

Esquema de los Métodos Estadísticos

Por el Ingeniero

EMILIO ALANIS PATIÑO

Irrigación en México publicó en sus dos números anteriores un artículo del ingeniero Emilio Alanís Patiño, titulado "La Estadística al Servicio de los Aprovechamientos Hidráulicos", y otro sobre "Estadísticas de Costos Unitarios". Varios compañeros, refiriéndose a estos artículos, manifestaron deseos de que el autor diera mayores explicaciones acerca de los métodos estadísticos, y este es el objeto del presente artículo y dos más que le seguirán.

I. Elementos y Naturaleza

EL ingeniero que estudia la posibilidad de aprovechar las aguas de un río, necesitará conocer el régimen de la corriente entre otras cosas; pero para entender bien el problema que trata, será preciso que conozca por lo menos los principios fundamentales de la Hidrología, así como las características meteorológicas, topográficas, geológicas, agrológicas e hidrológicas de la cuenca respectiva. Un cúmulo de teorías generales y de observaciones regionales obtenidas, interpretadas y aplicadas convenientemente, permitirán conocer con exactitud satisfactoria, los elementos básicos para proyectar las obras que garantizarán el mejor aprovechamiento de aquella corriente de aguas. El ingeniero tendrá éxito en la medida que disponga de informes o conocimientos y en el grado que sepa emplearlos racionalmente.

Indispensables para el conocimiento del régimen fluvial son los datos hidrométricos, resultantes de muchas observaciones realizadas en una o en varias estaciones, de manera uniforme, sin interrupciones o lagunas, y registrados o anotados sistemáticamente. Los datos serán mejores mientras más cuidadoso sea el trabajo de observación y más prolongado sea el período de observación.

Los frutos de esta lenta labor pueden exponerse en gráficas y en cuadros estadísticos con arreglos distintos, que en sí mismos no constituyen un fin, sino una información que debe interpretarse mediante una técnica especial, para sintetizar una ley, o sea el régimen del río, y sus probables desviaciones.

Conocimientos importantes para el estudio de una cuenca, se obtienen de múltiples observaciones termoplumiométricas y de evaporación, realizadas en varias estaciones distribuidas convenientemente dentro de la cuenca, de acuerdo con procedimientos y aparatos uniformes, sin discontinuidad y anotados sistemáticamente. Esta labor cooperativa produce cifras que se pueden exponer en gráficas y cuadros de tipo estadístico, más o menos elaborados, disponibles para usos diversos, pero adecuados para un tratamiento matemático que conduzca a su interpretación general. El ingeniero hidrólogo tiene un apoyo valioso en los datos meteorológicos, y aun puede cuantificar las relaciones que hay entre lluvias, temperaturas, evaporaciones y volúmenes escurridos.

Los datos hidrométricos y termoplumiométricos son ejemplo de lo que constituye el material con que trabaja la Estadística, cuyo campo de aplicación es tan vasto como lo son los hechos y las actividades en que es posible hacer y registrar observaciones numerosas, en forma sistemática. Centenas o millares de muestras de cemento, sometidas a una serie de pruebas bajo un proceso uniforme, aportan informaciones numéricas de tipo estadístico que son insustituibles para juzgar de la homogeneidad y las características medias del lote. Las centenas o millares de trabajadores ocupados en la construcción de una obra pueden ser motivo de observación en sus características personales, desde puntos de vista muy diversos, como los de edades, ocupaciones, cuotas de salarios, tiempo trabajado, antigüedad, estado de salud, etc., produciendo una masa considerable de informes, que en forma global sólo pueden tratarse y aprovechar-

se mediante la técnica estadística. El universo y la actividad humana ofrecen un variadísimo y muy extenso campo, cuyo conocimiento se obtiene o se precisa mediante los métodos estadísticos.

