# BIOESTADISTICA

**Dr. Carlos Chiriboga**
Profesor Agregado de la Facultad de
Medicina - Quito

# BIOESTADISTICA

Conforme van avanzando las ciencias modernas en el *análisis y conocimiento de la realidad* que estudian, van necesitando día a día más de la representación numérica de sus datos y de la comparación de las realidades cuantitativas encontradas. Es decir, van utilizando los métodos estadísticos como medio informativo y comparativo de los elementos numéricos estudiados. De allí que casi todos los autores estén de acuerdo en sus opiniones de que la Estadística no es una ciencia sino mas bien un conjunto de métodos auxiliares, y quizás podemos opinar hoy indispensables de las otras ciencias.

Se ha definido a la Estadística como el "conjunto de métodos especialmente adaptados a la clasificación de datos cuantitativos afectados por una multitud de causas" o como la colección, presentación, análisis e interpretación de datos numéricos.

Podemos decir que la Estadística es un conjunto de métodos mediante los cuales conocemos en forma objetiva y sintética una realidad mesurable y los cambios que la afectan. En otras palabras la Estadística será pues la que nos preste en forma muy rápida, ágil y fácil de apreciación, un sinnúmero de datos numéricos que representen a un todo disperso del cual no podríamos tener exacta idea por la sola apreciación de las partes que nos son conocidas y cuya captación puede ser deformada por los factores subjetivos.

Sin embargo, en la Estadística también pueden haber deformaciones de la realidad por diferentes causas, como por ejemplo por la influencia subjetiva que pueda recaer sobre la interpretación de los datos, pero esta es la parte con la que culmina un estudio estadístico y por lo tanto debe estar a cargo de personas perfectamente preparadas para la correcta valoración de datos.

Decimos que la Estadística nos dá una idea objetiva de la realidad. Así por ejemplo, si queremos conocer en qué forma afecta el problema de la malaria al Ecuador, por el solo conocimiento aproximado de las zonas conocidas como palúdicas y el examen de más o menos numerosos grupos de enfermos palúdicos. Nuestra visión será seguramente poco exacta, pero si luego de un chequeo de la población de las zonas palúdicas y un examen del número de enfermos que reciben tratamientos médicos, de los días de trabajo que dichos enfermos pierden, del número de defunciones por causa del paludismo, no sólo que tendríamos una clara idea del grave daño que esta enfermedad representa para el País, sino que aún más podríamos conocer en que grado afecta a la economía y al progreso social. Podríamos luego conocer por comparación con las cifras de otros países nuestra situación exacta de adelanto en el campo de la defensa de la salud, y aún más en lo futuro, las nuevas cifras nos irán demostrando la utilidad de las inversiones hechas con objeto de atacar esta plaga.

Así, pues, la Estadística nos proporciona métodos mediante los cuales podemos conocer en forma cuantitativa una realidad y los cambios que ésta experimenta. De allí que hoy sea elemento indispensable de toda ciencia. No es sólo la Medicina la que recurre a Métodos Estadísticos, sino también las demás ciencias Biológicas.

Ahora bien, puesto que hemos hecho una brevísima revisión de los principales razonamientos para considerar a la Estadística como un método que presta un incalculable aporte a la Medicina, hagamos también una ligera descripción de las fases de un proceso estadístico, sin que por ningún concepto pretendamos querer siquiera enseñar en tan poco tiempo como el que disponemos para nuestras charlas, algo de este conjunto de métodos. Sólo revisaremos brevemente el trayecto que sigue ordinariamente el dato para la obtención de conclusiones de carácter general, a partir de un número de observaciones.

Podemos esquematizar los procedimientos estadísticos al agruparlos de la siguiente manera:

1º—Recolección de datos.

2º—Tabulación y presentación de los mismos; y

3º—Análisis.

La primera fase, la de Recolección de datos, es tan importante como las demás pero con la especial circunstancia de que los errores que se comete en ellas afectan el resto de todo el trabajo estadístico, haciéndole perder validez. En la recolección de estos datos intervienen personas que muchas veces están desvinculadas del trabajo estadístico y que no comprender la necesidad de enmendar errores que pueden parecer pequeños pero que al sumarse deforman los resultados. Por esta razón al comenzar un trabajo de esta naturaleza deben tenerse en cuenta disposiciones generales tendientes a mejorarlo o facilitarlo.

En primer lugar todo estudio debe realizarse mediante un previo planteamiento de objetivos a alcanzarse y medios con los que se cuenta. De otra manera el investigador encontraría que muchos datos que recogió no le sirven o que le faltan otros que ya no los puede obtener. En este sentido cabe mencionar el grave defecto que puede observarse en muchos centros médicos en los cuales las fichas o historias clínicas cuentan con datos que realmente constituyen un modelo ideal, pero cuya recolección es poco menos que imposible dadas las posibilidades del personal con el que cuentan. He tenido ocasión de observar en alguna parte fichas de Salud en las que había que recoger más de 300 datos, eran realmente muy completas, pero obviamente impracticables. Aún en el caso de que se pudieran recoger todos esos datos, no pueden utilizarse después. Al proyectar un trabajo debe fijarse cuáles son los datos que se necesitan para alcanzar el objetivo que se busca. Es preferible hacer trabajos sencillos en los que se recogen bien los datos y todos se utilicen, que hacer otros laboriosísimos para luego utilizar parte de estos datos y despreciar el resto.

