

Manual de Estadística

David Ruiz Muñoz

Editado por eumed.net

2004

ISBN: 84-688-6153-7

ÍNDICE

Capítulo I:	Historia de la Estadística
Capítulo II:	Características de una distribución de frecuencias
Capítulo III:	Distribuciones bidimensionales
Capítulo IV:	Números índices
Capítulo V:	Series temporales
Capítulo VI:	Variables aleatorias
Capítulo VII:	Probabilidad

David Ruiz Muñoz
Profesor Departamento Economía y Empresa
Universidad Pablo de Olavide

**Capítulo
I****HISTORIA DE LA ESTADISTICA**

Como dijera Huntsberger: "La palabra estadística a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y así sucesivamente. Huntsberger tiene razón pues al instante de escuchar esta palabra estas son las imágenes que llegan a nuestra cabeza.

La Estadística es mucho más que sólo números apilados y gráficas bonitas. Es una ciencia con tanta antigüedad como la escritura, y es por sí misma auxiliar de todas las demás ciencias. Los mercados, la medicina, la ingeniería, los gobiernos, etc. Se nombran entre los más destacados clientes de ésta.

La ausencia de ésta conllevaría a un caos generalizado, dejando a los administradores y ejecutivos sin información vital a la hora de tomar decisiones en tiempos de incertidumbre.

La Estadística que conocemos hoy en día debe gran parte de su realización a los trabajos matemáticos de aquellos hombres que desarrollaron la teoría de las probabilidades, con la cual se adhirió a la Estadística a las ciencias formales.

En este breve material se expone los conceptos, la historia, la división así como algunos errores básicos cometidos al momento de analizar datos Estadísticos.

Definición de Estadística

La Estadística es la ciencia cuyo objetivo es reunir una información cuantitativa concerniente a individuos, grupos, series de hechos, etc. y deducir de ello gracias al análisis de estos datos unos significados precisos o unas previsiones para el futuro.

La estadística, en general, es la ciencia que trata de la recopilación, organización presentación, análisis e interpretación de datos numéricos con e fin de realizar una toma de decisión más efectiva.

Otros autores tienen definiciones de la Estadística semejantes a las anteriores, y algunos otros no tan semejantes. Para Chacón esta se define como "la ciencia que tiene por objeto el estudio cuantitativo de los colectivos"; otros la definen como la expresión cuantitativa del conocimiento dispuesta en forma adecuada para el escrutinio y análisis. La más aceptada, sin embargo, es la de Minguez, que define la Estadística como "La ciencia que tiene por objeto aplicar las leyes de la cantidad a los hechos sociales para medir su intensidad, deducir las leyes que los rigen y hacer su predicción próxima".

Los estudiantes confunden comúnmente los demás términos asociados con las Estadísticas, una confusión que es conveniente aclarar debido a que esta palabra tiene tres significados: la palabra estadística, en primer término se usa para referirse a la información estadística; también se utiliza para referirse al conjunto de técnicas y métodos que se utilizan para analizar la información estadística; y el término estadístico, en singular y en masculino, se refiere a una medida derivada de una muestra.

Utilidad e Importancia

Los métodos estadísticos tradicionalmente se utilizan para propósitos descriptivos, para organizar y resumir datos numéricos. La estadística descriptiva, por ejemplo trata de la tabulación de datos, su presentación en forma gráfica o ilustrativa y el cálculo de medidas descriptivas.

Ahora bien, las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad y en otras actividades; estudios de consumidores; análisis de resultados en deportes; administradores de instituciones; en la educación; organismos políticos; médicos; y por otras personas que intervienen en la toma de decisiones.

Historia de la Estadística

Los comienzos de la estadística pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo, prolijos datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, dicho registro de riqueza y población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

En el antiguo Israel la Biblia da referencias, en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David por otra parte, ordenó a Joab, general del ejército hacer un censo de Israel con la finalidad de conocer el número de la población.

También los chinos efectuaron censos hace más de cuarenta siglos. Los griegos efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años realizaban un censo de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. Para el nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del imperio.

Durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones Estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve en el 758 y por Carlomagno en el 762 DC. Durante el siglo IX se realizaron en Francia algunos censos parciales de siervos. En Inglaterra, Guillermo el Conquistador recopiló el Domesday Book o libro del Gran Catastro para el año 1086, un documento de la propiedad, extensión y valor de las tierras de Inglaterra. Esa obra fue el primer compendio estadístico de Inglaterra.

Aunque Carlomagno, en Francia; y Guillermo el Conquistador, en Inglaterra, trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media.

Durante los siglos XV, XVI, y XVII, hombres como Leonardo de Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Sir Francis Bacon y René Descartes, hicieron grandes operaciones al método científico, de tal forma que cuando se crearon los Estados Nacionales y surgió como fuerza el comercio internacional existía ya un método capaz de aplicarse a los datos económicos.

Para el año 1532 empezaron a registrarse en Inglaterra las defunciones debido al temor que Enrique VII tenía por la peste. Más o menos por la misma época, en Francia la ley exigió a los clérigos registrar los bautismos, fallecimientos y matrimonios. Durante un brote de peste que apareció a fines de la década de 1500, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos. Esa costumbre continuó muchos años, y en 1632 estos Bills of Mortality (Cuentas de Mortalidad) contenían los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar. El trabajo de Graunt, condensado en su obra Natural and Political Observations...Made upon the Bills of Mortality (Observaciones Políticas y Naturales ... Hechas a partir de las Cuentas de Mortalidad), fue un esfuerzo innovador en el análisis estadístico.

Por el año 1540 el alemán Sebastián Muster realizó una compilación estadística de los recursos nacionales, comprensiva de datos sobre organización política, instrucciones sociales, comercio y poderío militar. Durante el siglo XVII aportó indicaciones más

concretas de métodos de observación y análisis cuantitativo y amplió los campos de la inferencia y la teoría Estadística.

Los eruditos del siglo XVII demostraron especial interés por la Estadística Demográfica como resultado de la especulación sobre si la población aumentaba, decrecía o permanecía estática.

En los tiempos modernos tales métodos fueron resucitados por algunos reyes que necesitaban conocer las riquezas monetarias y el potencial humano de sus respectivos países. El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en siete moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después de revisar miles de partidas de defunción pudo demostrar que en tales años no fallecían más personas que en los demás. Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros.

Durante el siglo XVII y principios del XVIII, matemáticos como Bernoulli, Francis Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante durante cierto tiempo, la teoría de las probabilidades limitó su aplicación a los juegos de azar y hasta el siglo XVIII no comenzó a aplicarse a los grandes problemas científicos.

Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia serían el aliado más eficaz del gobernante consciente. La raíz remota de la palabra se halla, por otra parte, en el término latino status, que significa estado o situación; Esta etimología aumenta el valor intrínseco de la palabra, por cuanto la estadística revela el sentido cuantitativo de las más variadas situaciones.

Jacques Quételet es quien aplica las Estadísticas a las ciencias sociales. Este interpretó la teoría de la probabilidad para su uso en las ciencias sociales y resolver la aplicación del principio de promedios y de la variabilidad a los fenómenos sociales. Quételet fue el primero en realizar la aplicación práctica de todo el método Estadístico, entonces conocido, a las diversas ramas de la ciencia.

Entretanto, en el período del 1800 al 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría Estadística; la teoría de los errores de observación, aportada por Laplace y Gauss; y la teoría de los mínimos cuadrados desarrollada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Gaston ideó el método conocido por Correlación, que tenía por objeto medir la influencia relativa de los factores sobre las variables. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones.

Los progresos más recientes en el campo de la Estadística se refieren al ulterior desarrollo del cálculo de probabilidades, particularmente en la rama denominada indeterminismo o relatividad, se ha demostrado que el determinismo fue reconocido en la Física como resultado de las investigaciones atómicas y que este principio se juzga aplicable tanto a las ciencias sociales como a las físicas.

Etapas de Desarrollo de la Estadística

La historia de la estadística está resumida en tres grandes etapas o fases.

1.- Primera Fase: Los Censos:

Desde el momento en que se constituye una autoridad política, la idea de inventariar de una forma más o menos regular la población y las riquezas existentes en el territorio está ligada a la conciencia de soberanía y a los primeros esfuerzos administrativos.