La observación estadística se hace correctamente sirviéndose de métodos experimentados en el curso de milenios y mejorados notablemente en los últimos 50 años. Ahora los métodos estadísticos constituyen una técnica especial, en la que suele separarse la Estadística Metodológica y la Estadística Aplicada, subdividiéndose esta última en tantas ramas como actividades o campos hay en que es posible observar un conjunto numeroso de casos individuales. La Estadística en ambas categorías abarca desde las normas para una correcta obtención y presentación de los datos, hasta el aprovechamiento del material numérico y gráfico, en distintos grados de intensidad.

Las estadísticas son números que no deben manejarse únicamente por reglas aritméticas, sino mediante procesos lógicos que tengan en cuenta el origen, el significado y el alcance de cada cifra. Las simples listas o relaciones de datos individuales, cualitativos o cuantitativos, generalmente son la fuente inicial de las estadísticas, pero en sí mismas no son estadísticas. Las cantidades que figuran en documentos o estados contables, con frecuencia también se aprovechan para formar estadísticas, pero éstas adquieren su carácter sólo cuando se ordenan, agrupan o relacionan aquellas cantidades según determinados procedimientos. Por otra parte, las estadísticas se expresan no sólo en unidades monetarias, sino también en unidades lineales, de superficie, de volumen, de peso, de tiempo, de calor, de electricidad, etc., o en unidades específicas como personas, caballos, automóviles, lápices, etc., dando lugar a lo que en Estadística se llaman frecuencias o número de casos, o bien se expresan en unidades compuestas como hora-obrero, tonelada-kilómetro, hectárea-riego, metro cúbico por segundo, kilogramos por centímetro cuadrado, etc.

La comparación o la simple suma de cantidades, sólo puede hacerse cuando están expresadas en unidades iguales; por esto en Estadística es frecuente convertir pulgadas en centímetros, millas cuadradas en kilómetros cuadrados, barriles de petróleo, cemento o harina en toneladas, fanegas de sembradura en hectáreas, acre-pie en hectárea-metro, etc. Las unidades monetarias son diversas de país a país, y pueden convertirse unas en otras mediante los tipos de cambio, pero ocurre que en un mismo país la moneda varía de poder adquisi-

tivo, es decir, tiene una magnitud diferente de unas épocas a otras. Este problema adquiere gran importancia en series estadísticas referentes a largos períodos de tiempo, y se ha resuelto en distintas formas, siendo principales la reducción de cantidades monetarias a un patrón como el dólar oro o el franco oro, o la reducción a monedas del poder adquisitivo observado en determinada época, obteniéndose lo que se llama series deflacionadas. Se debe recordar que una divisa monetaria tiene un poder adquisitivo en el interior del país y otro poder adquisitivo en el exterior, pudiendo variar paralelamente o no; según el objeto, será necesario deflacionar los datos monetarios con respecto a una base del poder adquisitivo en el interior, o con relación al poder adquisitivo exterior.

Se ha dicho que las estadísticas dan una historia numérica, o bien que son cifras históricas, y efectivamente abundan los datos que cuantifican el desarrollo de un hecho a través del tiempo, y son muchas las estadísticas que revelan situaciones estáticas de una época determinada. Propiamente, toda estadística se refiere a un hecho o actividad que ocurre en algún lugar y en cierta época o durante determinado período. La producción de hierro en el país puede conocerse día por día, o mensualmente o cada año o por períodos de tiempo diversos, pero también puede darse como promedio por día, por mes o por año, o bajo la forma acumulada sucesivamente desde el 1º de enero hasta el día último de cada mes. Los precios mensuales del trigo se pueden obtener como promedio de observaciones diarias, o de registros semanales o simplemente como precios de cada día 15, y pueden referirse a la ciudad de México o a Chicago o a Buenos Aires, etc. Los volúmenes excavados con dragas de la C. N. I. también pueden referirse a distintos períodos de tiempo y a obras o lugares diversos, y así ocurre con muchísimos hechos.