El siguiente paso que debe darse al realizar un trabajo es el de revisión de material colectado para prepararlo para la tabulación. Esta revisión debe consistir principalmente en la búsqueda de errores o contradicciones que a veces escapan a las personas encargadas de recoger los datos. Cero que casi no hace falta mencionar la necesidad de que cada evento que se estudia sea perfectamente uniforme en cuanto al sistema de recolección de datos generales, pues de otro modo se obtienen elementos que no pueden ser comparados entre sí. Por ejemplo, si tratamos de buscar el peso medio de un grupo de personas lógicamente debemos clasificarlos de acuerdo con edad y sexo y en ningún caso vamos a tomar el peso de unas personas con vestidos y de otras desnudas ya que nos daría datos que no pueden compararse.

Una vez que se ha revisado el material colectado, viene el trabajo de la tabulación es decir de recuento de resultados, el que a veces es sencillo y puede efectuarse mediante trabajo manual pero que en otras ocasiones por el enorme número de casos requiere el empleo de maquinarias especiales producidas por compañías industriales, máquinas que efectúan la clasificación y contaje de los datos, para lo cual se requiere trasladar dichos datos obtenidos a tarjetas perforadas en las que cada número de cada tarjeta o de cada columna representa una respuesta, de modo que su contaje previa clasificación mecánica proporciona de manera muy fácil resultados que ponen la base para la elaboración de Tablas Maestras o Cuadros Estadísticos.

CUADROS ESTADISTICOS.—Dichos cuadros estadísticos son pues resumen del trabajo efectuado y deben elaborarse de modo tal que su comprensión sea fácil. El dato estadístico se puede presentar al público en forma de texto descriptivo o a manera de presentación tabular es decir mediante tablas en las que se enumeran las diferentes categorías de datos clasificados y sus respectivos valores en columnas. La elaboración de estas tablas debe sujetarse a normas generales que facilitan su captación como por ejemplo deben numerarse, si en una publicación hay varias tablas, debe ponerse claramente el título que expone clara y sintéticamente el problema considerado, debe citarse la fuente de donde se obtuvieron los datos y la fecha, y en la ordenación se debe tratar de colocar los puntos clasificados en orden de importancia. Debe indicarse las unidades que se utilizaron de modo que su valoración sea clara y, finalmente se debe señalar con claridad los subtotales y los totales.

GRAFICOS.—Además se presentan los datos estadísticos mediante gráficos que aunque no sustituyen a las tablas, sirven enormemente para su comprensión pues el dato es más fácilmente captado ya que los elementos notables se presentan en forma visual y objetiva.

En general se pueden señalar hechos importantes que deben tenerse en cuenta tanto para la elaboración de Gráficos como para su lectura, ya que de otro modo es fácil incurrir en graves errores.

Lo establecido es que en la línea vertical u ordenada (Y) se represente la frecuencia o valor de las características estudiadas cuya representación se hace en la abcisa o línea horizontal (X); debe evitarse ante todo la tendencia a alargar o ensanchar demasiado el gráfico por la arbitraria distribución de las unidades al espaciarlos en la abcisa o en la ordenada. Al igual que las tablas estadísticas los gráficos deben ser autoexplicativos, es decir que su comprensión puede hacerse sin recurrir al texto para lo que aparte de títulos y subtítulos debe señalarse la fuente de donde se obtuvieron los datos, su fecha y la significación del tipo de figura usada.

Debe evitarse la aglomeración de datos en un mismo gráfico, pues esto dificulta la interpretación en vez de facilitarla.

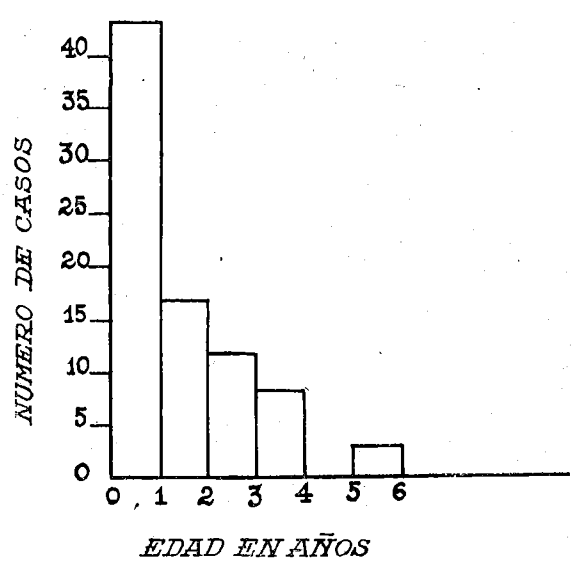
Los tipos de gráficos más usados son el Polígono de frecuencia y el Histograma.

El Polígono de frecuencia se elabora mediante puntos colocados en la frecuencia y en el centro de la característica que representa la abcisa, luego se unen los puntos; en cambio en el Histograma, la frecuencia de cada carácter se representa mediante un rectángulo de modo que, quizás en esta forma se facilite su interpretación. Debe tenerse en cuenta en ambos casos que cierto tipo de gráficos, iguales distancias representan iguales cambios, pues hay otros que se hacen en papel semilogarítmico en los que, iguales distancias representan iguales cambios proporcianales, pues en las líneas de ordenadas se coloca la frecuencia de modo que las distancias sean iguales para un número y su logaritmo.

Debemos también señalar que hay gráficos en los que se representan observaciones de escala contínua y en otros de escala discontínua.

