7 kgrs., a menos que hayamos expresado de un modo previo, cuál ha sido nuestra conducta al proceder y efectuar esta distribución. Sin embargo, los métodos estadísticos, de acuerdo con aspectos simplemente convencionales, permiten una incorporación adecuada de tales observaciones y lo hacen de una manera que no queden vacíos de inclusión y satisfagan, por otra parte, todos aquellos que en el curso del estudio pudieran presentarse.

Uno de tales procedimientos, basado en la inexactitud de las mediciones observadas, de los errores de observación, de registro y de recuento, permite establecer la forma de que nos valemos para incluir dentro de un grupo determinado todo aquel conjunto de valores que no tienen, aparentemente, por lo menos, una ubicación precisa. Así, por ejemplo, en el primero de los grupos anotados, 6.7 — 6.9, se incluyen todas las observaciones cuyo peso verdadero se halla entre 6.65 y 6.94, las que integran el segundo de los grupos estarían comprendidas dentro de los límites 6.95 y 7.24 y así, sucesivamente. Estadísticamente expresamos este hecho diciendo que hemos aproximado los valores a la unidad más cercana.

Con el objeto de obtener una mejor visión de lo que hemos expresado, reproducimos a continuación la Tabla de este último ejemplo y colocamos al lado de ella la forma en que debe ser interpretada:

6.7 — 6.9	6.65 — 6.94
7 — 7.2	6.95 — 7.24
7.3 — 7.5	7.25 — 7.54
7.6 — 7.8	7.55 — 7.84
etc.	etc.

Extremos y punto medio del intervalo de clase. Centro de clase. — Los valores límites del intervalo de clase se llaman extremos o puntos límites del intervalo. El punto medio de dicho intervalo se obtiene de la semisuma de los valores ex-

tremos. Corrientemente se calcula este valor dividiendo por 2 el intervalo de clase que hemos elegido y sumando este cociente al límite inferior de un intervalo cualquiera, siempre que éstos sean de la misma naturaleza o medida.

Frecuencia absoluta y frecuencia relativa. — El número de individuos u objetos a los que corresponde un valor de la variable comprendidos entre los extremos del intervalo de clase recibe el nombre de Frecuencia absoluta. En nuestro ejemplo de la Tabla 3, 14 niños presentaron un peso comprendido entre 8.200 y 8.499 grs. 14 niños, entonces, constituirían la frecuencia absoluta de este grupo. Frecuencia relativa es, en cambio, la relación que existe entre la frecuencia absoluta y el número total de términos que se consideran. En el ejemplo anterior, la frecuencia relativa para el peso de dichos niños sería $14/200$, lo que es igual a 0.07. Como estas cifras se expresan en tanto por ciento, diríamos simplemente que el 7 % de los niños presentó un peso comprendido entre 8.200 y 8.499 grs.

Promedios. — El estudio de toda Tabla de frecuencias nos conduce inevitablemente a la obtención por medio del cálculo de ciertos y determinados valores que sirven para representar, de una manera fundamental, las características más esenciales de toda serie numérica. Ya nosotros hemos visto que los métodos estadísticos persiguen en muchos de sus aspectos la simplificación de los datos emanados del estudio de las observaciones y esta simplificación es posible obtenerla muchas veces mediante el conocimiento de algunas constantes, que sirven para describir acertadamente una distribución y, en último término, para comparar las constantes de una distribución que se observan o proceden de otras series. Cualquiera expresión que llega de esta manera a constituirse en representante de la totalidad de los valores integrantes de una serie es lo que se denomina generalmente Promedio. En el lenguaje corriente esta expresión se confunde muy a menudo con la Media aritmética de una serie cualquiera. Debemos, sí, advertir que existen varias clases de promedios y, en ocasiones, técnicamente y dada la índole o aspecto de los problemas en estudio, el estadístico usa uno o alguno de ellos, con preferencia a otros.

Las constantes estadísticas de tendencia central más conocidas son la Media aritmética, llamada también promedio aritmético o simplemente media, la Mediana y el Modo. Menos empleadas en los trabajos médicos corrientes existen otras dos representantes: la Media geométrica y la Media

armónica, de las cuales no nos ocuparemos por el momento.

Media aritmética. — No es otra cosa que el cociente de la división de la suma de una serie de valores por el número de los términos que la constituyen. Podríamos concretar esta definición en la forma siguiente:

$$\text{Media (M)} = \frac{S(X)}{N}$$

En la que X viene a representar el valor individual de cada observación.

N el número total de casos u observaciones.

S significa "la suma de".

M representa la media aritmética de las observaciones.