Un dato puede corresponder a determinado centro de población (ciudad, villa, pueblo, etc.) o parte de él (cuartel, colonia, manzana, etc.), a cierta jurisdicción territorial administrativa (municipio, distrito rentístico, distrito electoral, distrito judicial, cantón, provincia, Estado, etc.), a especiales divisiones dentro de un país (provincia climatérica, cuenca hidrográfica, región natural, región económica, región agrícola, zona militar, zona estadística, etc.), a países diversos y aun a grupos de naciones. Estas múltiples divisiones geográficas con frecuencia tienen límites indefinidos, o señalados empíricamente, o con criterios carentes de uni-

formidad y de estabilidad. Muchos municipios aumentan o disminuyen su jurisdicción territorial, y varias naciones también modifican sus fronteras por efecto de guerras o de presiones políticas. Algo similar ocurre con las obras de la C. N. I., o con distritos de riego, que en ocasiones cambian la amplitud de su jurisdicción territorial.

La separación de datos por unidades geográficas es sólo uno de los problemas que se presentan para clasificar personas, animales, cosas, hechos, actividades, ideas, etc. En la vida real no existen divisiones netas y grupos aislados de categorías diversas, como hombres de estatura alta y baja, vacas y terneras, tierras arenosas y arcillosas, etc., pero el hombre ha necesitado introducir artificialmente clases, categorías o grupos relativamente homogéneos, respecto a cierta característica, con el fin de facilitar su estudio. Cualquier texto de Mineralogía, Geología, Botánica, Zoología o Antropología, se apoya en dos conocimientos básicos, que son su terminología y su clasificación. Las actividades ingenieriles, médicas, u otras profesionales, así como las comerciales e industriales, se desarrollan sobre la base de terminologías y clasificaciones especiales, auxiliares poderosos en la evolución de su técnica. La clasificación es la base lógica de la comparabilidad, puesto que correctamente sólo son comparables dos o más cosas similares en algún aspecto, es decir, pertenecientes a la misma categoría o clase. La Estadística halla en muchos campos, clasificaciones formadas y obligatorias, pero en otras ocasiones debe formular clasificaciones propias. Cuando se forman estadísticas con clasificaciones adoptadas para otros fines, se corre el riesgo de subordinarse a cambios que dificultan o hacen imposible las comparaciones, y cuando la Estadística adopta clasificaciones propias, se estará en peligro de que éstas sean poco útiles para otros fines. En toda clasificación se presentan tres problemas, que al resolverse dan origen a múltiples soluciones: 1º, su finalidad; 2º, su criterio básico; 3º, su grado de subdivisión que puede ser hasta llegar a los casos individuales.

La Estadística trabaja con cantidades expresadas en unidades muy diversas, referentes a épocas distintas, a divisiones geográficas diferentes o a grupos y categorías diversas, pero además, trabaja con cifras relativas, derivadas de las anteriores. Los tanto por ciento, tanto por mil o tanto por diez mil se emplean frecuentemente para facilitar las comparaciones. Los números índices simples, sintéticos y compuestos son otras expresiones de los números

relativos y sirven para comparar una situación simple o compleja en determinada época o lugar, con la situación semejante ocurrida en otra época o lugar que se toman como base. Las medias aritméticas, geométrica y armónica también pueden considerarse como valores relativos, por obtenerse de cifras absolutas. La densidad demográfica, o sea el cociente entre el número de habitantes y la superficie total del territorio que ocupan, es otra cifra relativa, e igualmente lo son otros cocientes de dos cantidades, de las que por lo menos una es estadística. En fin, la Estadística trabaja con valores obtenidos mediante procesos de elaboración más complicada, como las desviaciones cuadráticas medias absolutas y relativas y otras medidas de la variabilidad, los índices de concentración, las tendencias y los índices de variación estacional, los índices de concordancia entre los que se destacan las medidas de correlación lineal y curvilínea o parcial y múltiple, etc., etc.

Existen muchas situaciones y actividades que se observan estadísticamente en todas sus partes y unidades, mientras que otras sólo es posible conocerlas en algunas de sus partes o en cierto número de unidades que representan una porción más o menos grande del total. En este último caso los datos disponibles adquieren un significado especial, pues pueden ser representativos o no, del conjunto a que pertenecen; serán representativos cuando sus características sean iguales o sumamente parecidas a las del grupo de donde fueron obtenidos, pero los datos totales se ignoran, de modo que el grado de representatividad de la muestra sólo puede estimarse por procedimientos matemáticos basados en el Cálculo de las Probabilidades. Pocas personas están capacitadas para manejar datos de muestras, y sin embargo, muchas hay que los emplean, generalizando sus indicaciones.