Se llaman series de escala Contínua aquellas en las que se utilizan atributos que pueden ser medidos con unidades más pequeñas, o mejor dicho, cuyas unidades de medida pueden ser divisibles en otras más pequeñas; por ejemplo: al clasificar cualquier característica en in-

**Muertes por Tos ferina por edad en Tennesse U.S. 1.950**




**(HISTOGRAMA)**
**Ejemplo de Histograma con datos de una serie de escala contínua.**

dividuos de diferente edad; hacemos grupos, digamos, de personas de, 1, 2, 3, 4, etc. años de edad; pero ello implica que en cada grupo están personas que difieren en su edad por fracciones menores de un año. La edad podría medirse en meses, y por lo mismo al considerar cada grupo se debe tener en cuenta su amplitud. En el ejemplo, los niños de 2 años tendrán en realidad una edad entre 2 años y 2 años 11 meses.

Son series de Escala Discontínua las que utilizan *caracteres separables y que no tienen valores intermedios*. Por ejemplo: mortalidad por una determinada causa según la raza. Cada grupo es completamente separado del otro sin que haya un paso gradual como en los grupos de edad.

Generalmente la frecuencia (porcentajes o número de casos) se representa en la línea de las ordenadas, y la característica medida en la línea de las abcisas.

Los gráficos tienen modalidades tendientes a facilitar la presentación de los datos. Hay gráficos en área, gráficos sólidos, en mapa, etc. Los más conocidos son el Histograma y el Polígono de Frecuencia.

El Histograma se construye con barras cuya anchura corresponde a la amplitud de cada grupo, y su altura a la frecuencia representada en la línea de las ordenadas. *Es el gráfico más útil y fácil de comprender.*

El Polígono de Frecuencia se elabora colocando puntos en la parte central de cada grupo y en la altura de la frecuencia correspondiente y uniendo luego dichos puntos.

*En la elaboración de gráficos debe tenerse siempre* en cuenta que los grupos en que se ha dividido una observación sean iguales de modo que el gráfico no se deforme por la presencia de grupos más amplios a los cuales les correspondería una frecuencia menor si se dividieran en grupos iguales a los restantes.

## LOS PROMEDIOS

Una vez que los datos de una observación han sido *tabulados y elaborados los cuadros correspondientes*, el siguiente paso de un trabajo estadístico, es el cálculo de valores que informen de los hechos que se producen más frecuentemente en un conjunto de observaciones. Estos valores son los promedios entre los que tenemos: la Media Aritmética, (M); la Mediana (Ma); y el Modo (Mo).

LA MEDIA ARITMETICA.—Es el valor promedial que se calcula dividiendo la suma de los valores encontrados para el número de casos:

$$\overline{M} = \frac{(\Sigma\ X^1 + X^2 + X^3)\ n}{n}$$

Cuando el cálculo de la Media Aritmética se hace en observaciones agrupadas, como en el ejemplo adjunto, hay que señalar el valor promedial del grupo, y así las 32 personas cuyo peso está comprendido, entre 52 y 52,9 kilos, suponemos que pesarán 52,5 kilos, las 35 personas cuyo peso se señala como de 53 kg. pesan entre 53 y 53,9 kg. y se les señala un peso medio de 53,5 kg. y así sucesivamente.

**Cálculo de la Media del Peso de 425 personas**

| 1<br>**Nº Casos** | 2<br>**Peso** | 3<br>**Peso medio *c/ frecuencia*** | 4<br>**Peso total *del grupo*** |
|---|---|---|---|
| 32 | 52 | 52,5 | 1.680,0 |
| 35 | 53 | 53,5 | 1.872,5 |
| 42 | 54 | 54,5 | 2.289,0 |
| 44 | 55 | 55,5 | 2.442,0 |
| 50 | 56 | 56,5 | 2.825,0 |
| 48 | 57 | 57,5 | 2.760,0 |
| 49 | 58 | 58,5 | 2.866,0 |
| 41 | 59 | 59,5 | 2.439,5 |
| 30 | 60 | 60,5 | 1.815,0 |
| 34 | 61 | 61,5 | 2.091,0 |
| 20 | 62 | 62,5 | 1.250,0 |
| 425 | | | 24.330,5 |

Este valor intermedio de cada frecuencia se multiplica por el número de casos de cada grupo.— Así, en el cuadro adjunto tenemos que multiplicar 32 casos X 52; 35 casos X 53,5 y sucesivamente para los grupos restantes.

Los valores resultantes que constan en la columna 4 y que son valores del peso total de las personas de cada grupo, se suman; y así tenemos 24.330,5 kilos que es el peso total de las 425 personas.

La media aritmética se obtiene de dividir este peso total para el número de personas:

$$M = \frac{24.330,5}{425} = 57,25 \text{kg. La Media ha sido: } 57,25 \text{ Kg.}$$

LA MEDIANA.— Es el valor que corresponde a la observación central o media de un grupo, es decir que por encima de la observación a la cual corresponde la mediana y por debajo de ella se encuentra igual número de casos.

Cuando hay pocos casos y cuando las observaciones se refieren a series de atributo discontínuo el valor de la mediana no corresponde estrictamente a la definición.

**Cuadro imaginario de la Distribución del Peso en 425 personas**

| Nº de Casos | Peso en Kgs. |
|---|---|
| 32 | 52 |
| 35 | 53 |
| 42 | 54 |
| 44 | 55 |
| 50 | 56 |
| 48 | 57 |
| 49 | 58 |
| 41 | 59 |
| 30 | 60 |
| 34 | 61 |
| 20 | 62 |
| 425 | |

CALCULO DE LA MEDIANA.—Para el cálculo de la mediana se coloca a las observaciones en orden de menor a mayor acuerdo con sus valores y se aplica la fórmula:

$$\text{MEDIANA} = \frac{\text{N}^\circ \text{ Observac.} + 1}{2}$$

con la cual si es una serie corta de observaciones el cálculo es fácil; por ejemplo, si tenemos 35 casos el cálculo será:

$$\frac{35 + 1}{2} = 18$$

La mediana será la observación Nº 18 por encima de la cual habrán 17 observaciones y otras 17 debajo de ella.

