

**ESTADÍSTICA APLICADA
UNA VISIÓN INSTRUMENTAL**

**MARÍA TERESA GONZÁLEZ MANTEIGA
ALBERTO PÉREZ DE VARGAS**

**ESTADÍSTICA APLICADA
UNA VISIÓN INSTRUMENTAL**
**Teoría y más de 500 problemas
resueltos o propuestos con solución**



© María Teresa González Manteiga y Alberto Pérez de Vargas, 2009

Reservados todos los derechos.

«No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.»

Ediciones Díaz de Santos
E-mail: ediciones@diazdesantos.es
Internet://<http://www.diazdesantos.es>

ISBN: 978-84-7978-913-8
Depósito legal: M. 20.883-2009

Diseño de cubierta: Ángel Calvete
Fotocomposición e impresión: Fernández Ciudad
Encuadernación: Rústica - Hilo

Impreso en España

ÍNDICE

POR QUÉ Y PARA QUÉ SE HA ESCRITO ESTE LIBRO	XIX
PRÓLOGO	XXI
1. Capítulo preliminar	
1.1. Introducción	1
1.2. Conceptos generales básicos	12
1.3. Tipos de muestreo aleatorio en poblaciones finitas	14
1.4. Tipos de caracteres	17
1.4.1. Caracteres cualitativos	17
1.4.2. Caracteres cuantitativos	18
1.5. Nociones básicas del Programa STATGRAPHICS Plus 5.1 para Windows	20
1.6. Etapas de un estudio estadístico	26

PRIMERA PARTE ESTADÍSTICA DESCRIPTIVA

2. Estadística descriptiva unidimensional	
2.1. Introducción	31
2.2. Tablas de frecuencias	33
2.3. Representaciones gráficas	37
2.3.1. Caracteres cualitativos	37
2.3.2. Caracteres cuantitativos discretos	40
2.3.3. Caracteres cuantitativos continuos	42
2.4. Análisis numérico	48
2.4.1. Medidas de posición	48

2.4.2. Medidas de dispersión	61
2.4.3. Momentos	66
2.4.4. Características de forma	67
2.5. Transformaciones lineales de una variable estadística	71
2.6. Diagrama de caja y bigotes	74
2.7. Cálculo de los estadísticos con <i>STATGRAPHICS</i>	75
2.8. Números Índices. Breves nociones	85
2.9. Problemas propuestos	92
3. Estadística descriptiva bidimensional	
3.1. Introducción	99
3.2. Tablas de frecuencias de una distribución bidimensional	100
3.3. Distribuciones marginales y condicionadas	103
3.4. Momentos bidimensionales: medias y varianzas marginales, co- varianza	104
3.5. Vector de medias y matriz de covarianza	106
3.6. Regresión y correlación	108
3.7. Regresión lineal mínimo cuadrática. Rectas de regresión. Coe- ficiente de correlación lineal	110
3.7.1. Recta de regresión de y sobre x	110
3.7.2. Recta de regresión de x sobre y	113
3.7.3. Coeficiente de determinación lineal	115
3.7.4. Coeficiente de correlación lineal	119
3.8. Otras regresiones reducibles a lineales	129
3.8.1. Línea de regresión de tipo exponencial $y = ab^x$	130
3.8.2. Línea de regresión de tipo potencial $y = ax^m$	134
3.8.3. Línea de regresión de tipo homográfica o hiperbólica ...	135
3.9. Otras curvas de regresión mínimo-cuadrática. Regresión para- bólica. Coeficiente de determinación parabólico	136
3.9.1. Ajuste a una parábola de grado 2 por el método de míni- mos cuadrados	136
3.9.2. Coeficiente de determinación parabólico	137
3.10. Problemas propuestos	144

SEGUNDA PARTE

PROBABILIDAD. DISTRIBUCIONES DE PROBABILIDAD

4. Probabilidad	
4.1. Introducción	155
4.2. Experimentos aleatorios	157
4.3. Álgebra de sucesos y σ -álgebra	159
4.4. Frecuencia relativa de un suceso. Probabilidad en espacios muestrales finitos	162