Alguien que desee aprovechar debidamente las estadísticas debe tener en cuenta todo lo expuesto en los párrafos anteriores, pero además debe disponer de cuatro informaciones fundamentales: 1ª, de dónde proceden las estadísticas; 2ª, para qué se hicieron; 3ª, por qué procedimiento se obtuvieron; 4ª, qué grado de exactitud y eficiencia se les puede atribuir. Las estadísticas proceden de fuentes diversas, más o menos calificadas para formarlas, según su naturaleza, su organización, sus recursos técnicos y económicos, su fuerza moral, etc.; importa conocer la fuente de cada estadística y sus condiciones de trabajo. El fin que

determinó la formación de una estadística, puede señalar su naturaleza y sus características. Los procedimientos que se siguieron para obtener una estadística, indican hasta cierto punto el grado de confianza que se le puede conceder. Sin embargo, el juicio definitivo sobre el grado de exactitud y eficiencia de las estadísticas, puede darse hasta después de haber examinado y aun utilizado dichas estadísticas. Estas cuatro informaciones se deben tener presentes siempre que se usen estadísticas formadas por otras personas. Habrá casos en que sea del todo inconveniente su empleo, y otros en que las condiciones deben formularse con reservas.

En una estadística formada correctamente pueden introducirse errores graves durante su escritura mecánica, su edición mimeográfica o su impresión tipográfica. Si errores frecuentes de este tipo se agregan a defectos originales de las estadísticas, éstas pueden llegar a ser totalmente inútiles o aprovechables con dificultad y peligrosamente. Casi todas las cifras estadísticas están destinadas a publicarse en revistas, anuarios, folletos, periódicos, informes y otros documentos, de modo que el peligro de errores tipográficos es muy grande. Los países con buena tradición en sus servicios estadísticos y en sus industrias impresas, logran casi anular los errores tipográficos de los cuadros estadísticos. En este campo, los errores más comunes son: el cambio de una o varias cifras, en ocasiones alternándose, como 58 por 85; el cambio completo de un renglón por otro, o de una columna por otra, pero no de los encabezados; la adición u omisión de alguna cifra, especialmente a la izquierda de alguna cantidad; la omisión o colocación incorrecta del punto decimal; la sustitución del 0/0 por el 0/00; la anotación indebida de unidades de medida o de sus símbolos, por ejemplo: toneladas por kilos, millones de pesos por pesos, etc., y otras más. Los errores de este tipo en muchos casos se descubren fácilmente y aún pueden corregirse, pero en otros muchos no se identifican o no se pueden corregir.

Después de haber leído lo anterior, se comprenderá que para formar estadísticas, manejar las que otras personas obtienen y aprovecharlas convenientemente, se requiere cierta especialización asimilable a cualquier profesión. Sin embargo, personas de toda edad, instrucción y calidad mental reciben constantemente cifras estadísticas en periódicos, revistas, folletos, obras, discursos, conferencias, etc., y están en libertad de usarlas e interpretarlas como les place. El peligro de un uso

inconveniente de las estadísticas es mayor entre personas con cultura general, técnica o profesional no estadística, porque les parece lo más natural opinar sobre las cifras que se les presentan, y obtener conclusiones de ellas; estas personas no expresarán su juicio acerca de las ventajas que ofrece determinado tipo de presa, o a propósito de los efectos terapéuticos que producen los rayos ultravioletados, o sobre la teoría cinética de los gases, pero sí darán sus ideas sobre la descripción estadística del analfabetismo, o de los rendimientos de las cosechas, o de la elevación de los precios. La Estadística, por su propia naturaleza, por los campos en que se aplica y por la forma aparentemente accesible de sus resultados, está más propensa a usarse indebidamente, con perjuicio de su propio prestigio.

