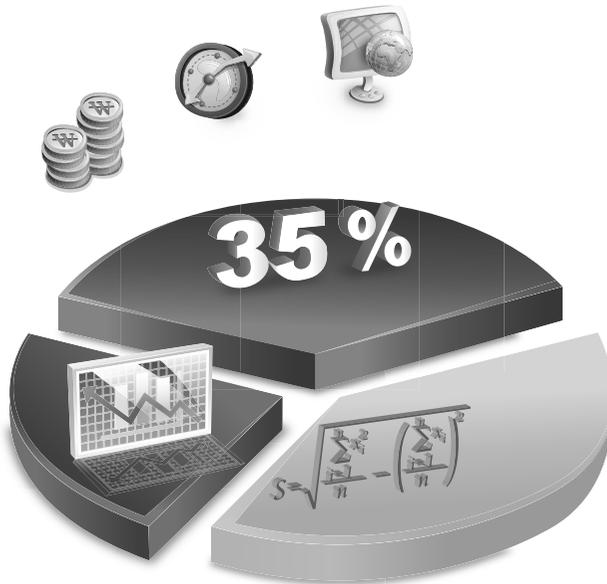




ESTADÍSTICA

Descriptiva y Probabilidades
para Ingenieros



Estadística Descriptiva y Probabilidades para Ingenieros

Autor: Celestino García Oré

© Derecho de autor reservado
Empresa Editora Macro E.I.R.L.

© Derecho de edición, arte gráfico y diagramación reservados
Empresa Editora Macro E.I.R.L.

Edición a cargo de:
Empresa Editora Macro E.I.R.L.
Av. Paseo de la República 5613 – Miraflores
Lima - Perú
(511) 719-9700

☎ ventas@editorialmacro.com

✉ <http://www.editorialmacro.com>

Primera edición: Octubre 2011 - 1000 ejemplares

Impreso en los Talleres Gráficos de
Empresa Editora Macro E.I.R.L.
Lima - Perú

ISBN Nº 978-612-304-027-7

Hecho el Depósito Legal en la Biblioteca Nacional del Perú Nº 2011-12578

Prohibida la reproducción parcial o total, por cualquier medio o método de este libro sin
previa autorización de la Empresa Editora Macro E.I.R.L.



CELESTINO GARCÍA ORÉ

Celestino García Oré nació en el distrito de Concepción, provincia de Vilcashuamán, departamento de Ayacucho-Perú.

Realizó sus estudios profesionales en La Cantuta (1954-1958). El año 1966 fue becado por la OEA para asistir al Instituto de Verano de Matemática en la Universidad de Puerto Rico. Con una beca otorgada por la Fundación Nacional de Ciencias, siguió estudios en el Instituto del Año Académico de Matemática y en la Escuela de Postgrado de Matemática de la Universidad de Puerto Rico. El año 1976, nuevamente becado por la OEA hizo estudios de postgrado en Estadística Matemática en la Universidad de Chile (CIENES), Santiago de Chile. El año 1980 asistió al Programa de Investigación Estadística (PIA) en la Universidad de Chile (CIENES), Santiago de Chile.

Se inició en la docencia universitaria en la Universidad Nacional de San Cristóbal de Huamanga (1963), y también perteneció al cuerpo docente de la Universidad de Puerto Rico (1969-1972). Después de 5 años de permanencia en el extranjero retornó al Perú el año 1972 y se desempeñó como docente en las siguientes universidades: Universidad Nacional de Ingeniería (1972-1989), Universidad Peruana Cayetano Heredia (1977-1981), Universidad de Lima (1982-1986), Universidad Ricardo Palma (1989-2001), Escuela de Postgrado de la Universidad de San Cristóbal de Huamanga (2000-2001), Escuela de Postgrado de la Universidad Nacional de Educación, La Cantuta (2002-2004).

Entre los años 1985-1987 fue presidente de la Asociación de Docentes de la Universidad Nacional de Ingeniería.