4.5.	Axiomática para la probabilidad. Propiedades de la probabilidad ...	164
4.6.	La probabilidad condicionada	170
4.7.	Sucesos dependientes e independientes	171
4.8.	Probabilidad de la intersección de dos o más sucesos	173
4.9.	El teorema de la probabilidad total	174
4.10.	El teorema de Bayes	175
4.11.	Probabilidades en espacios infinito-numerables	179
4.12.	Problemas propuestos	183
5.	Distribuciones de probabilidad discretas	
5.1.	Introducción	191
5.2.	Definición de variable aleatoria discreta	192
5.3.	Función de probabilidad de una variable aleatoria discreta	197
5.4.	Función de probabilidad acumulada de una variable aleatoria discreta	199
5.5.	Características de una variable aleatoria discreta	201
5.5.1.	Esperanza matemática	201
5.5.2.	Momentos para variables aleatorias	203
5.5.3.	Otros parámetros de interés para variables aleatorias discretas	206
5.5.4.	Características de forma para variables aleatorias discretas ..	207
5.6.	Principales distribuciones para variables aleatorias discretas	208
5.6.1.	Distribución discreta uniforme	208
5.6.2.	Distribución de Bernoulli	210
5.6.3.	Distribución binomial	212
5.6.4.	Distribución geométrica	215
5.6.5.	Distribución binomial negativa	218
5.6.6.	Distribución hipergeométrica	220
5.6.7.	Distribución de Poisson	226
5.7.	Aproximación de una distribución binomial por una de Poisson	230
	Tablas de distribuciones discretas	233
5.8.	Problemas propuestos	234
6.	Distribuciones de probabilidad continuas	
6.1.	Definición de variable aleatoria continua	241
6.2.	Función de distribución de una variable aleatoria continua	242
6.3.	Probabilidad asociada a un punto	243
6.4.	Función de densidad de probabilidad de una variable aleatoria continua	244
6.5.	Probabilidad asociada a un intervalo	246
6.6.	Esperanza matemática o valor esperado de una variable aleatoria continua	250
6.7.	Momentos de una variable aleatoria continua	251
6.8.	Otros parámetros de interés para variables aleatorias continuas	253

6.9. Características de forma	254
6.10. Acotación de Chebyshev	255
6.11. Algunas distribuciones continuas	257
6.11.1. Distribución continua uniforme	257
6.11.2. Distribución normal o de Laplace-Gauss	258
Tabla 6.1 de distribuciones continuas	264
6.12. Problemas propuestos	265
7. Distribuciones bidimensionales y multidimensionales	
7.1. Distribuciones bidimensionales y multidimensionales discretas ...	273
7.1.1. Distribuciones marginales	275
7.1.2. Distribuciones condicionadas	277
7.1.3. Independencia de variables aleatorias discretas	278
7.1.4. Distribución de probabilidad para funciones de las variables aleatorias discretas X e Y	280
7.1.5. Esperanza de una función de X . Función generatriz de momentos respecto del origen	282
7.2. Distribución multinomial	285
7.3. Distribuciones bidimensionales y multidimensionales continuas	287
7.3.1. Distribuciones marginales continuas	292
7.3.2. Distribuciones condicionadas continuas	294
7.3.3. Independencia de variables aleatorias continuas	295
7.3.4. Covarianza	296
7.3.5. Varianza de la suma de variables aleatorias	298
7.3.6. Coeficiente de correlación lineal	299
7.3.7. Esperanza de una función de X . Función generatriz de momentos respecto del origen	300
7.4. Función característica de una variable aleatoria X	301
7.5. Función de probabilidad de funciones de X e Y . Cambio de variable...	302
7.6. Distribución normal bidimensional	305
7.7. Distribución normal multidimensional	308
7.8. Problemas propuestos	309

TERCERA PARTE INFERENCIA ESTADÍSTICA

8. Inferencia estadística. Teoremas del límite	
8.1. Introducción	319
8.2. Población de las muestras	320
8.3. La distribución normal y la inferencia estadística	322
8.4. Teoremas del límite	325
8.4.1. El teorema central	325
8.4.2. Teorema de Lindeberg-Lévy	326