Sin embargo, si bien es cierto que la Media aritmética tiene en su favor la evidente facilidad con que se la determina, presenta también algunas desventajas, que es necesario y conveniente puntualizar. Desde luego, es un valor que se deja influir fácilmente por los términos extremos, hasta el punto que puede desfigurar por completo el concepto que lógicamente pueda derivarse del estudio de una serie. No debemos olvidar que las medias son simplemente expresiones numéricas, que intentan simplificar una serie estadística, manifestando en un dato breve la característica esencial y, al mismo tiempo, fundamental de aquella serie numérica. Supongamos, por ejemplo, que en una localidad cualquiera, existe una endemia de Fiebre tifoidea y que el número de casos registrados en los diferentes meses del año sean los siguientes:

Enero	16
Febrero	18
Marzo	15
Abril	12
Mayo	8
Junio	7
Julio	5
Agosto	5
Setiembre	6
Octubre	11
Noviembre	15
Diciembre	17

La suma de los casos es igual a 135, los que divididos por 12 (meses), nos da una media de 11.2 enfermos, cifra que corresponde con bastante aproximación al número de enfermos que se observa cada mes. Pero supongamos, ahora, que por una contaminación masiva o la presencia de un brote epidémico durante los meses de noviembre a diciembre, se hubiesen registrado 80 y 70 casos, respectivamente. Si en estas condiciones calculamos la Media aritmética, ésta resultará ser ahora igual a 22. Así, lo que ha ocurrido, un hecho excepcional, viene a desfigurar la realidad y lo que es habitual dentro del carácter endémico que se observa en dicha localidad.

De la facilidad con que la Media se deja influir por los movimientos extremos de la serie se deduce otro de los inconvenientes de esta expresión numérica y es de que no sigue las variaciones proporcionales de las series. Supongamos, por ejemplo, que en el curso del año 1928, en una población cualquiera, se han observado 40 casos de difteria laríngea en el curso del primer semestre y otros 40 casos en el segundo semestre. En el año siguiente, 1929, se ven 80 casos en el primer semestre y 20 en el segundo. Vemos aquí que el incremento experimentado en el primer semestre fué compensado posteriormente por reducción del número a la mitad de lo que se observa habitualmente en el curso del segundo semestre. Esta compensación natural no queda, en manera alguna, expresada por la Media, ya que ésta es de 50 casos (año 1929), en tanto que en el año anterior fué de 40 casos.

Es evidente, además, que la variabilidad que se observa en los valores individuales de cualquiera serie de observaciones, no puede ser reflejada en la cifra o media representativa de dicha serie de valores. En realidad, hay muchos hechos respecto de la naturaleza íntima de un fenómeno que el simple estudio de la media aritmética no puede ni permite establecer. Supongamos que deseemos obtener la media aritmética de las siguientes series:

48, 49, 50, 51, 52
30, 40, 50, 60, 70
10, 20, 50, 80, 90

El cálculo nos demuestra que la media aritmética de cualquiera de ellas es igual a 50 y en muchos casos no puede ser indiferente, como se comprende, que los valores de una serie oscilen en límites tan amplios, como en el ejemplo úl-

timo (10 a 90) o en límites tan reducidos como en la primera (48 a 52).

Otra de las características que puede desprenderse del estudio de las series expresadas en el ejemplo anterior y sobre cuyo aspecto fundamental nada dice el promedio aritmético obtenido es la distinta repartición de los valores que se observan en las series en estudio, muy concentrados alrededor del valor central en el primer ejemplo, igualmente repartidos a lo largo de la serie en el segundo y bastante alejados del término central en el tercero. Para una mayor ilustración de lo que acabamos de expresar, supongamos que el promedio de las notas obtenidas durante una promoción de dos grupos de estudiantes, hombres y mujeres, sea aproximadamente igual a 75. Esta cifra, en realidad, nada nos dice acerca del comportamiento de la naturaleza íntima de ambos grupos. Es posible y puede acontecer que los individuos integrantes de uno de estos grupos hayan obtenido individualmente notas excelentes, cuyos promedios generales se hallen muy cerca o muy vecinos del promedio más general, en tanto que en el otro grupo pudo ocurrir que la existencia de un número selecto de individuos haya alcanzado un alto valor en las notas obtenidas, anulando con su influencia el comportamiento de los valores bajos obtenidos por individuos constituyentes del mismo grupo. Es fácil comprender en este caso que la media aritmética, no obstante ser la misma en ambos grupos, nada nos dice acerca del comportamiento de los diferentes individuos que constituyen ambos grupos y en el estudio de esta serie como en cualquiera otra serie de valores o medidas, no puede, en manera alguna, ser indiferente el desconocimiento de la variabilidad de las medidas que las constituyen ni puede ser indiferente cualquier intento de penetrar en el conocimiento íntimo del fenómeno que se estudia.

Finalmente, la experiencia de los que se ocupan de estudios de esta naturaleza aconsejan no calcular este valor cuando los términos que la componen se hallen muy alejados unos de otros, en cuyo caso es aconsejable el empleo de otros promedios. No debe calcularse tampoco este valor cuando el número de observaciones sea inferior a 15, debido a las perturbaciones que suelen observarse por el simple juego de las variaciones del azar.