Otras obras del autor: *Ajuste de curvas de Pearson, Distribuciones y estadística inferencial, Distribuciones multivariantes. Regresión y correlación, Métodos Estadísticos en la evaluación educacional, Evaluación en el contexto de la educación actual.*

Dedicatoria

A mi esposa:

Blanca Almonacid Cisneros

A mis hijos:

Carlos Pavel

Nina Karina

Iván Alex

Omar

A mis nietos:

Jaime Uriel

Eduardo Mijael

Adham Pavel

Diego Franco

Said Aarón

Sebastián Iván

Carmina

Samira

Ana Paula

Omara Paola

Caissa Marisol

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología "CONCYTEC" por su apoyo integral a la primera edición (1989).

A la Empresa Gráfica PRINCELINNESS E.I.R.L. (1994).

A la Editorial MACRO por su apoyo decisivo a esta tercera edición revisada y aumentada.

Prólogo

El propósito del libro es presentar una introducción moderna de la Estadística en base a una larga experiencia docente en las universidades del país y del extranjero. Está dirigido a estudiantes de ingeniería, economía, estadística y especialidades afines.

Considerando la Estadística como una Metodología Científica de Trabajo antes que una Teoría (Teoría Estadística) se desarrolla en forma práctica y en base a ejemplos. Los problemas resueltos y propuestos fueron seleccionados de los Exámenes y Prácticas tomados en la Universidad Nacional de Ingeniería, Universidad Peruana Cayetano Heredia, Universidad de Lima y la Universidad Ricardo Palma.

La acogida favorable que tuvo la Primera y Segunda Edición hicieron posible que esta Edición revisada y aumentada, a la luz de los tiempos de la Informática y la Computación, supere ampliamente a la anterior, aunque conserva su estructura y objetivos trazados. Acogiendo la recomendación de los colegas que enseñan la asignatura de Estadística, tanto los problemas resueltos como los propuestos, se presentan en orden creciente de su dificultad.

Lima, Junio del 2011

El Autor

Índice

Capítulo 1

INTRODUCCIÓN	9
1.1 Breve reseña histórica	9
1.2 ¿La estadística es una ciencia o una metodología?	9
1.3 Población y muestra. Unidad estadística, dato u observación	10
1.4 Parámetro, estadígrafo y estadístico	10
1.5 Algunas definiciones de estadística	11
1.6 Estadística descriptiva y estadístico inferencial	11
1.7 Ciclo metodológico del quehacer estadístico	11
1.8 Variables	12
1.9 Vector observación o vector de medidas	13
1.10 Datos univariantes, bivariantes y multivariantes	13

Capítulo 2

ESTADÍSTICA DESCRIPTIVA	15
2.1 Recopilación	15
2.2 Clasificación	20
2.3 Presentación de datos	23
2.4 Descripción de datos	36
2.4.1 Media Aritmética o Promedio Aritmético	36
2.4.2 Mediana	44
2.4.3 Moda	48
2.4.4 Media Geométrica	56
2.4.5 Media Armónica	58
2.4.6 Cuantilas	60
2.4.7 Varianza	68
2.4.8 Desviación Típica	72
2.4.9 Coeficiente de Variación	72
2.4.10 Desviación Media	75
2.4.11 Rango	76
2.4.12 Rangos Intercuantílicos	76
2.4.13 Desviación Cuartílica	77
2.4.14 Medidas de Asimetría y Kurtosis	77
2.4.15 Momentos	80
2.4.16 Algunas relaciones más relevantes entre las Medidas Descriptivas	84
PROBLEMAS RESUELTOS	86
PROBLEMAS PROPUESTOS	129

Capítulo 3

DISTRIBUCIONES BIDIMENSIONALES	141
3.1 Distribuciones bidimensionales de datos no clasificados	142
3.1.1 Correlación Lineal, Covarianza	142
3.1.2 Matriz de Covarianzas y Matriz de Correlaciones	143
3.1.3 Correlación por Rangos	146
3.1.4 Regresión Lineal	149
3.2 Distribuciones bidimensionales de datos clasificados	154
3.2.1 Distribuciones Marginales	155
3.2.2 Medias y Varianzas, Covarianzas, Coeficiente de Correlación y Coeficiente de Regresión	157
3.3 Distribuciones de frecuencias absolutas y relativas acumuladas	168
3.4 Observaciones cualitativas y observaciones mixtas multivariadas	170
PROBLEMAS RESUELTOS	174
PROBLEMAS PROPUESTOS	188