8.4.3. Teorema de De Moivre	327
8.4.4. Aproximación de la distribución binomial por la normal ...	327
8.4.5. Aproximación de la distribución de Poisson por la normal	332
8.5. Distribución de la variable aleatoria \bar{X} para muestras de tamaño n en poblaciones no normales	335
8.6. Distribución de una proporción observada \hat{p}	336
8.7. Estimadores y estimaciones	337
8.8. Problemas propuestos	340
9. Estimación de los parámetros de una población	
9.1. Estimación puntual de un parámetro de una población	345
9.1.1. Clases de estimadores	345
9.1.2. Métodos de obtención de estimadores	348
9.1.3. Estimación de la media μ	351
9.1.4. Estimación de una proporción p	353
9.1.5. Estimación de la varianza σ^2	354
9.2. Principales distribuciones teóricas asociadas al proceso de muestreo	355
9.2.1. Función gamma	355
9.2.2. Distribución gamma	356
9.2.3. Distribución exponencial	357
9.2.4. Distribución χ^2	358
9.2.5. Distribución <i>t de Student</i>	361
9.2.6. Distribución <i>F de Snedecor-Fisher</i>	365
9.3. Las funciones de distribución y el programa <i>STATGRAPHICS</i> ..	371
9.4. Distribuciones de los estadísticos en el muestreo	372
9.5. Intervalos de probabilidad	375
9.5.1. Intervalo de probabilidad con riesgo α para \bar{X}	375
9.5.2. Intervalo de probabilidad con riesgo α para \hat{p}	378
9.5.3. Intervalo de probabilidad con riesgo α para s^2	381
9.5.4. Intervalo de probabilidad con riesgo α para la cuasidesviación típica s	382
9.5.5. Intervalo de probabilidad con riesgo α para la diferencia de medias $\bar{X}_1 - \bar{X}_2$	383
9.5.6. Intervalo de probabilidad con riesgo α para el cociente de cuasivarianzas $\frac{s_1^2}{s_2^2}$	386
9.5.7. Intervalo de probabilidad con riesgo α para la diferencia $\hat{p}_1 - \hat{p}_2$ de proporciones muestrales	387
Tabla II. Distribuciones continuas asociadas al proceso de muestreo....	389
9.6. Problemas propuestos	390

10. Intervalos de confianza de los parámetros de una población	
10.1. Introducción	395
10.2. Intervalo de confianza para μ en una población normal con σ^2 conocida	396
10.3. Intervalo de confianza para μ en una población normal con σ^2 desconocida	398
10.4. Intervalo de confianza para la proporción p	400
10.5. Intervalo de confianza para σ^2 en una población normal	401
10.6. Intervalo de confianza para σ en una población normal	403
10.7. Relación entre la precisión y el tamaño de la muestra	403
10.7.1. ¿Cuál debe ser el tamaño de la muestra que se necesita para estimar μ con confianza fijada, $1 - \alpha$, y con precisión $ \bar{X} - \mu \leq d$?	404
10.7.2. ¿Cuál debe ser el tamaño de la muestra que se necesita para estimar p con confianza $1 - \alpha$, y con precisión $ \hat{p} - p \leq d$?	405
10.8. Intervalo de confianza para la diferencia de medias $\mu_1 - \mu_2$	406
10.8.1. Si se conocen las varianzas poblacionales σ_1^2 y σ_2^2	407
10.8.2. Si no se conocen las varianzas poblacionales pero se pueden suponer iguales	407
10.8.3. Si no se conocen las varianzas poblacionales σ_1^2 y σ_2^2 y no se pueden suponer iguales	408
10.9. Intervalo de confianza para la diferencia de proporciones $p_1 - p_2$..	409
10.10. Intervalo de confianza para el cociente de varianzas	412
Tabla de intervalos de confianza (distribución de los estadísticos en el muestreo)	414
10.11. Problemas propuestos	417
11. Contrastes de hipótesis paramétricos	
11.1. Introducción	423
11.2. Etapas a seguir en un contraste paramétrico	424
11.3. La noción de riesgo. Errores	429
11.4. El P-valor	437
11.5. Prueba de conformidad con una proporción teórica (binomial)	439
11.5.1. Prueba bilateral para p	439
11.5.2. Pruebas unilaterales para p	441
11.6. Prueba de conformidad con una media teórica	443
11.6.1. Pruebas bilaterales para μ	443
11.6.2. Pruebas unilaterales para μ	445
11.7. Prueba de conformidad con una varianza teórica	447
11.7.1. Prueba bilateral para σ^2	447
11.7.2. Pruebas unilaterales para σ^2	448
11.8. Prueba de comparación de dos proporciones	450
11.8.1. Contraste bilateral	450