Uno de los aspectos importantes que se desprenden del estudio de la media aritmética es el de que las sumas de las desviaciones o diferencias de los valores individuales de las observaciones con respecto a la media de la serie respectiva es

igual a cero y esto es evidente desde el momento que la media aritmética representa el valor central de la serie completa de los términos. Sea la variable $X_1, X_2, X_3, X_4, X_5, \dots, X_n$.

La media aritmética, por definición, es

$$M = S X / N$$

Ahora bien, si restamos la media de cada observación, tendremos

$$X_1 - M = d_1$$

$$X_2 - M = d_2$$

$$X_3 - M = d_3$$

$$\dots\dots\dots$$

$$\dots\dots\dots$$

$$X_n - M = d_n$$

$SX_i - NM = S d_i$ (sumando los términos de la serie).

La letra d (minúscula) se usa convencionalmente para expresar con ella las desviaciones de la variable a partir de la media aritmética. En la suma anterior $SX - NM = S d_i$ podemos sustituir la media (M) por su valor correspondiente SX/N , puesto que M es igual a dicha expresión y tendremos

2) $SX_i - N(SX_i/N) = S d_i$ dividiendo por N
tenemos

3) $SX_i - SX_i = S d_i$ luego

4) $S d_i = 0$.

Resulta también importante conocer el hecho de que la suma de los cuadrados de las desviaciones de las observaciones con respecto a la media aritmética de la serie es siempre de un valor inferior al total que puede obtenerse de la suma de los cuadrados de las desviaciones con respecto a otro valor cualquiera, diferente de la media aritmética verdadera. Tomemos a título de ilustración, como un ejemplo, los números 5, 8 y 14. Si comparamos estas cifras con el número 10, encontramos de inmediato que ellas difieren de él en (-5) , (-2) y $(+4)$. Dichas diferencias elevadas al cuadrado nos dan 25, 4 y 16 y la suma de estos valores 45. Si ahora comparamos estos mismos números 5, 8 y 14, con su media aritmética verdadera $= 9$ ($5 + 8 + 14 = 27$; $27 : 3 = 9$), habríamos encontrado que las diferencias o desviaciones con respecto a la media son de -4 , -1 y $+5$. Los cuadrados de esas diferencias son 16, 1 y 25 y la suma de ellos 42, que es.

como puede apreciarse menor que la suma de los cuadrados que se obtuvieron cuando se calcularon las desviaciones a partir de la cifra 10 elegida como media arbitraria de la serie.

Cualquiera que sea el número que se elija en ésta como en cualquier otro tipo de observaciones se encontrará invariablemente que la suma de los cuadrados de las desviaciones obtenidos a partir de la media aritmética es un valor mínimo comparado con el que se obtiene al efectuar este cálculo a base de otra media cualquiera. Esto viene a justificar el hecho que la media aritmética de una serie de observaciones está ajustada de acuerdo con el proceder de los mínimos cuadrados, expresión a la que nos referiremos en otra oportunidad, al considerar algunos de los aspectos más fundamentales que se derivan del estudio y del empleo del coeficiente de correlación.

Una demostración más general de este aspecto puede deducirse del estudio que ofrecemos a continuación. Supongamos que hemos elegido un valor cualquiera (M), a partir del cual debemos medir las desviaciones (d). En consecuencia, para cada uno de los valores (X) de nuestra serie original tendríamos que

$$\begin{aligned} X + d &= M: \text{ de donde} \\ d &= M - X; \text{ cuadrando ambos términos} \\ &\text{tendremos} \\ d^2 &= (M - X)^2 \\ d^2 &= M^2 - 2MX + X^2 \end{aligned}$$

Si sumamos, ahora, las observaciones, tendremos que

$$\begin{aligned} \sum d^2 &= \sum (M^2 - 2MX + X^2) \\ \sum d^2 &= NM^2 - 2M \sum SX + \sum SX^2, \text{ la cual viene} \\ &\text{a ser la suma de los cuadrados de las desviaciones que nos} \\ &\text{otros debemos reducir a un mínimo y alcanzarán este valor} \\ &\text{cuando la primera diferencial sea igual a cero.} \end{aligned}$$

Si representamos dicha función por (f), tendremos que $df/dM = 2NM - 2\sum SX$, la cual hacemos igual a cero, de donde

$$Nm = \sum SX, \text{ luego } M = \sum SX/N \quad (q. e. d.)$$

En otros términos, nosotros obtendremos la menor suma posible de los cuadrados de las desviaciones cuando nuestro valor M sea igual a $\sum SX/N$, esto es, cuando sea igual a la media aritmética.

