



Sobre el autor

Sheldon M. Ross obtuvo su doctorado en Estadística por la Universidad de Stanford en 1968, tras ello se unió al Departamento de Ingeniería Industrial e Investigación Operativa de la Universidad de California, en Berkeley. Allí permaneció hasta el otoño de 2004, año en que obtuvo la plaza de Profesor de Ingeniería Industrial y de Sistemas en el Departamento Daniel J. Epstein de la Universidad del Sur de California. Ha publicado un gran número de artículos técnicos y de libros de texto en distintas áreas de Estadística y Probabilidad Aplicada. Entre sus libros de texto figuran *Un primer curso de Probabilidad* (sexta edición), *Introducción a los modelos de Probabilidad* (octava edición), *Simulación* (tercera edición) e *Introducción a la Probabilidad y a la Estadística para ingenieros y científicos* (tercera edición).

El profesor Ross es fundador de la revista *Probability in the Engineering and Informational Sciences*, revista de la que continúa siendo editor. Es miembro del Instituto de Estadística Matemática y ha recibido el Premio Humboldt para los Científicos Senior de Estados Unidos.

ción y la síntesis de los datos. La segunda rama, denominada Estadística Inferencial, tiene por objeto la extracción de conclusiones a partir de los datos.

Los capítulos 2 y 3 están dedicados a la Estadística Descriptiva. En el capítulo 2 se analizan los métodos gráficos y tabulares que permiten presentar los conjuntos de datos. Se ve cómo una presentación efectiva de los conjuntos de datos facilita a menudo el descubrir algunas de sus características esenciales. El capítulo 3 muestra cómo se pueden resumir o sintetizar determinadas características de los conjuntos de datos.

Para poder sacar conclusiones a partir de los datos es preciso entender qué representan. Por ejemplo, habitualmente se asume que los datos constituyen una “muestra aleatoria procedente de una determinada población”. Para entender exactamente lo que significa esta frase, y otras similares, es necesario tener ciertos conocimientos de Probabilidad; ésta es la finalidad del capítulo 4. El estudio de la Probabilidad es controvertido en una clase de introducción a la Estadística, porque suele ser bastante difícil para los estudiantes. Como resultado, ciertos libros de texto rebajan la importancia de este tema y lo presentan de forma superficial. Nosotros hemos elegido un enfoque diferente: hemos intentado concentrar la atención en los aspectos esenciales y presentarlos de forma clara y fácilmente entendible. Así pues, hemos tratado de forma breve, aunque cuidadosamente, los sucesos de un experimento, las propiedades de las probabilidades asignadas a dichos sucesos y los conceptos de probabilidad condicionada e independencia. El estudio de la Probabilidad continúa en el capítulo 5, donde se introducen las variables aleatorias discretas, y en el capítulo 6, dedicado a las variables aleatorias normales y a otras variables aleatorias continuas.

En el capítulo 7 se tratan las distribuciones de probabilidad de los estadísticos asociados al muestreo. También se analiza por qué la distribución normal tiene una importancia fundamental en la Estadística.

El capítulo 8 aborda el problema de utilizar los datos para estimar determinados parámetros de interés. Por ejemplo, podríamos desear estimar la proporción de personas que están en la actualidad a favor de una determinada medida política. Se estudian dos tipos de estimadores: el primero de ellos estima una magnitud de interés mediante un número (por ejemplo, se podría estimar que el 52% de la población está a favor de la medida política); el segundo tipo proporciona el estimador en la forma de un intervalo (por ejemplo, se podría estimar que el porcentaje de la población a favor de la medida política está comprendido entre el 49% y el 55%).

El capítulo 9 introduce un tema importante dedicado a los contrastes estadísticos de hipótesis, en el que se utilizan los datos para contrastar la plausibilidad de determinadas hipótesis. Por ejemplo, en un determinado contraste se podría rechazar la hipótesis de que más de un 60% de la población de votantes está a favor de una propuesta dada. También se incorpora el concepto de p -valor, que mide el grado de plausibilidad de una hipótesis una vez que se han observado los datos.

Mientras que los contrastes del capítulo 9 afectan a una sola población, los del capítulo 10 hacen referencia a dos poblaciones distintas. Por ejemplo, podemos estar interesados en contrastar si las proporciones de hombres y de mujeres a favor de una determinada propuesta coinciden.

Probablemente la técnica de inferencia estadística más extendida es la conocida como análisis de la varianza, que se expone en el capítulo 11. Dicha técnica nos permite contrastar hipótesis sobre parámetros que dependen de distintos factores. En este capítulo se analizan situaciones de análisis de la varianza unifactoriales y bifactoriales.