11.8.2. Contrastes unilaterales	452
11.9. Prueba de comparación de dos varianzas	453
11.9.1. Contraste bilateral	454
11.9.2. Contrastes unilaterales	455
11.10. Prueba de comparación de dos medias	456
11.10.1. Contrastes bilaterales	456
11.10.2. Contrastes unilaterales	460
11.11. ¿Cómo hacer los contrastes con <i>STATGRAPHICS</i> ?	464
Tablas de contrastes de hipótesis paramétricos	471
11.12. Problemas propuestos	473
12. Contrastes no paramétricos	
12.1. Introducción	479
12.2. Prueba χ^2 de bondad de ajuste para el contraste de proporciones. Ajuste a un modelo multinomial	480
12.3. Prueba χ^2 de bondad de ajuste de una distribución observada a una teórica	487
12.3.1. El modelo especifica completamente la distribución a ajustar	487
12.3.2. El modelo sólo especifica el tipo de distribución a ajustar, pero no se conocen los valores de los parámetros del modelo	488
12.4. Prueba χ^2 de independencia de dos variables	496
12.5. Prueba χ^2 de homogeneidad	500
12.6. Prueba de Kolmogorov-Smirnov para una muestra	503
12.6.1. ¿Cómo hacerlo con <i>STATGRAPHICS</i> ?	507
12.7. Contraste χ^2 de McNemar para dos muestras apareadas	509
12.8. Prueba de los rangos con signo de Wilcoxon para dos muestras apareadas	512
12.9. Contraste U de Mann-Whitney	515
12.10. Prueba Q de Cochran para $k \geq 3$ muestras apareadas	519
12.11. Contraste de rangos de Friedman para $k \geq 3$ muestras apareadas	521
12.12. Prueba de Kruskal-Wallis para más de dos muestras independientes	524
12.12.1. ¿Cómo hacerlo con <i>STATGRAPHICS</i> ?	527
12.13. Problemas propuestos	530
13. Análisis de la varianza y análisis de la regresión	
13.1. Introducción al ANOVA	539
13.2. ANOVA con un factor de variación	542
Tabla ANOVA con un factor de variación	546
13.2.1. Contraste de Hartley	548
13.2.2. Prueba de Bartlett	549

13.2.3. ¿Cómo hacerlo con STATGRAPHICS?	554
13.2.4. Comparaciones múltiples	556
13.3. Análisis de la regresión lineal simple	559
13.3.1. Contraste de hipótesis para el parámetro β	563
13.3.2. Contraste de significación de la regresión lineal simple	566
Tabla de contraste de significación de la regresión lineal simple	567
13.3.3. Análisis de la regresión lineal simple con varias medidas de la variable Y para un mismo valor de X	573
Tabla para análisis de la regresión lineal simple	576
13.3.4. Intervalos de confianza para β	578
13.3.5. Intervalos de confianza para $\mu_{Y X=x} = \alpha + \beta x$	579
13.3.6. ¿Cómo hacerlo con STATGRAPHICS?	581
13.4. Problemas propuestos	584

CUARTA PARTE

UNA INTRODUCCIÓN A LOS PROCESOS ESTOCÁSTICOS Y A LA GEOESTADÍSTICA

14. Una introducción a los procesos estocásticos y a la estadística espacial: Métodos geoestadísticos

14.1. Introducción	595
14.2. Un modelo de proceso estocástico temporal en tiempo discreto. Cadenas de Markov finitas	598
14.3. Un modelo de proceso estocástico temporal en tiempo conti- nuo. Proceso de Poisson temporal	604
14.4. Un modelo de proceso estocástico espacial. Proceso espacial de Poisson	610
14.5. Estadística espacial. Métodos geoestadísticos	612
14.6. Funciones aleatorias. Variables regionalizadas	614
14.7. Análisis exploratorio de los datos	619
14.8. Análisis estructural. Variograma y semivariograma	622
14.9. Semivariograma experimental	623
14.10. Funciones aleatorias estacionarias	628
14.11. Funciones aleatorias intrínsecas	631
14.12. Modelos teóricos de semivariograma	632
14.12.1. Modelo efecto pepita puro	633
14.12.2. Modelo lineal	634
14.12.3. Modelo potencial	634
14.12.4. Modelo lineal acotado	635
14.12.5. Modelo esférico	636
14.12.6. Modelo esférico con efecto pepita	636

14.12.7. Modelo gaussiano	637
14.12.8. Modelo exponencial	638
14.12.9. Modelo exponencial con pepita	638
14.13. Análisis estructural de los datos	640
14.14. Estimación espacial. Krigeado	641
14.15. Krigeado simple	644
14.16. El krigeado ordinario	650
14.17. Regularización. Variables regularizadas	660
14.18. Ecuaciones del krigeado por bloques	663
14.19. Intervalos de confianza para las estimaciones	668
14.20. Representaciones gráficas con el programa SURFER para ordenador personal	669
14.21. ¿Cómo hacer representaciones gráficas con SURFER?	671
14.22. Problemas propuestos	677