Medias ponderadas. — De bastante interés en los problemas estadísticos, como puede deducirse del estudio de las series siguientes:

$$\begin{aligned} A &= 6, 5, 8, 4 \text{ y } 6 \\ N &= 10, 14, 13 \text{ y } 15 \end{aligned}$$

Las medias aritméticas respectivas son 6, 16 y 13. Si quisiéramos, ahora, establecer una media de conjunto, la operación que corrientemente hacemos es la siguiente: sumamos 6, 16 y 13 y dividimos la suma por 3, lo que viene a ser igual a 9.58. Pero este número no representa en verdad la media que deseamos y el cálculo hecho en la forma anterior envuelve un error de hecho. Por definición hemos aceptado que la media aritmética es igual a la suma de los términos de la serie dividido por el número total de términos. Nuestro proceder correcto sería, en consecuencia, en este caso, el siguiente:

$$\frac{6 + 5 + 8 + 4 + 10 + 14 + 13 + 15}{10}, \text{ lo que es}$$

igual a 8.9, un valor diferente del que hemos encontrado con anterioridad. En términos generales, la operación que deberíamos efectuar para obtener con propiedad el valor que buscamos, se reduce a multiplicar el número de términos de cada una de las series por su media aritmética, sumar los productos obtenidos y luego dividirlos por el número total de términos. En el ejemplo anterior procederíamos de la manera siguiente:

$$\begin{aligned} 6 \times 6.16 &= 36.96 \\ 4 \times 13 &= 52 \end{aligned}$$

$36.96 + 52 = 88.96$; $88.96 : 10 = 8.9$, con lo que se obtiene la media aritmética verdadera de las dos series de observaciones.

De una manera general, entonces, nosotros podemos calcular la media aritmética de varias series de observaciones independientemente del número de valores que integran cada serie. Si designamos las distintas series de observaciones bajo los rubros 1, 2, 3, 4, hasta N y representamos los totales de los valores de las observaciones de las diferentes series S S S hasta S, resulta evidente que el total ge-

neral de los valores de todos los grupos en conjunto, debe ser igual a la suma de los totales parciales de cada una de las diferentes series o, lo que es lo mismo, escrito en forma de ecuación, tendríamos

$$SX = SX_1 + SX_2 + SX_3 + \dots + SX_n$$

Pero nosotros sabemos que $SX = NM$ $SX_1 = NM_1$ etc., puesto que por definición la media aritmética M en este caso es igual a la suma de los valores de las observaciones (X) dividida por el número de términos (N) que la integran.

$M = SX/N$, de donde $SX = NM$, pasando N al primer miembro de la igualdad y cambiando el orden de los términos.

En consecuencia, tendríamos que

$NM = N_1 M_1 + N_2 M_2 + N_3 M_3 + \dots + N_n M_n$,
luego el valor buscado M debe ser

$$M = \frac{N_1 M_1 + N_2 M_2 + N_3 M_3 + \dots + N_n M_n}{N}$$

En otros términos, hemos multiplicado la media aritmética de cada uno de los subgrupos o series de observaciones por el número de casos correspondientes a cada serie y hemos dividido la suma por el número total de casos de las series consideradas en conjunto. De la misma manera que hemos procedido para obtener la media aritmética de dos series numéricas o de un mayor número de ellas, podemos proceder para obtener la media de series que dependen de otras series. Supongamos, por ejemplo, que se nos solicite la media aritmética de una serie de tasas, cada una de las cuales corresponde o representa, a su vez, la tasa media de cierto número de poblaciones, como es el caso que sigue:

4	poblaciones han observado una tasa de natalidad de 20 o/oo
2	" " " " " " " " " 22 "
7	" " " " " " " " " 21 "
5	" " " " " " " " " 25 "
4	" " " " " " " " " 27 "

en vez de sumar las tasas ($20 + 22 + 21 + 25 + 27 = 115$; $115 : 5 = 23$), cuya media vendría a ser 23, lo que debemos hacer es multiplicar cada tasa por el número de poblaciones

a la que ella corresponde, sumar luego todos estos productos y dividir la suma por el número total de poblaciones, con lo cual obtendríamos una media ponderada de 22 o/oo, cifra diferente a la que encontramos en el procedimiento anterior, a todas luces incorrecto.

Media geométrica. — Es la raíz del producto de una serie de términos, cuyo exponente es igual al número de términos. Dicha expresión es más estable y escapa, desde luego, a varias de las objeciones que se hacen a la media aritmética. Cuando nos referimos anteriormente a esta última, dijimos que ella no seguía las variaciones proporcionales de las series, hecho que resultó evidente en el ejemplo propuesto de los casos de difterialaríngea, en el cual, a pesar de la compensación habida entre el incremento al doble y el descenso a la mitad del número de casos en los dos semestres, la Media aritmética subía sin hallarse en relación con la variación proporcional.