Pero, para calcular la mediana en series agrupadas es preciso hallar el valor de la observación que ocupando el centro de todas se encuentre en un grupo; por ej. en el cuadro adjunto tenemos que aplicar la fórmula

$$\frac{\text{N}^{\circ}\text{ de casos} + 1}{2} = \frac{425 + 1}{2} = 213$$

y obtenemos el número 213 que corresponde a la observación central por encima y por debajo de la cual habrán 212 casos.

La observación Nº 213 será encontrada sumando los casos hasta obtener el grupo en el que se halla: 32 + 35 + 42 + 44 + 50 = 203, si a los 203 sumamos el siguiente grupo de 48 casos excederemos a 213, pues 203 + 48 = 251 por lo cual decimos que el caso Nº 213 se halla dentro del grupo de 48 personas que pesan 57 Kgs. Pero puesto que es una serie de atributo contínuo debemos suponer que no todas estas personas pesan exactamente 57 kg, sino que pesan entre 57 kgs. y 57,9 kgs., se han agrupado juntos pues la escala sólo ha medido diferencias de 1 kg. Si colocáramos a estas 48 personas en orden estricto de peso hallaríamos el que corresponde exactamente a la Nº 213 para lo cual debemos encontrar en qué lugar del grupo de 48 está ubicada; es decir: 32 + 35 + 42 + 44 + 50 = 203, faltando 10 para llegar a 213, porque 213 — 203 = 10; es decir que la persona Nº 10 del grupo de 48 es la 213 en todo el conjunto de observaciones y a ella corresponde el valor de la mediana el cual será igual a 57 kgs. que es el peso de todas las personas del grupo, más la fracción de kilogramo que le corresponde a la Nº 10, lo cual se halla dividiendo:

$$\frac{10}{48} = 0{,}21$$

, o sea 57 + 0,21 = 57,21 kgs. De modo que el valor de la mediana será 57,21 kgs.

Resumiendo las operaciones tenemos:

$$\frac{\text{N}^{\circ}\text{ de Casos} + 1}{2} = \frac{425 + 1}{2} = 213$$

(observación central)

Sumando los grupos 32 + 35 + 42 + 44 + 50 = 203; 213 — 203 = 10, como estos 10 casos están en el grupo de 48 que pesa n57 kgs. tenemos:

$$\frac{10}{48} = 0{,}21,$$ que es la fracción de kilo que se suma a 57 y tenemos el valor de la mediana 57,21 kgs.

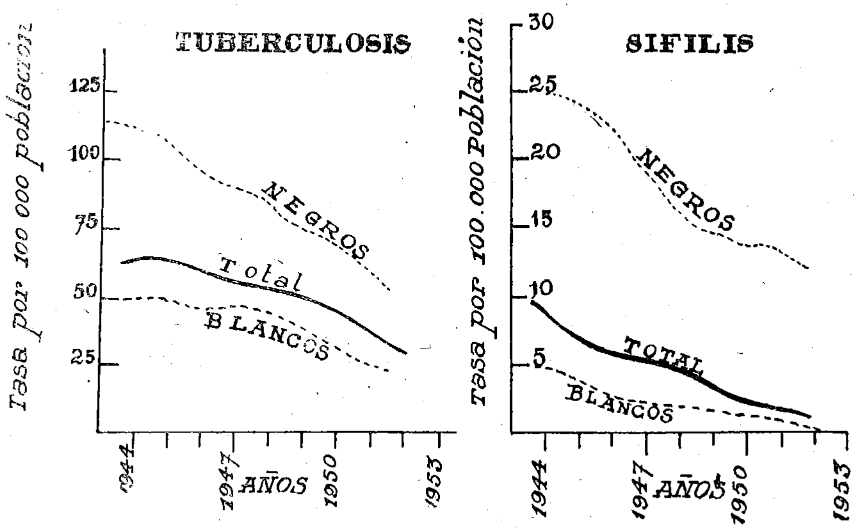
EL MODO.— Llamado también valor dominante, es en muchos casos muy útil si al calcular la media, ésta se desvía demasiado por la existencia de pocas observaciones extremas, su valor no dará una idea precisa de la realidad; por ej. al calcular la duración media de tiempo de incubación de una enfermedad, o su duración, o días de permanencia en un servicio hospitalario, etc., pocas observaciones con valores elevados, es decir con permanencia de muchos días podrían hacer que la mediana dé una idea exacta de la duración usual de la incubación o la duración de la enfermedad; en estos casos el cálculo del modo o valor dominante es muy útil pues informa de la manera más usual de presentarse una serie de observaciones, o mejor dicho lo que se presenta más comunmente, con más frecuencia. Ej. la duración más frecuente de una enfermedad. Su cálculo se basa en la fórmula:

$$Mo = M - 3\ (M - Ma).$$

MODO = MEDIA ARITMETICA — 3 veces la diferencia entre media y mediana.