APÉNDICE I

Tabla I. Áreas bajo la curva normal	685
Tabla II. Áreas bajo la curva de densidad de la χ^2_n a la izquierda de $\chi^2_{\alpha,n}$...	686
Tabla III. Áreas bajo la curva de densidad de la t_n a la izquierda de $t_{\alpha,n}$...	687
Tabla IV. Áreas bajo la curva de densidad de la $F_{[n_1, n_2]}$ a la izquierda de $F_{[n_1, n_2];\alpha}$, $\alpha = 0,05$	688
Tabla V. Áreas bajo la curva de densidad de la $F_{[n_1, n_2]}$ a la izquierda de $F_{[n_1, n_2];\alpha}$, $\alpha = 0,01$	689
Tabla VI. Áreas bajo la curva de densidad de la $F_{[n_1, n_2]}$ a la izquierda de $F_{[n_1, n_2];\alpha}$, $\alpha = 0,025$	690
Tabla VII. Áreas bajo la curva de densidad de la $F_{[n_1, n_2]}$ a la izquierda de $F_{[n_1, n_2];\alpha}$, $\alpha = 0,005$	691
Tabla VIII. Números aleatorios	692
Tabla IX. Valores críticos $H_{\alpha,[k,n-1]}$ del contraste de Hartley de homogeneidad de las varianzas	693
Tabla X. Valores críticos $D_{n;\alpha}$ del contraste de Kolmogorov-Smirnov para una muestra	694

APÉNDICE II

Soluciones a los problemas propuestos	697
Bibliografía	747
Índice de materias	753

POR QUÉ Y PARA QUÉ SE HA ESCRITO ESTE LIBRO

¿Un libro más de Estadística? No, un libro distinto. No es un libro teórico, no se hace hincapié en las demostraciones, no es un manual de un programa estadístico. Se trata de dar *una visión instrumental* que combina el método de resolución manual con el uso de software de estadística.

Para muchas personas que necesitan utilizar la estadística ésta es un misterio. Recurren al *software* estadístico que les proporciona resultados. Al desconocer la materia, creen en la información que obtienen sin entender cómo se ha llegado a ella ni si es correcto ni apropiado el análisis realizado. Esta situación es similar a la de quien no conoce una escritura jeroglífica y pretende traducir e interpretar un texto en ese lenguaje.

En este libro se presentan los temas de forma gradual y secuenciada, especificando las hipótesis necesarias y razonando los procedimientos. Se intercalan ejemplos y ejercicios resueltos, conjugando los métodos de resolución manual con el uso de programas de estadística para ordenador. De esta forma se ayuda a comprender el procedimiento estadístico y a valorar las ventajas de los programas que facilitan los cálculos y el manejo de grandes cantidades de datos.

¿Por qué ahora? Los nuevos planes de estudio, como consecuencia de la adecuación de las titulaciones a las nuevas exigencias derivadas de la Declaración de Bolonia, que sienta las bases para la Construcción de un Espacio Europeo de Educación Superior, EEES, hacen hincapié en la competencia. Los futuros titulados, para abrirse camino en el campo profesional tanto en España como en el resto del mundo globalizado en el que vivimos, han de demostrar ser expertos, conocedores de las materias no sólo de forma teórica sino también acreditar ser capaces de aplicar los conocimientos adquiridos.

Otra de las exigencias de la Convergencia Europea es que el alumno ha de ser sujeto activo y responsable en su formación, es decir, el alumno universitario tiene que ser cada vez más gestor de su aprendizaje. Unido esto a la reducción del número de horas lectivas dedicadas a la explicación por parte del profesor, refuerza la importancia de los libros y demás herramientas útiles para el aprendizaje.

El profesor tendrá un papel más de tutor que de transmisor de conocimientos y ha de proporcionar al alumno una visión práctica e interdisciplinar.

Especialmente útiles y necesarios son los métodos y las técnicas estadísticas tanto en las ciencias experimentales, como en la medicina, en las ciencias sociales, en economía, en la ingeniería o en la industria. La estadística es una materia fundamental en la Ciencia y en la Técnica que permite dar informaciones objetivas en todos los campos.