2.- Segunda Fase: De la Descripción de los Conjuntos a la Aritmética Política:

Las ideas mercantilistas extrañan una intensificación de este tipo de investigación. Colbert multiplica las encuestas sobre artículos manufacturados, el comercio y la población: los intendentes del Reino envían a París sus memorias. Vauban, más conocido por sus fortificaciones o su Dime Royale, que es la primera propuesta de un impuesto sobre los ingresos, se señala como el verdadero precursor de los sondeos. Más tarde, Bufón se preocupa de esos problemas antes de dedicarse a la historia natural.

La escuela inglesa proporciona un nuevo progreso al superar la fase puramente descriptiva. Sus tres principales representantes son Graunt, Petty y Halley. El penúltimo es autor de la famosa Aritmética Política.

Chaptal, ministro del interior francés, publica en 1801 el primer censo general de población, desarrolla los estudios industriales, de las producciones y los cambios, haciéndose sistemáticos durante las dos terceras partes del siglo XIX.

3.- Tercera Fase: Estadística y Cálculo de Probabilidades:

El cálculo de probabilidades se incorpora rápidamente como un instrumento de análisis extremadamente poderoso para el estudio de los fenómenos económicos y sociales y en general para el estudio de fenómenos "cuyas causas son demasiado complejas para conocerlos totalmente y hacer posible su análisis".

División de la Estadística

La Estadística para su mejor estudio se ha dividido en dos grandes ramas: la Estadística Descriptiva y la Inferencial.

Estadística Descriptiva: consiste sobre todo en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describir los mismos sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos, como tales.

Estadística Inferencial: se deriva de muestras, de observaciones hechas sólo acerca de una parte de un conjunto numeroso de elementos y esto implica que su análisis requiere de generalizaciones que van más allá de los datos. Como consecuencia, la característica más importante del reciente crecimiento de la estadística ha sido un cambio en el énfasis de los métodos que describen a métodos que sirven para hacer generalizaciones. La Estadística Inferencial investiga o analiza una población partiendo de una muestra tomada.

Método Estadístico

El conjunto de los métodos que se utilizan para medir las características de la información, para resumir los valores individuales, y para analizar los datos a fin de extraerles el máximo de información, es lo que se llama métodos estadísticos. Los métodos de análisis para la información cuantitativa se pueden dividir en los siguientes seis pasos:

1. Definición del problema.
2. Recopilación de la información existente.
3. Obtención de información original.
4. Clasificación.
5. Presentación.
6. Análisis.

Errores Estadísticos Comunes

Al momento de recopilar los datos que serán procesados se es susceptible de cometer errores así como durante los cálculos de los mismos. No obstante, hay otros errores que no tienen nada que ver con la digitación y que no son tan fácilmente identificables. Algunos de éstos errores son:

Sesgo: Es imposible ser completamente objetivo o no tener ideas preconcebidas antes de comenzar a estudiar un problema, y existen muchas maneras en que una perspectiva o

estado mental pueda influir en la recopilación y en el análisis de la información. En estos casos se dice que hay un sesgo cuando el individuo da mayor peso a los datos que apoyan su opinión que a aquellos que la contradicen. Un caso extremo de sesgo sería la situación donde primero se toma una decisión y después se utiliza el análisis estadístico para justificar la decisión ya tomada.

Datos no comparables: el establecer comparaciones es una de las partes más importantes del análisis estadístico, pero es extremadamente importante que tales comparaciones se hagan entre datos que sean comparables.

Proyección descuidada de tendencias: la proyección simplista de tendencias pasadas hacia el futuro es uno de los errores que más ha desacreditado el uso del análisis estadístico.

Muestreo Incorrecto: en la mayoría de los estudios sucede que el volumen de información disponible es tan inmenso que se hace necesario estudiar muestras, para derivar conclusiones acerca de la población a que pertenece la muestra. Si la muestra se selecciona correctamente, tendrá básicamente las mismas propiedades que la población de la cual fue extraída; pero si el muestreo se realiza incorrectamente, entonces puede suceder que los resultados no signifiquen nada

Capítulo II	CARACTERÍSTICAS DE UNA DISTRIBUCIÓN DE FRECUENCIAS
------------------------	---

2.1. Introducción

La fase previa de cualquier estudio estadístico se basa en la recogida y ordenación de datos; esto se realiza con la ayuda de los resúmenes numéricos y gráficos visto en los temas anteriores.