En el capítulo 12 se presenta la regresión lineal y cómo ésta se puede utilizar para relacionar el valor de una variable (digamos la altura de un hombre) con el de otra (la altura de su padre). Se analiza el concepto de regresión a la media y también se introduce y se explica detalladamente la falacia de la regresión. Se examina la relación entre regresión y correlación. Finalmente, en una sección opcional, se utiliza la regresión a la media junto con el teorema central del límite para presentar un razonamiento simple y original que explica por qué los conjuntos de datos biológicos suelen habitualmente seguir una distribución normal.

En el capítulo 13 se exponen los contrastes de bondad de ajuste, que sirven para contrastar si un determinado modelo propuesto es consistente con los datos. En este capítulo también se consideran poblaciones clasificadas según dos características y se muestra cómo contrastar si las características de un miembro de la población elegido aleatoriamente son independientes.

El capítulo 14 se centra en los contrastes de hipótesis no paramétricos, que son contrastes que se pueden usar en situaciones en las que los contrastes de los capítulos precedentes resultan inapropiados.

En el capítulo 15 se introduce el control de calidad, una técnica estadística clave relacionada con los procesos de transformación y producción.

Novedades de esta edición

Esta edición incluye muchos ejemplos y ejercicios nuevos y actualizados. Entre las secciones nuevas se hallan las siguientes:

- La sección 4.7, que es opcional y está dedicada a los principios de conteo.
- La sección 5.7, igualmente opcional, en la que se introducen las variables aleatorias de Poisson.
- La sección 12.10, en la que se evalúan los modelos de regresión mediante el análisis de los residuos.
- Las nuevas secciones del capítulo 15, sobre el control de calidad, en las que se introducen los gráficos de control de medias móviles ponderadas exponencialmente y de sumas acumuladas.

Contenido

Sobre el autor	v
Prefacio	xiii
Agradecimientos	xvii

1 Introducción a la Estadística 1

1.1	Introducción	1
1.2	La naturaleza de la Estadística	3
1.3	Poblaciones y muestras	5
1.4	Breve historia de la Estadística	7
	<i>Problemas</i>	10
	<i>Distintas definiciones de la Estadística</i>	13
	<i>Términos clave</i>	13

2 Descripción de los conjuntos de datos 15

2.1	Introducción	15
2.2	Tablas y gráficos de frecuencias	16
2.3	Datos agrupados e histogramas	28
2.4	Gráficos de tallos y hojas	40
2.5	Conjuntos de datos apareados	49
2.6	Comentarios históricos	56
	<i>Términos clave</i>	57
	<i>Resumen</i>	58
	<i>Problemas de repaso</i>	61

3 Uso de la Estadística para sintetizar conjuntos de datos 69

3.1	Introducción	70
3.2	Media muestral	71
3.3	Mediana muestral	80
3.4	Moda muestral	96
3.5	Varianza muestral y desviación típica muestral	98

3.6	Conjuntos de datos normales y la regla empírica	108
3.7	Coefficiente de correlación muestral	121
	<i>Términos clave</i>	135
	<i>Resumen</i>	136
	<i>Problemas de repaso</i>	138

4 Probabilidad 143

4.1	Introducción	143
4.2	Espacio muestral y sucesos de un experimento	144
4.3	Propiedades de la Probabilidad	151
4.4	Experimentos con resultados igualmente probables	159
4.5	Probabilidad condicionada e independencia	166
*4.6	Teorema de Bayes	184
*4.7	Principios de recuento	189
	<i>Términos clave</i>	198
	<i>Resumen</i>	199
	<i>Problemas de repaso</i>	201

5 Variables aleatorias discretas 209

5.1	Introducción	209
5.2	Variables aleatorias	210
5.3	Valor esperado	217
5.4	Varianza de las variables aleatorias	230
5.5	Variables aleatorias binomiales	237
*5.6	Variables aleatorias hipergeométricas	246
*5.7	Variables aleatorias de Poisson	248
	<i>Términos clave</i>	252
	<i>Resumen</i>	252
	<i>Problemas de repaso</i>	254

6 Variables aleatorias normales 259

6.1	Introducción	260
6.2	Variables aleatorias continuas	260
6.3	Variables aleatorias normales	264
6.4	Probabilidades asociadas a la variable aleatoria normal estándar	269
6.5	Búsqueda de las probabilidades de la normal: conversión a la normal estándar	276
6.6	Propiedad aditiva de las variables aleatorias normales	278
6.7	Percentiles de las variables aleatorias normales	283
	<i>Términos clave</i>	289
	<i>Resumen</i>	289
	<i>Problemas de repaso</i>	292