Este libro contiene más de 500 problemas entre los resueltos con detalle y los propuestos, de los que se incluye la solución de todos en el Apéndice II. Los ejemplos y los problemas tratan de hacer más sencilla la comprensión de la teoría, de hacer visible lo que es abstracto y más ameno y atractivo el aprendizaje, a la vez que proporcionan aplicaciones que pueden sugerir otras en el lector.

¿A quién va dirigido? A todas aquellas personas que necesiten hacer uso de la estadística, alumnos de grado y postgrado de las facultades y escuelas técnicas, investigadores y a todos aquellos que sientan curiosidad por las técnicas estadísticas y sus aplicaciones. Esperamos que a todos les anime y les haga disfrutar al obtener resultados que pueden interpretar en su campo de trabajo.

PRÓLOGO

El propósito principal de esta obra es presentar la estadística desde el punto de vista de sus aplicaciones, sin detenerse en demostraciones ni profundizar en temas muy especializados. Como dice Mood: «La estadística es la tecnología de la investigación científica».

En el siglo XXI no se puede concebir una investigación en las ciencias experimentales, en medicina, en las ciencias sociales, en la técnica, en la industria, que no utilice la estadística y el ordenador, que ha facilitado el cálculo y el manejo de gran cantidad de datos. El estudiante y el investigador de hoy necesitan analizar los datos que recogen en su campo de trabajo y se encuentran, en numerosas ocasiones, en situaciones de incertidumbre, lo que hace necesaria la utilización de métodos estadísticos para sacar de su estudio mejores conclusiones.

Los métodos estadísticos son de dos tipos: descriptivos e inductivos. El objeto de los descriptivos es ordenar, resumir y analizar los datos recogidos, mientras que los inductivos tratan de obtener conclusiones a partir de los datos de la muestra analizada, es decir, a partir del conocimiento de una muestra se establecen inferencias sobre la población de la que se ha obtenido dicha muestra y se contrastan.

Este libro está dividido en cuatro partes. La primera se dedica a la estadística descriptiva, unidimensional, si de cada elemento de la muestra se recoge información de una sola característica, bidimensional, si se analizan dos características de cada elemento observado. En la segunda parte se introduce la probabilidad y se presentan distribuciones de probabilidad, para estudiar en la tercera parte la inferencia estadística que permite hacer estimaciones, formular hipótesis y contrastarlas. La cuarta y última parte del libro se dedica a una introducción de los procesos estocásticos, la estadística espacial y la geoestadística. La geoestadística estudia las variables numéricas $Z(x)$ distribuidas en el espacio, o variables regionalizadas, porque a cada valor observado o desconocido se asocia una posición en el espacio. La geoestadística es una técnica relativamente reciente que tiene su origen en el estudio de las explotaciones mineras, y que hoy ha encontrado aplica-

ciones en la biología, en la medicina, en hidrología, en edafología, y en otros campos muy diversos, dando lugar a una rama de investigación fascinante y en continuo desarrollo.

Los catorce temas de los que se compone el libro están ordenados de forma secuencial. Es conveniente, y recomendable, seguir el orden presentado y hacer un razonable esfuerzo por comprender los ejemplos. Los capítulos se ilustran con ejercicios y problemas resueltos tomados de diferentes campos, y al final de cada uno se insertan problemas propuestos cuya solución se incluye en el Apéndice II.

Se comienza con ejercicios sencillos para pasar a resolver otros más complicados o que requieren hacer uso del ordenador para facilitar los cálculos. Son muchos los programas que se pueden utilizar para el análisis estadístico en un ordenador personal, de entre ellos hemos elegido *STATGRAPHICS*^{TM1} Plus para Windows, Versión 5.1, cuyo uso está ampliamente extendido, por ser ésta la versión del programa más empleada en la actualidad, por su facilidad de manejo y por los resultados gráficos que ofrece; *Excel*^{TM2}, Versión 2003, porque permite formar tablas detalladas de los cálculos; *DERIVE*^{TM3}, Versión 6.10, para efectuar cálculos y para el dibujo de algunas gráficas, *SURFER*^{TM4}, Versión 8, y la Versión 4 de *GRAPHER*TM, para la geoestadística.

Agradecemos al analista del Departamento de Apoyo a Investigación de los Servicios Informáticos de Apoyo a la Docencia e Investigación, Área de Informática y Comunicaciones de la UCM, Dr. D. Santiago Cano Alsúa, su continua orientación y asesoramiento en la elección y uso del software.