Capítulo 4

PROBABILIDADES	197
4.1 Experimento aleatorio o fenómeno aleatorio	197
4.2 Espacio muestral o espacio de muestra	198
4.3 Evento o suceso	200
4.4 Función de probabilidad	204
4.5 Análisis combinatorio	209
4.5.1 Principios de Conteo	209
4.5.2 Permutaciones	210
4.5.3 Combinaciones	214
4.6 Probabilidad condicional	216
4.7 Independencia estocástica	221
4.7.1 Independencia Conjunta	223
4.8 Partición del espacio muestral	226
PROBLEMAS RESUELTOS	233
PROBLEMAS PROPUESTOS	262

1.1 BREVE RESEÑA HISTÓRICA

El desarrollo histórico de la estadística atraviesa por tres etapas:

- a) **Etapla inicial.** Se extiende desde la antigüedad hasta mediados del siglo XVIII. Se caracteriza porque la estadística está asociada a los censos poblacionales y registro de bienes y servicios de un estado. Estadística deriva del vocablo 'estado'. La cultura egipcia contó con recopilaciones regulares de datos de su administración estatal, incluso divinizó a Saphkit, diosa de los libros y de las cuentas. Los romanos llevaron registros numéricos con fines tributarios. En el Imperio Incaico llevaron registros de su población y otras cuentas mediante los **QUIPUS**. Los quipus son **numerales** de un sistema de Numeración Decimal Inca (base 10) que consideraban el valor de posición y conocieron el cero (un hilo de cualquier color sin nudos, representa el cero). Así, el Sistema de Numeración Inca es equivalente al Sistema de Numeración Decimal actual Hindo – Árábigo y ampliamente superior a los Sistemas Maya y Romana. Además, los quipus son equivalentes a los discos duros actuales, en los que se almacenaban la información requerida por el vasto imperio incaico.
- b) **Etapla de la sistematización.** Se caracteriza por la aparición de escuelas, destacándose tres:
- La escuela alemana. Creó la primera cátedra de estadística considerando esta disciplina como la descripción de los fenómenos concernientes al estado o administración.
 - La escuela inglesa. Cuantificaron las leyes que rigen los fenómenos sociales y como tal 'aritmetizaron la estadística'.
 - La escuela francesa. Introduce la teoría de las probabilidades como fundamento matemático-teórico de la estadística.
- c) **Etapla actual.** Comprendido entre principios del siglo XIX hasta nuestros días. En esta etapa **la matemática se plasma como la columna vertebral de la estadística** y se caracteriza por el gran desarrollo alcanzado como ciencia y como una metodología de investigación científica aplicada a todas las ramas del saber humano: ingeniería, economía, biología, agronomía, industria, comercio, educación, etc.

1.2 ¿LA ESTADÍSTICA ES CIENCIA O UNA METODOLOGÍA?

La Estadística es ciencia, tan ciencia como la física y la biología, porque su fundamentación teórica la encontramos en una de las ramas más 'puras' de la matemática: La teoría de las probabilidades.

Es también una metodología de trabajo científico que se justifica y resalta en el uso obligatorio de los métodos estadísticos en todo trabajo de investigación.

Estaremos interesados en presentar una estadística como una metodología de trabajo científico, con un contenido más práctico y aplicativo que teórico, en razón a que el presente texto está dirigido a estudiantes que siguen la especialidad de ingeniería, ciencias, educación y materias afines.

1.3 POBLACIÓN Y MUESTRA. UNIDAD ESTADÍSTICA, DATO U OBSERVACIÓN //

- a) **POBLACIÓN (O UNIVERSO)** es el conjunto mayor de 'objetos',¹ que tiene al menos una característica común, cuyo estudio nos interesa o acerca de los cuales se desea información. La población puede ser finita o infinita.
- b) **MUESTRA.** Si la población es infinita será imposible tener una información completa sobre ella; si la población es finita pero numerosa (las poblaciones de interés son generalmente numerosas), igualmente no será posible obtener una información completa sobre ella, entonces se acude a una técnica estadística llamada **Técnica de Muestreo** que nos permitirá seleccionar una parte representativa y finita de la población denominada **MUESTRA**.
- c) **UNIDAD ESTADÍSTICA, DATO U OBSERVACIÓN.** Llamaremos **unidad estadística** a los elementos que conforman la población, y a las unidades estadísticas disponibles las denominaremos **datos u observaciones**; aunque esta última denominación está reservada a una unidad estadística obtenida experimentalmente, ambos términos se toman como sinónimos.