Si determinamos la Media geométrica en el ejemplo anterior, obtendríamos:

$$\text{Media del primer año} = \sqrt{40 \times 40} = 40$$

Media del segundo año = $\sqrt{80 \times 20} = 40$, resultando que, como puede apreciarse, sigue y traduce dicha variación proporcional.

Mediana. — Es el valor central de una serie de observaciones. Distinguimos su empleo en las series simples y en las agrupadas. En una serie simple la mediana es aquella de las observaciones que ocupa el lugar central de ella una vez dispuestas en orden numérico. Sea la serie siguiente:

$$7 - 8.5 - 8 - 10 - 11 - 6 - 7.5 - 10.5 - 9$$

Colocados en el orden numérico ordinal tendríamos:

$$6 - 7 - 7.5 - 8 - 8.5 - 9 - 10 - 10.5 - 11$$

El término central, que es 8.5, viene a ser, por definición, la mediana de la serie. Cuando la serie numérica es de orden par, la mediana es igual, a la semisuma de los dos valores centrales.

Fácilmente se concibe la escasa influencia que sobre esta constante ejercen los valores extremos de la serie, de suerte que ella viene a dar una idea bastante ajustada de la realidad de los fenómenos que se estudian. Al referirnos anteriormente a los inconvenientes que presenta algunas veces la interpre-

tación de los valores obtenidos en el cálculo de la media aritmética, dijimos que es ésta una constante que se deja influir fácilmente por aquellos términos que expresan valores bastante diferentes de los que se observan habitualmente dentro de la serie. Así, si en el ejemplo citado antes, que se refería al brote anormal de fiebre tifoidea, observado en una localidad X, hubiésemos calculado la mediana, habríamos obtenido una cifra que no habría sido influida en la forma en que ocurrió, con el promedio aritmético, reflejando, entonces, con más propiedad, la realidad endémica de dicha población. En efecto, el cálculo de la mediana de la serie habría sido:

$$5 - 5 - 6 - 7 - 11 - 15 - 15 - 16 - 17 - 18 \\ - 70 - 80$$

Ya que la serie es de un orden par de términos, la mediana vendría a ser igual a la semisuma de los valores centrales:

$$15 + 15 = 30; 30 : 2 = 15$$

Si recordamos, el promedio aritmético calculado en esa oportunidad fué igual a 22.

Para obtener la mediana de una serie de frecuencias (series agrupadas), debemos proceder de otra manera y expresar que en este caso la mediana viene a ser aquel valor de la distribución por encima y por debajo del cual se distribuyen la mitad de las frecuencias. Consideremos para este objeto los datos expresados en la Tabla 3. El total de frecuencias o número de observaciones es igual a 200. Con el objeto de encontrar aproximadamente la mediana de dicha distribución, deberíamos suponer que las frecuencias absolutas y las medidas que cada una de ellas representa, se hallan uniformemente distribuidas a lo largo de cada intervalo de clase. Ahora bien, si en la distribución que analizamos tenemos un total de 200 observaciones, la mediana de la serie vendría a ser aquel valor por encima y por debajo del cual hay 100 observaciones repartidas.

Si contamos el número de observaciones (Tabla 3, columna 4), desde la clase superior observamos que la observación número 100 cae necesariamente en el intervalo de clase, cuya frecuencia corresponde a 38 (8ª fila, columna 4). Si es evidente que la observación número 100 cae dentro de

este intervalo de clase, la mediana debe estar necesariamente entre las observaciones número 20 y número 21 de este grupo. Sólo nos resta, entonces, calcular la parte proporcional que debe corresponderle a cada una de las 38 observaciones existentes en este intervalo de clase, diciendo "si a 38 observaciones de este intervalo de clase corresponden 300 grs., a cada una de ellas corresponderá la 38ava parte del total (en el supuesto de que los valores se encuentren uniformemente distribuidos) y a 20 observaciones, que es el valor que buscamos, le corresponderá 20 veces esta cantidad". El cálculo establece que el valor buscado es 163.3. En consecuencia, a la observación número 100 le corresponderán 8.800 grs. + 163.3, lo que es igual a 8.963.3, y a la observación inmediatamente siguiente, al efectuar un cálculo análogo, 9.126.6. Dado que esta serie es de orden par (200 observaciones), la mediana será igual, entonces, a $8.963.3 + 9.126.6 = 18.089.9$, lo que dividido por 2 da un valor de 9.041.1. De donde la mediana es igual a 9.041.1, valor un tanto superior a la media aritmética, que es igual a 8.992.

Modo. — Distinguimos el modo crudo y el modo verdadero. El primero de ellos es el valor que se observa con mayor frecuencia en una serie de observaciones. Este valor no es muy preciso, puesto que depende de la magnitud de los intervalos de clase y de los límites de dichos intervalos. En nuestro ejemplo de la Tabla 3, el modo crudo vendría a ser el valor comprendido entre 8.800 — 9.099, puesto que hay 38 observaciones que lo representan. El modo verdadero es un promedio de escasa aplicación en nuestros trabajos médicos de estadística.