Los ejercicios se resuelven a mano, haciendo uso de calculadora, hojas de cálculo y tablas, que se incluyen en el Apéndice I, y también, si es posible, con *STATGRAPHICS*. El conjugar ambos procedimientos ayuda, entre otras cosas, a comprender mejor el método para obtener la solución, a interpretar los resultados que proporcionan los programas, a valorar la ayuda que supone poder hacer uso de *software*, que ahorra tiempo facilitando los cálculos y que permite manejar grandes cantidades de datos, y a desarrollar el espíritu crítico fomentando el hábito de no aceptar sin justificación ningún resultado. Familiarizarse con la utilización de algún programa estadístico e interpretar la salida facilita la comprensión de cualquier salida de otro programa y esto es hoy en día imprescindible en el trabajo científico.

Queremos agradecer a todas las personas que han contribuido a que este libro se haya podido escribir, a nuestros profesores que nos han ayudado a formarnos, a los compañeros principalmente de las facultades de Ciencias Biológicas y de Ciencias Geológicas de la UCM por animarnos a redactarlo y por sus comentarios, a los compañeros que nos han proporcionado datos, en especial a la Dra. D.^a Paloma Sevilla García y al Dr. D. Sergio Rodríguez García, ambos profesores del

¹ *STATGRAPHICS* es una marca registrada por Manugistics Inc. and Statistical Graphics Corporation. USA.

² *Excel* es una marca registrada por Microsoft Corporation.

³ *DERIVE* es una marca registrada por Texas Instruments Incorporated.

⁴ *SURFER* y *GRAPHER* son marcas registradas por Golden Software, Inc.

Departamento de Paleontología de la Facultad de Ciencias Geológicas, a los antiguos alumnos que nos han alentado en esta tarea, a la Editorial Díaz de Santos, por haber confiado de nuevo en nuestro trabajo, a la Imprenta Fernández Ciudad por su esmero en la fotocomposición y por su saber hacer, mejorando la presentación y el resultado final de esta obra, a nuestras familias por el tiempo que les hemos sustraído durante la elaboración de este libro.

Gracias al Dr. D. Benjamín Fernández Ruiz, catedrático del Departamento de Biología Celular, al Dr. D. Antonio Tormo Garrido, actual Decano de la Facultad de Ciencias Biológicas, al Dr. D. Eduardo de Juana Aranzana, Presidente de la Sociedad Española de Ornitología (SEO), y al Dr. D. Carlos Vicente Córdoba, catedrático del Departamento de Biología Vegetal I (Botánica y Fisiología Vegetal) por su generosidad al cedernos las fotografías que ilustran la portada del libro.

Nuestro agradecimiento muy especial al Dr. D. Francisco José Cano Sevilla, catedrático de Estadística e Investigación Operativa de la UCM y asesor del Consejo de Coordinación Universitaria, por su valiosa colaboración en la lectura y corrección del original. Sus acertadas observaciones nos han servido de gran ayuda.

Queremos hacer constar nuestro homenaje póstumo y nuestra gratitud a dos catedráticos: D. Rafael García Aráez, fallecido el pasado 28 de mayo, que fue catedrático de Matemáticas de Instituto y de la E.U.I.T. Forestal de la UPM, donde ocupó también el cargo de Jefe de Estudios desde 1975 hasta 1986, y D. Sixto Ríos García, el Padre de la Estadística en España, fallecido el pasado 8 de julio, que fue catedrático de Estadística en la Facultad de Matemáticas de la UCM, Académico numerario de la Real Academia de Ciencias Exactas, Físicas y Naturales desde 1961, Fundador y Director del Instituto de Estadística e Investigación Operativa del CSIC y de la primera Escuela de Estadística de España. Los que hemos tenido la suerte de conocerlos y de trabajar con ellos no los podemos olvidar y les estaremos siempre agradecidos.

Se lo dedicamos a todos ellos, a los futuros alumnos, a los investigadores que necesiten hacer uso de la estadística y a todos aquellos que precisen aplicarla, para los que hemos hecho esta obra de un modo especial, con el deseo de que les ayude en su formación y en sus trabajos.