De modo que en el ejemplo del cuadro anterior en que:

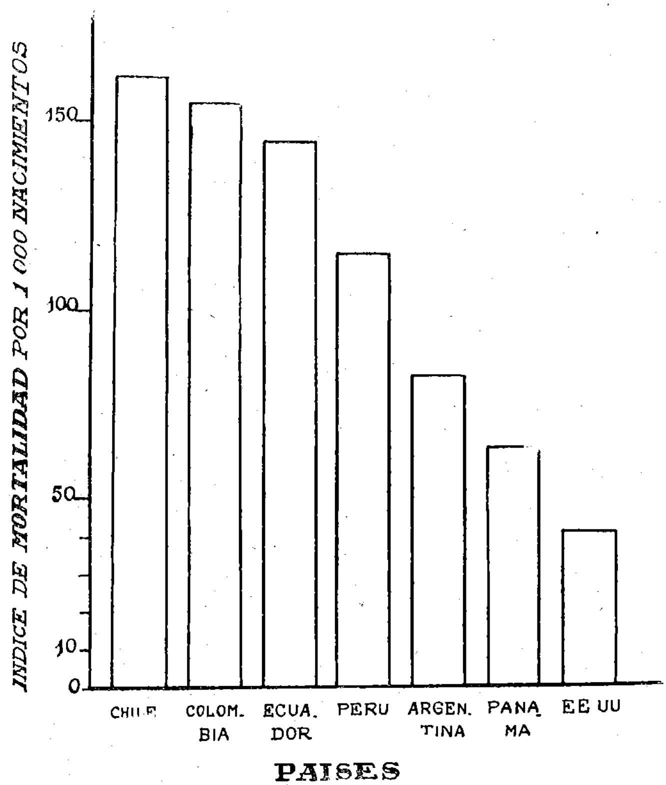
M=57,25kgs. tenemos Mo=57,25—3 (57,25—57,21)
Ma=57,21 Kgs, Mo=57,25—3 (0,04).
Mo=57,25—0,12=57,13




**Tasas de Mortalidad por Tuberculosis y Sífilis, según la raza. Por 100.000 de población.**

**Tennessee U.S. 1944-1953**

**Ejemplo de Serie Discontinua.**
**Gráfico comparativo de Mortalidad Infantil en varios países de América. 1946**




**Ejemplo de Histograma con datos de una serie de Escala Discontinua.**

VARIABILIDAD.— Es costumbre en las publicaciones médicas, referirse únicamente a los valores promediales o porcentajes obtenidos; sin embargo estos datos no informan al lector de la forma cómo las observaciones se han distribuído. La distribución constituye un dato valioso no sólo para el mejor conocimiento del fenómeno estudiado sino también para poder compararlo con otros similares.

Si decimos que la Presión Arterial Mx. promedial es de 120 mm. y la Mn. de 80 mm. de Hg. al hacer un estudio de tan importante elemento fisiológico, sólo estamos mencionando un valor que se encuentra en el centro de equilibrio del grupo estudiado, pero no sabemos de qué límites ha venido.

Si después de estudiar el peso en un grupo de niños de 6 años (Cuadro IV) concluímos que su peso promedio es de 18 Kg. sólo tendremos un conocimiento parcial de lo estudiado, pues no sabemos dentro de qué límites varía el peso de los niños de esta edad en nuestro medio, y aún más, si con este dato queremos juzgar acerca del estado nutricional de un niño que acude a nosotros y pesa 15 Kg. siendo de la misma edad, no estaremos en condiciones de poder asegurar nada respecto de la normalidad de su peso; pero si decimos que el peso de los 511 niños de 6 años oscila entre 12 y 26 Kg. es decir con un rango de 15 (pues cada grupo difiere del siguiente en 1 Kg.) ya tendremos una idea más cercana de lo que el valor 15 Kg. significa como peso del niño examinado.

EL RANGO.—Es pues la distancia entre la observación mayor y la menor.

En ocasiones es también insuficiente pues sólo informa acerca de los casos extremos y nó de cómo los casos se distribuyen alrededor de la media.

Esto es lo que informa, lo que se denomina Desvío Tipo o Desviación Estandar.

DESVIACION ESTANDAR O DESVIO TIPO.—Representado por la letra griega sigma y abreviado por D. T. o por S. D. es una buena medida de la distribución o variabilidad de las observaciones alredor de la media.

Su cálculo en observaciones no agrupadas difiere algo del que se realiza en las observaciones agrupadas.

Cuando tenemos pocas observaciones que no están agrupadas se lo efectúa con la fórmula

$$\sigma = \sqrt{\frac{\Sigma \ (\text{Desviación del Promedio})^2}{N}}$$

Es decir que la Desviación Estandar es igual a la raíz cuadrada de la suma de los cuadrados de las desviaciones de la media dividida para el número de casos.

Veamos su cálculo en el ejemplo adjunto (Cuadro Nº III) en el que tenemos la talla en centímetros de 23 señoritas de 22 años.

La talla promedio de las 23 señoritas es de 152 centímetros porque:

$$M = 3496 \div 23 = 152 \text{ centímetros.}$$

En la segunda columna del cuadro III, tenemos las cantidades en que cada observación se desvía de la Media; y por eso colocamos el 0 frente a 152;

—1 frente a 151 ya que el desvío con relación a la Media será en — 1 ctms.; — 3 frente a 149 y así sucesivamente. Las observaciones cuyo valor es inferior a la Media, tienen lógicamente signo (—) negativo y las observaciones cuyo valor es mayor son positivas. Puesto que la suma de estas cantidades daría 0, se las eleva al cuadrado de modo que desaparecen los signos (—) en las cantidades de observaciones menores a la M. y se suman estos cuadrados, (columna 3) lo que da un total de 850, que debe dividirse para el número de casos:

$$\div \ 23 = 36{,}95$$

de lo que debe extraerse la raíz cuadrada ya que eran cantidades que fueron elevadas al cuadrado.