En las distribuciones moderadamente asimétricas existe una relación aproximada entre las magnitudes de la media aritmética, la mediana y el modo. En aquellas otras, en que la distribución es enteramente simétrica, las tres medidas coinciden en un mismo punto. En circunstancias ordinarias la mediana se encuentra entre la media aritmética y el modo y su relación aproximada es la siguiente:

$$Mo = 3 \text{ Med.} - 2 \text{ media aritmética}$$

Si la distribución es moderadamente asimétrica y hemos calculado los valores representativos de la mediana y de la media aritmética de la serie, podemos substituir estos valores en la ecuación anterior, para deducir de ella el modo. Si nos remitimos a los resultados obtenidos en nuestro estudio, en

Tabla 3

Cálculo de la media aritmética y de la desviación Standard de
una distribución de frecuencia (Método abreviado) con el
Procedimiento de comprobación de Charlier

PESO EN GRAMOS	CENTRO DE CLASE	DESVIACIÓN	FRECUENCIA				
		(d)	(f)	(fd)	f(d+1)	f d ²	f(d-1) ²
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
6700 • 8999	6850	-7	3	-21	-18	147	108
7000 • 7299	7150	-6	4	-24	-20	144	100
7300 • 7599	7450	-5	9	-45	-38	225	144
7600 • 7899	7750	-4	11	-44	-33	176	99
7900 • 8199	8050	-3	11	-33	-22	99	44
8200 • 8499	8350	-2	14	-28	-14	56	14
8500 • 8799	8650	-1	28	-28	0	28	0
8800 • 9099	8950	0	38	0	38	0	38
9100 • 9399	9250	1	22	22	44	22	88
9400 • 9699	9550	2	22	44	66	88	198
9700 • 9999	9850	3	16	48	64	144	256
10000 • 10299	10150	4	10	40	50	160	250
10300 • 10599	10450	5	1	5	6	25	36
10600 • 10899	10750	6	3	18	21	108	147
10900 • 11199	11050	7	2	14	16	98	128
11200 • 11499	11350	8	1	8	9	64	81
11500 • 11799	11650	9	0	0	0	0	0
11800 • 12099	11950	10	4	40	44	400	484
12100 • 12399	12250	11	0	0	0	0	0
12400 • 12700	12550	12	1	12	13	144	169
Totales			200	+28	+228	2128	2384

$$M_a = \frac{28}{200} = 0.14$$

$$M_{real} = 8950 + (0.14 \cdot 300) = 8992$$

$$\sigma_a = \sqrt{\frac{2128}{200} - (0.14)^2} = 3.25$$

$$\sigma_r = 300 \cdot 3.25 = \underline{975}$$

que la media aritmética es $= 8.992$ y la mediana $= 9.041$ y reemplazamos estos valores en la fórmula, obtendríamos

$$\begin{aligned} Mo &= 3 (9.041) - 2 (8.992) \\ &= 9.139 \end{aligned}$$

Resultados.

Para interpretar estadísticamente los resultados del presente trabajo, necesitamos conocer ciertas constantes, que nos van a permitir describir la serie en su conjunto, estimar los límites de su variación y, aun, comparar estos resultados con los que puedan obtenerse del estudio de otras series. Dichas constantes son la media aritmética y su D. S. (desviación standard). Para calcular el primero de estos valores hemos acudido a un método abreviado, que, por otra parte, es el que se usa corrientemente en trabajos de esta índole. A fin de comprenderlo con facilidad, remitámonos a los datos expresados en la Tabla 3.

En dicha tabla se han inscrito las diferentes clases en la columna 1. Puesto que la primera clase incluye todos los ítems, cuya magnitud está comprendida entre 6.700 y 7.000, determinamos el centro del intervalo de la clase $= 6.859$ grs. y con él encabezamos los ítems de la columna 2. Procedemos de la misma manera para obtener los ítems que siguen a continuación. En un tercer tiempo nos fijamos una media aritmética arbitraria, eligiendo para este objeto uno de los valores (puede elegirse un valor cualquiera) que inscribimos como "punto medio de la clase" y frente a él, en una tercera columna, colocamos un 0.

En nuestro ejemplo (Tabla 3) hemos colocado el 0 frente al punto medio de la clase correspondiente a 8.950. Siempre es conveniente elegir un centro de clase que corresponda más o menos a la parte central de la distribución y su objeto es el de facilitar los cálculos aritméticos que es necesario hacer a continuación. Colocamos, en seguida, en serie ordinal los números 1, 2, 3, 4, etc., precedidos por el signo menos, por encima del 0 y los números 1, 2, 3, 4, etc., precedidos del signo más por debajo del 0. Dichos números representan desviaciones a partir de nuestra media arbitraria $= 0$ y representa cada una de ellas los centros de clases que se hallan inscritos en la columna 2. En la 4ª columna hemos inscrito la frecuencia de las distintas observaciones. La columna 5ª representa el producto de la desviación de la columna 3 por las frecuencias correspondientes (f d).