No basta la preparación técnica ni el trabajo cuidadoso para hacer un buen uso de las estadísticas, sino que es preciso, además, tener temperamento, actitud mental y posición moral adecuadas. La tendencia de conservar prejuicios rechazando sin examen cualquier dato o conclusión contrario a ellos, es uno de los enemigos de la Estadística. El estudio de las cuestiones más elementales debe hacerse con espíritu científico, cuya característica mejor quizá sea su amplitud de miras. La especialización técnica y profesional, cuando no se equilibra convenientemente con otras actividades o conocimientos, estrecha el criterio de las personas que le entregan su vida y las convierte en seres incomprensivos de todo lo que no es su campo de acción. Se supone que quienes usan estadísticas conocen bien el asunto a que ellas se refieren, y están convencidas de que las cifras de que disponen les ayudarán a conocer mejor dicho asunto y a formular conclusiones racionales. El empleo de métodos estadísticos en determinada materia, no excluye que ésta se trate o investigue también mediante otros métodos. Las cuestiones que se observan estadísticamente, casi siempre están ligadas con otras cuestiones numerosas y complejas y muchas veces se influyen recíprocamente; quien estudia tales cuestiones debe tener habilidad especial para descubrir y tener en cuenta casi todas las fuerzas importantes que intervienen para producir un hecho o determinada situación. El buen sentido, la lógica y la rectitud moral, son básicos en el empleo correcto de los métodos estadísticos, como lo son en cualquiera profesión y debieran serlo en toda actividad.

Millares de personas han dedicado sus esfuerzos al progreso de la Estadística y de su empleo. Son aplicaciones vastas de la Estadística la Demografía, la Biometría, la Psicometría, la Econometría, y otras ciencias similares. Las ciencias médicas, la Geografía Económica, la Hidrología, la Meteorología y otras disciplinas semejantes, pueden usar provechosamente los métodos estadísticos. En la Ingeniería pura, como en la Astronomía o la Geodesia, la Estadística tiene reducida aplicación, pero en las actividades que realizan los ingenieros sí hay campo para ella. En los negocios es posible y ventajoso el empleo de la Estadística para cuestiones técnicas, pero principalmente para las administrativas y las de operaciones comerciales. El hombre tenazmente lucha por medir o cuantificar lo que concierne a él mismo y a lo que le rodea de cerca y aun a distancias enormemente grandes, lucha por descubrir las leyes que gobiernan el universo y lucha también por reducir a pocas cifras, y si es posible a simples índices o a ecuaciones, cúmulos considerables de observaciones. En el afán de verlo todo y medirlo todo, el hombre ha construido y perfeccionado admirablemente el microscopio y el ultramicroscopio, ha ensayado métodos de medición indirecta para las magnitudes sumamente pequeñas o exageradamente grandes, y ha introducido nuevas unidades para tales magnitudes. En el afán de precisar las relaciones de unos hechos con otros, formular las leyes a las que el hombre debe hacer frente y dar en breves expresiones numéricas el resultado de múltiples observaciones realizadas sobre campos vastísimos y durante grandes períodos de tiempo, el hombre ha desarrollado los métodos estadísticos, a los que espera un futuro brillante, con las limitaciones únicas que imponen la Lógica y la Economía de los servicios estadísticos.

El profesor F. C. Mills expone del modo siguiente el carácter inductivo, el alcance y las limitaciones lógicas de los métodos estadísticos:

“El proceso lógico mediante el cual se obtienen conclusiones generales por medio del estudio de casos particulares se denomina “inducción”, en oposición a “deducción”, que significa la obtención de conclusiones derivadas de ciertos principios generales. Por “inducción estadística” o “inferencia estadística” se entiende la generalización de los resultados estadísticos, en el supuesto de que una cierta medida pueda aplicarse a un grupo más extenso (o conjunto) que aquél del cual se derivó. Se emplea este procedimiento constante-

mente en los trabajos de estadística práctica, aunque no siempre se satisfagan plenamente las hipótesis y las limitaciones inherentes a dicho proceso. Expongamos sucintamente una descripción de estas hipótesis y limitaciones”.