$$\alpha = \sqrt{36{,,}95} = 6{,}09$$

En síntesis para obtener la Desviación Estandar de observaciones no agrupadas se debe:

## CUADRO III

**Talla en centímetros de 23 señoritas de 22 años Universidad Central — Quito-1954**

| Talla en ctms. | Desviación del promedio | Desviaciones del promedio al cuadrado |
|---|---|---|
| 143 | — 9 | 81 |
| 144 | — 8 | 64 |
| 146 | — 6 | 36 |
| 146 | — 6 | 36 |
| 146 | — 6 | 36 |
| 147 | — 5 | 25 |
| 148 | — 4 | 16 |
| 148 | — 4 | 16 |
| 148 | — 4 | 16 |
| 149 | — 3 | 9 |
| 151 | — 1 | 1 |
| 152 | 0 | 0 |
| 153 | 1 | 1 |
| 153 | 1 | 1 |
| 153 | 1 | 1 |
| 153 | 1 | 1 |
| 154 | 2 | 4 |
| 155 | 3 | 9 |
| 156 | 4 | 16 |
| 157 | 5 | 25 |
| 158 | 6 | 36 |
| 166 | 14 | 196 |
| 170 | 18 | 324 |
| 3.496 | | 850 |

1º—Hallar la Media Aritmética (en el ejemplo 152 ctms.).

2º—Hallar las desviaciones de cada observación con relación a la Media (columna 2 del cuadro III).

3º—Elevar al cuadrado dichas desviaciones (columna 3).

4º—Sumar dichos cuadrados (850) y dividir el producto de dicha suma por el número de observaciones

$$\left( \frac{850}{23} = 36,95 \right)$$

5º—Extraer la raíz cuadrada

$$\sqrt{36.95} = 6,09$$

CORRECCION DE LA DESVIACION ESTANDAR. Cuando se ha calculado la Desviación Estandar del promedio en series cortas de observaciones, es preciso hacer una corrección de modo que sea un valor más preciso de la variabilidad de las obrervaciones.

Esta corrección consiste en multiplicar $\sigma$ obtenida por el procedimiento descrito

$$\times \sqrt{\frac{n}{n-1}}$$

$\sigma \sqrt{\frac{n}{n-1}}$ **en el caso anterior 6,09** $\sqrt{\frac{23}{23-1}}$

$$6,09 \sqrt{1.045} = 6,09 \times 1.02 = 6,2118$$

$$\sigma = 6,2118$$

CALCULO DE LA DESVIACION ESTANDAR EN SERIES AGRUPADAS.—Para calcular el Desvío Tipo o

Desviación Estandar en series agrupadas puede emplearse el siguiente procedimiento:

Elíjase un grupo cercano al del Promedio y frente a él póngase un **0** a partir del cual se numerarán los desvíos anteriores con signo negativo (—) y los siguientes con signo positivo.

En el cuadro IV B tenemos el 0 convencional frente al grupo que pesa 18 Kgs. y numeramos al grupo de 17 Kgs. como —1; al de 16 Kg. como —2 y así sucesivamente; los grupos siguientes como el de individuos que pesan 19 Kgs. tienen un desvío positivo y por eso le ponemos + 1; el de 20 Kgs. será 2; el de 21 Kgs. será 3 y así los restantes.

Este desvío debe llevarse al cuadrado y dicho cuadrado multiplicarse por el número de casos de cada frecuencia y así tendremos en el ejemplo:

$-6^2 = 36 \times 1 = 36$ y en el grupo siguiente: $-5^2 = 25 \times 6 = 150$; en el siguiente $-4^2 = 16 \times 23 = 368$ y así para los restantes, pero un procedimiento más simple y que por lo mismo facilita el cálculo de $\sigma$ consiste en multiplicar el desvío (columna 4) por el número de casos de esta frecuencia (columna 3) y así tenemos una serie de valores FX cuya suma ( $\Sigma$ FX) dividida para el número de casos nos dá un promedio "en unidades de cálculo" que lo representamos por el signo $\gamma_1$ . A su vez estos valores de la columna 5 ó FX al ser multiplicados por la desviación dan un valor igual al del cuadrado de dichas desviaciones X el número de casos; y así tenemos:

$$-6 \times -6 = 36$$
$$-5 \times -30 = 150$$
$$-4 \times -92 = 368, \text{ etc.}$$

Los valores de esta columna ($FX^2$) se suman ( $\Sigma$ $FX^2$) y su producto se divide para el número de casos obteniéndose un segundo promedio que lo llamamos $\gamma_2$

Si tenemos en cuenta que la desviación Estandar en estas series agrupadas se obtiene con la fórmula:

$$\sigma = \sqrt{\gamma_2 - \gamma^1{}_2}\ \text{xm.} \qquad \text{en donde,}$$

$$\gamma_2 = \frac{\Sigma\, FX^2}{N} \qquad y \qquad \gamma_1 = \frac{\Sigma\, FX}{N}$$

y m el modelo o amplitud de las frecuencias que en este caso es de 1 Kg. (cada grupo tiene una amplitud de 1 Kg.).