2.2. Medidas de posición

Son aquellas medidas que nos ayudan a saber *donde* están los datos pero sin indicar como se distribuyen.

2.2.1. Medidas de posición centrala) Media aritmética (\bar{X})

La media aritmética o simplemente media, que denotaremos por \bar{X} , es el número obtenido al dividir la suma de todos los valores de la variable entre el número total de observaciones, y se define por la siguiente expresión:

$$\bar{x} = \frac{\sum_{i=1}^n x_i n_i}{N}$$

Ejemplo:

Si tenemos la siguiente distribución, se pide hallar la media aritmética, de los siguientes datos expresados en kg.

xi	ni	xi ni
54	2	108
59	3	177
63	4	252
64	1	64
	N=10	601

$$\bar{X} = \frac{\sum_{i=1}^n x_i n_i}{N} = \frac{601}{10} = \underline{\underline{60,1}} \text{ kg}$$

Si los datos están agrupados en intervalos, la expresión de la media aritmética, es la misma, pero utilizando la marca de clase (X_i).

Ejemplo:

$(L_{i-1}, L_i]$	x_i	n_i	$x_i n_i$
			n_i

[30 , 40]	35	3	105
(40 , 50]	45	2	90
(50 , 60]	55	5	275
		10	470

$$\bar{X} = \frac{\sum_{i=1}^n x_i n_i}{N} = \frac{470}{10} = \underline{\underline{47}}$$

Propiedades:

1ª) Si sometemos a una variable estadística X, a un cambio de origen y escala $Y = a + bX$, la media aritmética de dicha variable X, varía en la misma proporción.

$$Y = a + bX \quad \longrightarrow \quad \bar{Y} = a + b\bar{X}$$

2ª) La suma de las desviaciones de los valores o datos de una variable X, respecto a su media aritmética es cero.

$$\sum_{i=1}^n (x_i - \bar{x}) n_i = 0$$

Ventajas e inconvenientes:

- La media aritmética viene expresada en las mismas unidades que la variable.
- En su cálculo intervienen todos los valores de la distribución.
- Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
- Es única.
- Su principal inconveniente es que se ve afectada por los valores extremadamente grandes o pequeños de la distribución.

- **Media aritmética ponderada**

Es una media aritmética que se emplea en distribuciones de tipo unitario, en las que se introducen unos coeficientes de ponderación, denominados w_i , que son valores positivos, que representan el número de veces que un valor de la variable es más importante que otro.

$$W = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

b) Media geométrica

Sea una distribución de frecuencias (x_i, n_i) . La media geométrica, que denotaremos por G, se define como la raíz N-ésima del producto de los N valores de la distribución.

$$G = \sqrt[N]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}}$$

Si los datos están agrupados en intervalos, la expresión de la media geométrica, es la misma, pero utilizando la marca de clase (Xi).

El empleo más frecuente de la media geométrica es el de promediar variables tales como porcentajes, tasas, números índices. etc., es decir, en los casos en los que se supone que la variable presenta variaciones acumulativas.

Ventajas e inconvenientes:

- En su cálculo intervienen todos los valores de la distribución.
- Los valores extremos tienen menor influencia que en la media aritmética.
- Es única.
- Su cálculo es más complicado que el de la media aritmética.

Además, cuando la variable toma al menos un $x_i = 0$ entonces G se anula, y si la variable toma valores negativos se pueden presentar una gama de casos particulares en los que tampoco queda determinada debido al problema de las raíces de índice par de números negativos.

c) Media armónica

La media armónica, que representaremos por H, se define como sigue:

$$H = \frac{N}{\sum_{i=1}^r \frac{1}{x_i} n_i}$$

Obsérvese que la inversa de la media armónica es la media aritmética de los inversos de los valores de la variable. No es aconsejable en distribuciones de variables con valores pequeños. Se suele utilizar para promediar variables tales como productividades, velocidades, tiempos, rendimientos, cambios, etc.

Ventajas e inconvenientes:

- En su cálculo intervienen todos los valores de la distribución.
- Su cálculo no tiene sentido cuando algún valor de la variable toma valor cero.
- Es única.

- Relación entre las medias:

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