Si ahora, sumamos algebraicamente esta columna, encontramos que la suma de los productos fd , o lo que es lo mismo, de las desviaciones de la columna 3 por las frecuencias de la columna 4, es igual a $\pm = 28$. Si dividimos este valor por

Gráfico N° 1

Distribución de pesos de un grupo de 200 lactantes alimentados al pecho al cumplir un año de edad (1948)
Caja del Seguro Obligatorio. Stgo.

la suma de las frecuencias (suma de la columna 4), tendremos:

$\pm 28 : 200 = \pm 0.14$, lo cual nos indica que el promedio verdadero (no arbitrario) de nuestras ob-

servaciones se encuentra a $+ 0.14$ intervalos de clase por encima de la media aritmética que hemos tomado como origen. Dado que el intervalo verdadero de la clase es igual a 300 grs., al valor $+ 0.14$ de las desviaciones de la columna 3 le corresponderá en unidades reales $0.14 \times 300 = 42$ grs. Nuestra media aritmética elegida arbitrariamente fué de 8.950 grs. (el valor frente al cual colocamos nuestro 0) y si a ella le agregamos 42 grs., que es la corrección necesaria que debemos hacer, obtendremos $8.950 + 42 = 8.992$ grs., que es el valor de la media aritmética que buscamos.

En los trabajos estadísticos en forma constante se efectúan operaciones aritméticas y es posible, muchas veces, que en alguno de los cálculos se deslicen errores, que invalidan, como se comprende la labor realizada. Afortunadamente, con determinados procesos o procedimientos, es posible comprobar la exactitud de los resultados que se obtienen y aun verificar la validez de los mismos a medida que se avanza en el trabajo, de manera que al final del mismo puede uno comprobar si ha habido o no errores aritméticos u otros errores de distinta naturaleza.

En el caso de la media aritmética calculada a partir de una distribución de frecuencia es posible poner en práctica un procedimiento conocido con el nombre de método de comprobación de Charlier y que tiene por objeto determinar la exactitud de los cálculos efectuados. Para usar este procedimiento agregamos una columna más a nuestra tabla (columna 6), colocando en ella los valores de $f(d + 1)$, en la que f representa las frecuencias y d las desviaciones, como ya hemos visto. La columna 6 de la Tabla 2 nos permite de inmediato observar la forma en que trabaja dicho método. El primer número que encontramos en la columna 6 es -18 , número que se obtiene agregando la unidad (1) al valor de d (-7), con lo cual obtenemos $1 + (-7) = -6$, que multiplicado por el valor correspondiente de $f = (3)$, nos da como producto -18 .

Una vez aplicado este método, si nuestros cálculos han sido efectuados correctamente, encontramos que la suma algebraica de los valores de la columna 6, en el caso de la Tabla 3, $\sum f(d + 1) = + 228$, debe ser igual a la suma de los totales de las dos columnas que la preceden. En efecto:

$$+ 228 = 200 + 28$$

Como en algunas ocasiones la suma de una o de dos de las columnas pueden presentar signos negativos, debe ob-

servarse cuidadosamente al efectuar esta suma, la regla de los signos algebraicos. Supongamos, por ejemplo, que la suma de las frecuencias hubiese sido = 1945 y que la suma de los valores de fd hubiese arrojado (-18) . En tal caso, la suma de los valores de $f(d+1)$, habría sido simplemente $1945 + (-18) = 1927$. La demostración de este hecho es muy sencilla:

$$\begin{aligned} f(d+1) &= fd + f, \text{ luego} \\ \sum f(d+1) &= \sum fd + \sum f. \end{aligned}$$

El método de comprobación de Charlier puede aplicarse todavía como método de comprobación en el caso del cálculo de la D. S. Si nos fijamos en la columna 8 de la Tabla 3, observaremos en ella los valores de $f(d+1)$ al cuadrado. Para hallar dichos valores agregamos la unidad (1) a cada uno de los términos de la columna encabezada por d (3ª columna), con lo cual obtenemos el valor $d \pm 1$. Elevemos esta expresión al cuadrado y el resultado lo multiplicamos por las frecuencias respectivas (f). Así, por ejemplo, en la primera fila de la columna 3, correspondiente a las desviaciones, encontramos el número -7 , al cual si le agregamos 1 obtenemos -6 . Este número elevado al cuadrado nos da 36. Multiplicamos 36 por 3 (que es el valor correspondiente a f) y obtenemos 108. Una vez que obtenemos los números de la columna 8 sumamos estas cifras para obtener su total. En nuestra Tabla esto es igual a 2.384. Aplicamos ahora el método de Charlier, cuya ecuación es:

$$\sum [f(d+1)^2] = \sum (fd)^2 + 2 \sum (fd) + \sum f$$

lo que significa que la suma de la última columna debe ser igual a la suma de la columna f más la suma de la columna $(fd)^2$ más dos veces la suma de la columna fd . En nuestro caso

$$\begin{aligned} \sum f(d+1)^2 &= 2.384 \\ \sum (fd)^2 &= 2.128 \\ \sum (fd) &= 28 \\ \sum f &= 200 \end{aligned}$$