Si a las personas que utilicen este libro les resulta provechoso habremos logrado nuestro objetivo y nos sentiremos satisfechos. También estaremos muy agradecidos a todos los que nos remitan sus opiniones, sugerencias o comentarios a la siguiente dirección: mtgmant@bio.ucm.es

LOS AUTORES

Capítulo preliminar

1

1.1. INTRODUCCIÓN

La estadística es una rama de las matemáticas con origen remoto y en continua evolución y desarrollo. Hasta el siglo XVII se entendía por estadística la recopilación de datos para la administración del Estado. Ya en el Imperio Romano se hizo enumeración y recuento de soldados, medios de transporte, riquezas... y se tiene constancia de que en tiempos de César Augusto se realizó un censo de la población. De *status*, o estado de las cosas, parece derivarse la palabra estadística. Aquí se encuentra el origen de la estadística descriptiva.

Vivimos inmersos en un mundo de cifras: evolución del paro, variaciones en los índices de precios (IPC), gastos familiares, índice de la bolsa, IBEX-35, cotizaciones bursátiles, número de accidentes de circulación, censo electoral, porcentaje de personas que padecen una enfermedad, predicción del tiempo, resultados de unas elecciones, porcentaje de hogares que utilizan Internet, eficacia de una campaña publicitaria, prospecciones petrolíferas y de minas, relación entre el número de hijos nacidos con Síndrome de Down y la edad de la madre, frecuencias genotípicas en una población y frecuencias fenotípicas que son objeto de estudio de la genética de poblaciones, gasto por usuario de teléfono móvil, etc. Hay dos formas de ver los datos, la del periodista, que se ocupa de la anécdota, y la del estadístico, que está interesado por la regularidad.

En la prensa escrita, en los medios audiovisuales, en los textos aparecen recuentos y porcentajes.

En la Figura 1.1 se puede observar doble información. En los sectores se representa el número de habitantes que eran usuarios de Internet en 2006, en cabeza está Asia. Además, se refleja el porcentaje de usuarios de Internet en relación al número total de habitantes de cada una de las poblaciones examinadas; en este aspecto Asia baja al quinto lugar de las siete porciones en las que se ha dividido la población mundial.

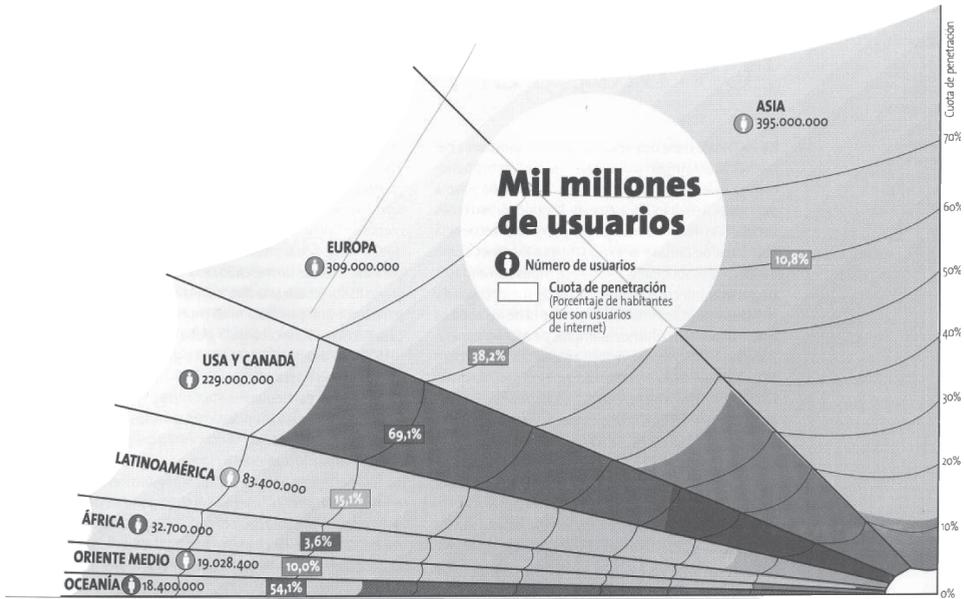


FIGURA 1.1. *Usuarios de Internet en 2006.*

Se advierte que las cantidades absolutas dan una información importante, pero más pobre que los porcentajes, si se pretende comparar resultados.

También se hace referencia a estadísticos como medias, medianas, cuartiles, etc., como se observa en las Figuras 1.2. y 1.3 que aparecen a continuación.

	01-02	02-03	03-04	Media l/m ²
Septiembre	43,0	42,3	39,4	46,4
Octubre	67,3	75,7	114,7	68,5
Noviembre	41,9	93,1	83,7	72,3
Diciembre	27,4	72,7	65,0	73,8

Infografía ABC (En negrita) Superior a la media desde 1930

Figura 1.2. *Datos históricos de pluviometría.*