7 Distribuciones de los estadísticos asociados al muestreo 295

7.1	Preámbulo	296
7.2	Introducción	296

7.3	Media muestral	297
7.4	Teorema central del límite	302
7.5	Muestreo de proporciones en poblaciones finitas	311
7.6	Distribución de la varianza muestral de una población normal	321
	<i>Términos clave</i>	324
	<i>Resumen</i>	324
	<i>Problemas de repaso</i>	325

8 Estimación 329

8.1	Introducción	329
8.2	Estimador puntual de la media de una población	330
8.3	Estimador puntual de una proporción poblacional	334
8.4	Estimación de la varianza de una población	340
8.5	Estimadores por intervalo para la media de una población normal con varianza conocida	345
8.6	Estimadores por intervalo para la media de una población normal con varianza desconocida	357
8.7	Estimadores por intervalo de una proporción poblacional	368
	<i>Términos clave</i>	378
	<i>Resumen</i>	378
	<i>Problemas de repaso</i>	381

9 Contraste de hipótesis estadísticas 385

9.1	Introducción	385
9.2	Contrastes de hipótesis y niveles de significación	386
9.3	Contrastes relativos a la media de una población normal: el caso de la varianza conocida	392
9.4	Contrastes de la t para la media de una población normal: el caso de la varianza desconocida	407
9.5	Contrastes de hipótesis sobre proporciones poblacionales	418
	<i>Términos clave</i>	428
	<i>Resumen</i>	428
	<i>Problemas de repaso</i>	432

10 Contrastes de hipótesis relativas a dos poblaciones 437

10.1	Introducción	437
10.2	Contraste de la igualdad de medias de dos poblaciones normales: caso de varianzas conocidas	439
10.3	Contraste de la igualdad de medias: varianzas desconocidas y tamaños muestrales grandes	446
10.4	Contraste de la igualdad de medias: contrastes con muestras pequeñas cuando las varianzas poblacionales son desconocidas pero iguales	455
10.5	Contraste de la t con muestras apareadas	463
10.6	Contraste de la igualdad de proporciones poblacionales	472
	<i>Términos clave</i>	484

<i>Resumen</i>	484
<i>Problemas de repaso</i>	488

11 Análisis de la varianza 493

11.1	Introducción	493
11.2	Análisis de la varianza unifactorial	495
11.3	Análisis de la varianza bifactorial: introducción y estimación de parámetros	503
11.4	Análisis de la varianza bifactorial: contraste de hipótesis	509
11.5	Comentarios finales	518
	<i>Términos clave</i>	518
	<i>Resumen</i>	519
	<i>Problemas de repaso</i>	522

12 Regresión lineal 525

12.1	Introducción	526
12.2	Modelo de regresión lineal simple	527
12.3	Estimación de los parámetros de regresión	531
12.4	Variable aleatoria de error	541
12.5	Contraste de la hipótesis de que $\beta = 0$	545
12.6	Regresión a la media	552
12.7	Intervalos de predicción para respuestas futuras	562
12.8	Coefficiente de determinación	567
12.9	Coefficiente de correlación muestral	571
12.10	Análisis de los residuos: evaluación del modelo	573
12.11	Modelo de regresión lineal múltiple	576
	<i>Términos clave</i>	582
	<i>Resumen</i>	582
	<i>Problemas de repaso</i>	586

13 Contrastes de bondad de ajuste de la chi-cuadrado 593

13.1	Introducción	594
13.2	Contrastes de bondad de ajuste de la chi-cuadrado	596
13.3	Contraste de la independencia en poblaciones clasificadas de acuerdo con dos características	608
13.4	Contraste de la independencia en las tablas de contingencia con los totales marginales fijos	618
	<i>Términos clave</i>	624
	<i>Resumen</i>	624
	<i>Problemas de repaso</i>	627

14 Contrastes de hipótesis no paramétricos 633

14.1	Introducción	633
14.2	Contraste de signos	634
14.3	Contraste de rangos signados	642

14.4	Contraste de la suma de rangos para comparar dos poblaciones	651
14.5	Contraste de rachas para la aleatoriedad	659
	<i>Términos clave</i>	666
	<i>Resumen</i>	666
	<i>Problemas de repaso</i>	669

15 Control de calidad 671

15.1	Introducción	671
15.2	Gráficos de control de \bar{X} para detectar un deslizamiento en la media	672
15.3	Gráficos de control para la fracción de defectos	687
15.4	Gráficos de control de medias móviles ponderadas exponencialmente	689
15.5	Gráficos de control de sumas acumuladas	694
	<i>Términos clave</i>	697
	<i>Resumen</i>	697
	<i>Problemas de repaso</i>	698