Sustituídos estos valores en la ecuación de Charlier, tendremos:

$$2.384 = 2.128 + 2 (+ 28) + 200$$

$$= 2.128 + 56 + 200$$

$= 2.384$, lo cual viene a probar que nuestros cálculos eran correctos.

Conclusión.

La media aritmética del conjunto de 200 niños de un año de edad, alimentados al pecho hasta los 6 meses, pertenecientes a la clientela de la Caja de Seguro, dió un valor igual a 8.992 grs. (± 975 grs.). En un trabajo anterior encontramos que la media aritmética de los niños alimentados en forma artificial era de 8.506.6 ($\pm 1.016.4$). Nos interesa saber, ahora, si la diferencia entre los promedios aritméticos de estos dos grupos es o no estadísticamente significativa. Puede demostrarse que la D. S. de la diferencia de estos promedios es igual a

$$\begin{aligned} \text{D.S. dif. M} &= \sqrt{\frac{(1.016.4)^2}{209} + \frac{(975)^2}{200}} \\ &= \sqrt{9.696.03} \\ &= 98.4 \end{aligned}$$

Dado que la diferencia de los promedios es igual a 485.4 ($8.992 - 8.506.6$), esta diferencia resulta de 4.9 veces su D.S. ($x/6 = 485.4 : 98.4 = 4.9$), valor que es altamente significativo.

Si quisiéramos establecer el peso medio del lactante chileno, perteneciente a la clientela asalariada y sujeto a las condiciones estipuladas en el trabajo que presentamos, podríamos decir que éste debe hallarse entre 9.129.8 y 8.854.2. En efecto, la D.S. del promedio de las muestras sería igual, en este caso, a $6 M = 975 : \sqrt{200} = 68.9$. Si aplicamos 2 D.S. a ambos lados del promedio aritmético aquí encontrado, 8.992 grs., el promedio real y verdadero de los niños alimentados en esta forma debería encontrarse dentro de estos límites.

Resumen.

Se presentan los resultados de un estudio del peso del lactante al cumplir un año de edad. El se basa en el análisis

de 200 observaciones pertenecientes a niños de la Caja de Seguro, sanos y alimentados al pecho durante los seis primeros meses de la vida. El trabajo se resume en la siguiente forma:

1. El peso medio resultó igual a 8.992 grs. (± 975).
2. La comparación de este valor con el obtenido en un conjunto de 209 niños alimentados artificialmente, 8.506.6 ($\pm 1.016.4$), revela una diferencia de 485.4 grs., la que es estadísticamente significativa ($X/6 = 4.9$).
3. Se hace una revisión de los métodos estadísticos de uso más corriente en estudios médicos.

Summary.

An study of the weight of healthy infants at their first birthday is reported. The work comprised 200 children of the Caja de Seguro fed by maternal nursing during at least the first six months of life. The following points may be emphasized:

1. The mean weight was 8.992 gm. (± 975).
2. A comparison between this value and that found in 209 infants reared by artificial feeding, 8.506.6 ($\pm 1.016.4$), reveals a difference of 485.4 gm., which is statistically significant ($X/6 = 4.9$).
3. A review of most important statistical methods used in medical studies is presented.

Bibliografía.

- 1.—ZENTENO, T. y CORREA, O. — Estudio del peso del lactante. Comparación entre lactantes de Hospital y de la Caja de Seguro. *Rev. Chilena de Pediatría* 19: 270-284, marzo de 1948.
- 2.—BRADFORD HILL, A. — Principles of Medical Statistics. The Lancet Limited. Londres. 1945.
- 3.—DE SHELLY HERNANDEZ, R. — La Estadística Aplicada a las Ciencias Biológicas. Litografía del Comercio. Caracas. 1941.
- 4.—KRAMER, E. — A first course in Educational Statistics. John Wiley & Sons Inc. New York. 1938.
- 5.—PEARL, R. — Medical Biometry and Statistics. Saunders. Philadelphia. 1940.
- 6.—PIZZI, M. — Los Métodos Estadísticos y su Aplicación a la Medicina. Imprenta Universitaria. Santiago de Chile. 1947.
- 7.—SMITH, J. — Elementary Statistics. Henry Holt & Co. New York. 1935.
- 8.—WAUGH, A. — Elements of Statistical Method. Mc Graw-Hill Book Co. New York. 1943.