El investigador o el estadístico define, delimita o determina una población. Por ejemplo, podemos decidir que la población esté constituida por los 40 alumnos que conforman un salón de clase o decidir que la población de interés esté constituida por todos los alumnos del colegio o que nuestro universo sea el conjunto de todas las microempresas que existen en el país o podríamos definir poblaciones cada vez más amplias como el conjunto de todos los estudiantes de Latinoamérica.

1.4 PARÁMETRO, ESTADÍGRAFO Y ESTADÍSTICO //

- a) **PARÁMETRO.** Es un número que describe alguna característica de la población. Para determinar su valor es necesario utilizar la información poblacional completa, y por lo tanto, las decisiones sobre las características de la población se toman con certidumbre total.
- b) **ESTADÍGRAFO.** Es un número que describe alguna característica de la muestra y se obtiene a partir de los datos muestrales; la toma de decisiones contiene un grado de incertidumbre.
- c) **ESTADÍSTICO O ESTADÍSTICA.** Es una variable cuyos valores son estadígrafos.

Ejemplo 1

Consideremos la población conformada por los 40 alumnos de la sección A y obtenemos 15 como la edad promedio, este número es un **parámetro** y afirmamos con certidumbre total que la edad promedio de los alumnos de la sección A es 15 o que el parámetro poblacional, llamado promedio, es 15. Si calculo la edad promedio de solamente cinco alumnos, elegidos al azar de la sección A y obtengo 16, este número es un **estadígrafo** y puedo afirmar, con un alto grado de incertidumbre, que la edad promedio de los alumnos de la sección A es 16. Si selecciono otra muestra del mismo tamaño cinco, posiblemente la edad promedio que obtenga sea un número distinto a 16, supongamos que resulte 14, este número es otro estadígrafo; puedo seguir calculando más medias (estadígrafos) a partir de muestras del mismo tamaño cinco, resultando éstas como 17, 15, etc. Estos números: 16, 14, 17, 15, ... constituyen valores de una **variable especial** llamada **estadístico o estadística**, que puede permitirnos estimar el parámetro poblacional (que en este ejemplo es 15) con un grado de incertidumbre mínimo.

1. Conjunto de 'objetos' equivale a un conjunto de personas, animales, cosas, precios de artículos, etc.

1.5 ALGUNAS DEFINICIONES DE ESTADÍSTICA //

Se han planteado muchas definiciones. Veamos algunas:

- "La estadística es una rama del método científico que trata de los datos reunidos al contar o medir las propiedades de alguna población". Kendall Stuart.
- "Estadística es la tecnología del método científico". Mood.
- "La estadística moderna es una teoría de la información con la inferencia como objetivo". Scheaffer y Mendenhall.
- "La estadística es una disciplina que proporciona un conjunto de métodos que permite recopilar, clasificar, presentar y describir datos en forma adecuada para tomar decisiones frente a la incertidumbre o afirmar algo (hacer inferencias) acerca de la población o sus parámetros a partir de los datos extraídos de la misma". C. García Oré.

Tomando la última definición (con fines estrictamente didácticos y no por ser la mejor definición) llegamos a distinguir la estadística descriptiva de la inferencial.