Podemos reemplazar y tenemos:

## CUADRO IV A.

**Cálculo de la Media Aritmética del Peso de 511 niñas de 6 años.— Quito - 1954**

| Peso en Kgs. (1) | Peso Medio de c/ frecia (2) | Nº de Casos (3) | Peso total de c/ frecuencia (2 x 3) |
|---|---|---|---|
| 12 | 12,5 | 1 | 12,5 |
| 13 | 13,5 | 6 | 81,0 |
| 14 | 14,5 | 23 | 333,5 |
| 15 | 15,5 | 47 | 728,5 |
| 16 | 16,5 | 80 | 1.320,0 |
| 17 | 17,5 | 96 | 1.680,0 |
| 18 | 18,5 | 98 | 1.813,0 |
| 19 | 19,5 | 69 | 1.345,5 |
| 20 | 20,5 | 50 | 1.025,0 |
| 21 | 21,5 | 19 | 408,5 |
| 22 | 25,5 | 11 | 247,5 |
| 23 | 23,5 | 5 | 117,5 |
| 24 | 24,5 | 3 | 73,5 |
| 25 | 25,5 | 2 | 51,0 |
| 26 | 26,5 | 1 | 26,5 |
| | | 511 | 9.163,5 |

$$M = \frac{9.163.5}{511} = 17{,}93 \text{ Kg.}$$

## CUADRO IV B

**Cálculo de la Desviación Estandar de la Media del Peso de 511 Niñas de 6 años.—Quito - 1954**

| Peso en Kgs. | Peso medio de cada grp. | Nº de Casos F | Desviación X | F X (3 x 4) | F $X^2$ (4 x 5) |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 12 | 12,5 | 1 | — 6 | — 6 | 36 |
| 13 | 13,5 | 6 | — 5 | — 30 | 150 |
| 14 | 14,5 | 23 | — 4 | — 92 | 368 |
| 15 | 15,5 | 47 | — 3 | — 141 | 423 |
| 16 | 16,5 | 80 | — 2 | — 160 | 320 |
| 17 | 17,5 | 96 | — 1 | — 96 | 96 |
| 18 | 18,5 | 98 | 0 | 0 | 0 |
| 19 | 19,5 | 69 | 1 | 69 | 69 |
| 20 | 20,5 | 50 | 2 | 100 | 200 |
| 21 | 21,5 | 19 | 3 | 57 | 171 |
| 22 | 22,5 | 11 | 4 | 44 | 176 |
| 23 | 23,5 | 5 | 5 | 25 | 125 |
| 24 | 24,5 | 3 | 6 | 18 | 108 |
| 25 | 25,5 | 2 | 7 | 14 | 98 |
| 26 | 26,5 | 1 | 8 | 8 | 64 |
| | | 511 | | 190 | 2.404 |

$$\sigma = \sqrt{\frac{\Sigma FX^2}{N} - \left(\frac{\Sigma FX}{N}\right)^2} \times M$$

que con los valores del cuadro IV B será:

$$\gamma_1 = \frac{\Sigma FX}{N} = \frac{190}{511} = 0,371$$

$$\gamma_2 = \frac{\Sigma FX^2}{N} = \frac{2.404}{511} = 4,704$$

y por tanto $\sigma = \sqrt{\gamma_2 - \gamma_1^2 \times M}$ que al reemplazar, da:

$$\sigma = \sqrt{4,704 - 0,371^2 \times 1}$$

$$\sigma = \sqrt{4,56656} = 2,136$$

$$\sigma = 2,136$$

VALOR DE LA DESVIACION ESTANDAR.—Hemos dicho que la Desviación Estandar da una idea precisa de la variabilidad de las observaciones y que señala el límite dentro del cual pueden variar los promedios obtenidos en otras muestras.

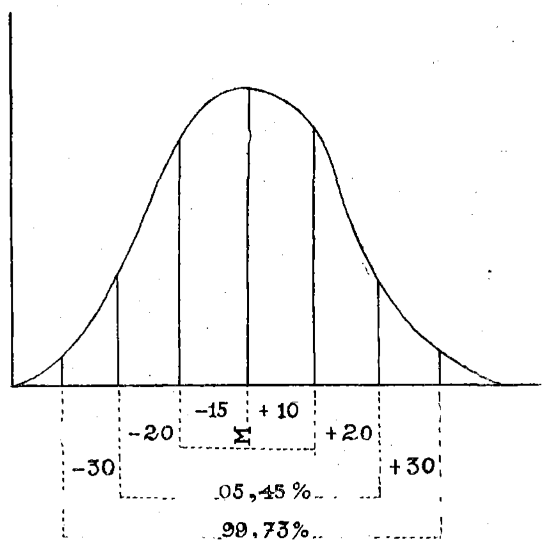
En efecto, si en una curva de "distribución normal" vemos el número de observaciones que caen al rededor de la Media en relación con el Desvío Tipo o Desviación Estandar, comprobaremos que:

1º—Un 68,27% de observaciones se encuentran alrededor de la Media, por dentro de una Desviación Estandar a cada lado del promedio ($\pm$ 1 D. E. = 68,27%)

2º—Un 95,45% se encuentra por dentro de 2 Desviaciones Estandar a cada lado de la Media ($\pm$ 2 D. E. =

3º—Un 99,73% de observaciones se hallan por dentro de 3 Desviaciones Estandar a cada lado de la Media ($\pm$ 3 D. E. = 99,73%).

Por esto es que se acepta como caso muy raro el hecho de que la Media de una muestra difiera de la ver-




dadera media del universo en más de 2 Desviaciones Estandar ($\pm$ 2 D. E.) calculadas en la muestra.

Es decir que si por ejemplo encontramos que el peso promedial de todas las niñas de 6 años de Quito es de, supongamos, 20 Kgs. y en nuestras observaciones encontramos, un peso promedial para niñas de esta edad de 17,9 Kgs. diremos que la diferencia entre ambos promedios no tiene "significación estadística", pues no es mayor de 2 D. E. (En la muestra de 511 niñas tenemos una D. E. = 2,136), pero si el peso promedial de todas las niñas de 6 años fuera de 23 Kgs. y en nuestra muestra tenemos 17,9 Kg., la diferencia sería "estadísticamente significativa" por ser mayor que 2 D. E.; ya que

$\sigma$ = 2,136 Kg.; 2 $\sigma$ = 4,272 Kgs.

y la diferencia entre 23 Kgs. y los 17,9 Kgs. de nuestro promedio sería de 5,1 Kgs. que es mayor que 4,272 Kgs.