“... la naturaleza no se caracteriza por un completo caos; en todos los procesos naturales aparecen principios de regularidad, orden y estabilidad, los cuales se acusan más fuertemente al operar con masas de datos cuantitativos. Por tanto, cuando generalizamos una medida cualquiera, tal como un índice de precios al por mayor, venimos a hacer, sobre poco más o menos, la siguiente hipótesis: es razonable suponer que, en el más amplio conjunto al cual ha de exponerse este resultado, exista uniformidad con respecto a la característica o relación medida. Como resultado de esta uniformidad puede esperarse que las medidas estadísticas obtenidas de grupos sucesivos procedentes de este conjunto fluctúen dentro de límites definidos, determinables de antemano”.

“Es evidente que al hacer esta hipótesis, esto es, al decir “es razonable suponer...”, introducimos un supuesto, incapaz de completa comprobación mediante el uso de métodos puramente estadísticos. Hay, pues, en toda inducción estadística un elemento *a priori*. Ninguna conclusión estadística puede jamás descansar por completo en su propia base, sino que debe estar respaldada por el juicio y el raciocinio, si es que ha de ser convincente. Si se obtuviera un elevado coeficiente positivo de correlación al hacer un estudio de las importaciones de plátanos y el coeficiente de suicidio, nadie se mostraría por esto convencido de la existencia de una relación causal o de contingencia entre ambas variables; no habría aquí una base razonable que permitiera suponer que en el universo de los fenómenos de que procede el grupo estudiado, existiese uniformidad con respecto a la correlación hallada”.

* * *

Las personas desconocedoras de los métodos estadísticos que se hayan tomado la molestia de leer las páginas anteriores, no por esto habrán adquirido especialidad en estadística; quienes desean lograrlo en grado apreciable, deberán asimilar amplios y firmes conocimientos de la técnica estadística, aplicarlos repetida e inteligentemente, estudiar con provecho la obra de las personas de diversos países que han hecho aportaciones valiosas a la Metodología Estadística, fomentar y encauzar su interés por la búsqueda de la verdad, en fin, crear hábitos de trabajo más que sabias doctrinas con

todo lo que se deriva de este artículo, cuyo contenido deja flotando muchas ideas para desarrollos más especializados.

* * *

En los artículos que siguen trataré esquemáticamente la técnica de los valores medios, la variabilidad y la correlación. Los lectores que deseen conocer obras y artículos de esta materia, tendrán buenas referencias consultando la "Bibliografía del Método Estadístico y sus Aplicaciones", que cita 1 811 títulos correspondientes a unos 1 000 autores que han escrito en español, inglés, italiano, alemán y francés, posteriormente a 1920. Esta Bibliografía se formó en la Dirección General de Estadística, S. E. N., y se editó por el D. A. P. P. el año de 1938. Para facilitar la tarea de buscar obras útiles y más o menos adaptadas a las necesidades más comunes en los trabajos de la Comisión Nacional de Irrigación, he seleccionado las referencias que siguen:

An engineer's manual of statistical methods. Por Leslie E. Simon. Editado en 1941 por John Wiley

& Sons., Inc. Nueva York. 231 páginas. Dólares 2.75.

Statistical determination of costs, with special reference to marginal costs. Por J. P. Dean. Editado el año de 1936 por la University of Chicago Press. Chicago.

Hand book of statistical monographs, tables and formulas. Por Dunlop J. W. y Kurtz A. K. Editado en 1932 por World Book Co. Nueva York.

Métodos estadísticos aplicados a la economía y a los negocios. Por Cecil Mills. Editado en 1940 por M. Aguilar. Madrid. 600 páginas. \$ 25.00, aproximadamente.

Principios de metodología estadística. Por Irving G. Gavett. Editado en 1941 por la Escuela Nacional de Economía, U. N. A. México, D. F. 327 páginas, \$ 6.00.

Graphic method for presenting business statistics. Por John Randolph Ringleman. Editado en 1936 por McGraw-Hill. Nueva York.