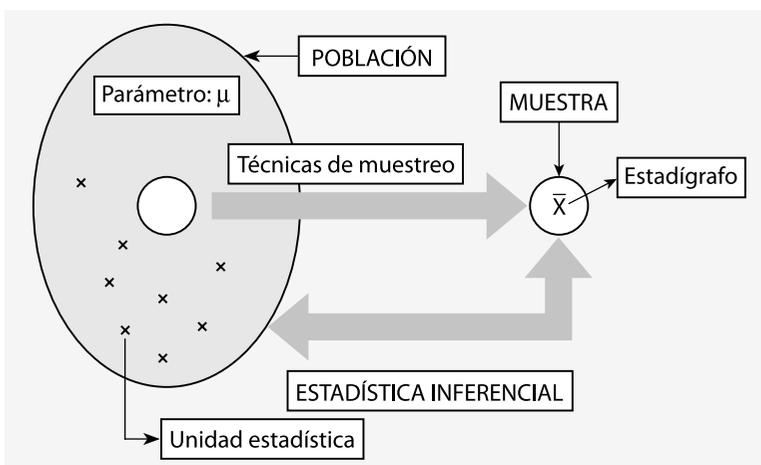
1.6 ESTADÍSTICA DESCRIPTIVA Y ESTADÍSTICA INFERENCIAL //

- ESTADÍSTICA DESCRIPTIVA:** Trata de la recopilación, clasificación, presentación y descripción de los datos.
- ESTADÍSTICA INFERENCIAL:** Proporciona la teoría necesaria para la toma de decisiones frente a la incertidumbre o hacer inferencias acerca de la población o sus parámetros, a partir de los datos muestrales.

1.7 CICLO METODOLÓGICO DEL QUEHACER ESTADÍSTICO //

Cuando no sea posible obtener una información completa de la **población** se extraen **muestras representativas** de dicha población mediante las **técnicas de muestreo**, y sobre la base del estudio o información obtenida de los datos muestrales se afirma algo acerca de la población, se toman decisiones frente a la incertidumbre o se estiman los parámetros poblacionales a partir de las estadísticas, mediante el uso de la **ESTADÍSTICA INFERENCIAL**.

Este ciclo se cumple en la mayoría de las veces del quehacer estadístico:



1.8 VARIABLES

Una variable es la representación literal, simbólica o matemática de alguna característica que posee un dato u observación, cuyo estudio nos interesa.

• VARIABLE CUANTITATIVA

Si la característica bajo estudio es **medible** o **contable**, la variable correspondiente se denomina **VARIABLE CUANTITATIVA**, cuyos valores son números. Las variables cuantitativas pueden ser discretas y continuas.

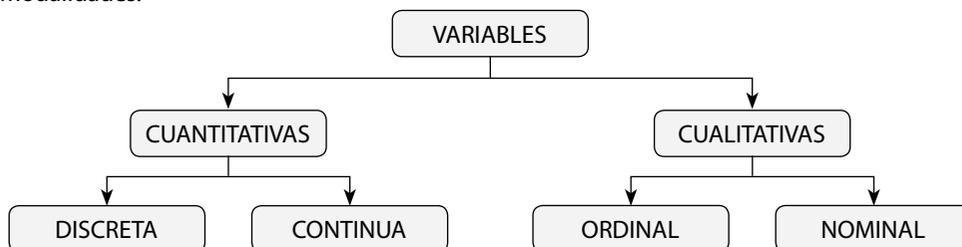
Las **variables discretas** surgen del proceso de conteo y toman valores de un conjunto discreto. Las **variables continuas** surgen de la medición y toman todos los valores reales, al menos teóricamente.

• VARIABLE CUALITATIVA

Si la característica bajo estudio no es medible ni contable se denomina **VARIABLE CUALITATIVA**, cuyos valores no son números. Sus valores son atributos, categorías, cualidades o modalidades. Las variables cualitativas pueden ser ordinales y nominales.

Las variables cualitativas ordinales presentan una jerarquía entre sus atributos, existe una categoría u orden entre sus valores o cualidades.

Las variables cualitativas nominales no presentan orden ni jerarquía entre sus atributos, cualidades o modalidades.



Ejemplo 2

Consideremos un 'alumno' como dato y estamos interesados en estudiar cuatro de sus características: número de hermanos, estatura, sexo, años de estudios.

Representemos estas características mediante las siguientes variables:

X: Número de hermanos.

Y: Estatura.

Z: Sexo.

W: Año de estudios.