Ello querría decir que la diferencia entre el promedio del universo (23 Kg.) y el de nuestra muestra (17,9 Kgs.) no se debe probablemente al azar, sino a algún otro factor que debe ser investigado claramente, por ejemplo, desnutrición, clima, enfermedades, etc.

Las diferencias entre promedios, que son menores a 2 D.E. deben aceptarse como debidas al azar aunque podrían deberse a otros factores especialmente en observaciones con reducido número de casos.

El cálculo de la D. E. y la relación que con ellos se deriva para apreciar diferencias entre promedios y porcentajes de observaciones similares y comparables es muy útil en las investigaciones médicas, ya que frecuentemente se comparan índices de mortalidad, índices biométricos, porcentajes de éxitos y fracazos con determinados métodos terapéuticos, etc. Muchas veces el investigador malogra valiosas experiencias con conclusiones erróneas por la falta de este cálculo.

LA DESVIACION ESTANDAR DE PROPORCIONES.— Como hemos dicho, la teoría del Desvío Tipo o Desviación Estandar se fundamenta en la de la Variabilidad y por consiguiente en la del cálculo de probabilidades. Así como para calcular la probabilidad de que una moneda caiga con el sello hacia arriba empleamos la fórmula P = ½ que significa que hay 50% de que esto suceda y si el número de monedas aumenta, aumenta también la probabilidad de que todas caigan de sello, lo cual será igual al producto de las probabilidades de cada moneda: $P = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \ldots (\frac{1}{2})^n$; así también puede calcularse la probabilidad de los coeficentes y proporciones que tan empleados son en investigaciones médicas.

Para el cálculo de la D. E. de las proporciones, es decir, de la variabilidad de las mismas, tenemos la fórmula siguiente:

$$\sigma = \sqrt{\frac{p\,q}{n}}$$

en la cual p y q indican las dos probabalidades del evento, es decir p es la probabilidad de que un hecho suceda; y q la probabilidad opuesta o sea de que no suceda.

Si decimos que el 20% de los operados de Lobectomía pulmonar mueren, esta es una probalidad; y la otra será del 80% que no mueren.
n = es el mismo de observaciones. De modo que en el ejemplo:

$$\sigma = \sqrt{\frac{20 \times 80}{100}} = 4$$

Lo que significa que la mortalidad por esta intervención tiene una D. E. de 4 y de acuerdo con lo dicho antes, una variación de ± 8% no tendrá una importancia estadística.

DESVIACION ESTANDAR DIFERENCIAL.—Hemos visto que puede calcularse la D. E. de una proporción obtenida al investigar en una muestra con el objeto de demostrar la validez de la diferencia con semejantes proporciones obtenidas del universo estudiado.

Pero cuando no es posible obtener proporciones de todo el conjunto o universo, lo que es frecuente, sino que al investigar emplea en un estudio muestras de tamaño mayor o menor, se puede calcular la D. E. de la diferencia de las dos proporciones encontradas en las muestras, para de este modo demostrar si la dicha diferencia se debe al azar o a otros factores que influyendo sobre el fenómeno estudiado en un grupo, modifican los resultados.

Si, por ejemplo, observamos 50 individuos que al recibir tratamiento de amebiasis con emetina, han sido curados en un 80% y comparamos este resultado con el obtenido al tratar otros 60 casos similares de amebiasis con otra droga, supongamos terramicina, cuyo porcentaje de curados ha sido 90%; debemos establecer si la diferencia entre estos dos porcentajes se debe al azar o realmente ha influído el tratamiento.

La D. E. de los tratados con emetina se calculará con la fórmula:

$$\sigma = \sqrt{\frac{p\ q}{n}} \qquad \sigma = \sqrt{\frac{80 \times 20}{50}}$$

e igualmente la de los tratados con terramicina:

$$\sigma \sqrt{\frac{p\,q}{n}} = \sigma \sqrt{\frac{90 \times 10}{60}}$$

y la D. E. de la diferencia entre las dos proporciones se calculará con la fórmula:

$$\sigma_{\text{dif}} = \sqrt{\frac{p\,q}{n_1} + \frac{p_2\,q_2}{n_2}}$$

en la que p y q son las probabilidades de curación o fracaso con cada tratamiento y $n_2$ y $n_1$ el número de casos en cada observación, que en el ejemplo es de 50 para el 1 grupo y 60 para el otro.

Reemplazando tenemos: $$\sigma_{\text{dif}} + = \sqrt{\frac{80 \times 20}{50} + \frac{90 \times 10}{60}}$$

$$\sigma_{\text{dif}} + = \sqrt{32 \times 15} = 6{,}8$$

Siendo la D. E. de 6,8, y la diferencia entre los dos porcentajes encontrados de 10 (porque 90% — 80%) tenemos que la diferencia entre los dos porcentajes es menor que 2 DE y por lo mismo puede ser debida al azar, de modo que deberíamos concluír que se necesita investigar en un número mayor de casos para sacar conclusiones más valederas.

DIFF = 10%.

$\sigma$ Dif. = 6,8; 2 $\sigma$ Dif = 13,6 y 10% es menor que 13,6.

+++++++