La variable X es discreta. Toma valores como 0, 1, 2, ..., n

La variable Y es continua. Toma valores reales del intervalo:

[Estatura mínima, Estatura Máxima.]; toma valores como 1.80, 1.67, 1.70, ...

La variable Z es cualitativa nominal. Tiene dos atributos: Masculino y femenino.

La variable W es cualitativa ordinal. Tiene valores como: Educación Inicial, Primer Grado, Segundo Grado, ..., Sexto Grado, Primer Año, ..., Quinto Año.

1.9 VECTOR OBSERVACIÓN O VECTOR DE MEDIDAS //

Es un vector cuyos componentes son las variables que representan las características de un dato u observación, bajo estudio.

El vector de medidas del ejemplo 2 se expresa así:

$$V = (X, Y, Z, W) \quad \text{o} \quad V = \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}$$

Ejemplo 3

Expresar como un vector de medidas V_0 , las características de un alumno que tiene 3 hermanos, cuenta con 16 años de edad, de sexo masculino y cursa el último año de educación primaria.

$$V_0 = (3, 16, M, 6^{\text{to}}) \quad \text{o} \quad V_0 = \begin{pmatrix} 3 \\ 16 \\ M \\ 6^{\text{to}} \end{pmatrix}$$

Obsérvese que los componentes del vector V_0 son una 'mezcla' de variables cuantitativas y cualitativas.

1.10 DATOS UNIVARIANTES, BIVARIANTES Y MULTIVARIANTES //

DATOS UNIVARIANTES: Si se estudia una sola característica y ésta se representa mediante una sola variable X , estamos frente a un dato univariante y consideramos que este dato proviene de una población también univariante o unidimensional y una muestra de datos es un conjunto de la forma:

$$M = \{x_1, x_2, x_3, \dots, x_n\}$$

Donde las x_i son valores particulares de la única variable X .

Ejemplo 4

Una muestra de 6 calificaciones de un alumno se expresa así:

$$M = \{12, 13, 09, 08, 15, 17\}$$

DATOS BIVARIANTES: Si estuviésemos interesados en estudiar dos características diferentes en cada estudiante, por ejemplo, calificaciones (X) y edad (Y), las representaríamos mediante dos variables:

$$(X, Y) \quad \text{o} \quad \begin{pmatrix} X \\ Y \end{pmatrix}$$

Estamos frente a un dato bivalente y consideramos que este dato proviene de una población bivalente o bidimensional y una muestra es un conjunto de la forma:

$$M = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$$

Donde $x_1, x_2, x_3, \dots, x_n$ son valores de la variable X

$y_1, y_2, y_3, \dots, y_n$ son valores de la variable Y

Ejemplo 5

Si a las calificaciones $X = 12, 13, 09, 08, 15, 17$ asociamos sus edades:

$Y = 13, 12, 12, 13, 14, 12$

La muestra es un conjunto de la forma:

$$M = \{(12, 13), (13, 12), (09, 12), (08, 13), (15, 14), (17, 12)\}$$

O en forma de Tabla:

X	Y
12	13
13	12
09	12
08	13
15	14
17	12

En general, podemos hablar de datos multivariantes provenientes de poblaciones multivariantes o multidimensionales expresables por n variables que representan n características que pueden ser observadas y estudiadas simultáneamente en cada dato.

Ejemplo 6

Consideremos un 'alumno' como dato y estamos interesados en estudiar cuatro de sus características (variables): número de hermanos (X), estatura (Y), sexo (Z) y año de estudios (W). En la siguiente tabla está la información obtenida de una muestra aleatoria de seis alumnos. Esta clase de información, generalizada a un número mayor de alumnos o individuos y más variables o características, adecuadamente codificadas, almacenadas en un disco duro, se denomina una **base de datos o una matriz de datos**.

Vectores de medida	Características o variables			
	X	Y	Z	W
V_1	3	1,53	M	4° Año
V_2	2	1,54	M	5° Año
V_3	0	1,61	F	5° Año
V_4	1	1,52	M	4° Año
V_5	4	1,64	F	3° Año
V_6	3	1,59	M	4° Año

Impreso en los Talleres Gráficos de



Surquillo

☎ 7199700 - 7199701

Octubre 2011