Apéndices 701

A.	Un conjunto de datos	703
B.	Preliminares matemáticos	709
C.	Cómo seleccionar una muestra aleatoria	713
D.	Tablas	717
	<i>Tabla D.1 Probabilidades de la normal estándar</i>	717
	<i>Tabla D.2 Percentiles $t_{n,\alpha}$ de las distribuciones t</i>	718
	<i>Tabla D.3 Percentiles $\chi^2_{n,\alpha}$ de las distribuciones chi-cuadrado</i>	720
	<i>Tabla D.4 Percentiles de las distribuciones F</i>	722
	<i>Tabla D.5 Funciones de distribución binomiales</i>	728
E.	Programas	735
	Respuestas a los problemas con número impar	737
	Índice	795



Introducción a la Estadística

Los estadísticos han invadido todas las ramas de la ciencia con una rapidez de conquista que sólo tiene como rivales a Atila, a Mahoma y al escarabajo de Colorado.

Maurice Kendall (estadístico británico)

1.1	Introducción	1
1.2	La naturaleza de la Estadística	3
1.3	Poblaciones y muestras	5
1.4	Breve historia de la Estadística	7
	Problemas	10
	Distintas definiciones de la Estadística	13
	Términos clave	13

Este capítulo introduce la materia objeto de la Estadística, el arte de aprender de los datos. Describe las dos ramas de la Estadística, la descriptiva y la inferencial. Se analiza la idea de aprender sobre una población a través de muestrear y estudiar a algunos de sus miembros. Finalmente se presentan algunos rasgos históricos.

1.1 Introducción

¿Es mejor que nuestros hijos sean escolarizados antes o después? Esta es una cuestión de interés para muchos padres y también para los gestores públicos. ¿Cómo se puede responder?

Inicialmente, parece razonable que nos planteemos esto a partir de nuestra propia experiencia y de algunas conversaciones mantenidas con los amigos. Sin embargo, si se quiere convencer a otras personas y obtener consensos, resulta necesario reunir algún tipo de información objetiva. Por ejemplo, en muchos Estados, los niños deben someterse a exámenes o a pruebas de conocimiento al final de su primer año de escolarización. Se pueden conseguir los resultados de los niños en dichas pruebas y analizarlos después para

1.2 La naturaleza de la Estadística

En el mundo de hoy, el que uno debe primero reunir datos para aprender sobre algo se ha convertido en un axioma. Por ejemplo, el primer paso para aprender sobre temas como

1. El estado actual de la economía.
2. El porcentaje de votantes a favor de una propuesta.
3. El número medio de kilómetros que puede recorrer un automóvil de nueva fabricación con un litro de gasolina.
4. La eficacia de un nuevo medicamento.
5. La utilidad de un nuevo método de enseñanza de lectura para niños de escuela elemental.

consiste en compilar los datos relevantes.

Definición

La *Estadística* es el arte de aprender a partir de los datos. Está relacionada con la recopilación de datos, su descripción subsiguiente y su análisis, lo que nos lleva a extraer conclusiones.

1.2.1 Obtención de datos

En ocasiones un análisis estadístico comienza con un conjunto de datos; por ejemplo, el gobierno habitualmente reúne datos sobre la tasa de desempleo y sobre el producto interior bruto. La Estadística se utiliza después para describir, clasificar y analizar esos datos.

En otras situaciones, los datos no están disponibles, y la Estadística se puede usar para diseñar un experimento apropiado para generar dichos datos. El experimento elegido dependería de la utilidad que se quiera obtener de los datos. Por ejemplo, si se acaba de desarrollar un medicamento reductor del colesterol y se quiere determinar su eficacia, se deben reclutar voluntarios y anotar sus niveles de colesterol. Después se les suministrará el medicamento durante cierto periodo de tiempo, y posteriormente se volverán a medir sus niveles de colesterol. Sin embargo, el experimento sería ineficaz si a *todos* los voluntarios reclutados se les suministrara el medicamento. Porque si fuera así, aunque los niveles de colesterol de todos los voluntarios se hubieran reducido significativamente, no estaría justificado concluir que las mejoras son debidas al medicamento en cuestión sino a alguna otra posibilidad. Es decir, está bien documentado el hecho de que cualquier medicación recibida por un paciente, tanto como si está o no directamente relacionada con la enfermedad sufrida, a menudo se traduce en mejoras en el estado del paciente. Esto se conoce como el *efecto placebo*, que no es tan sorprendente como podría parecer inicialmente, puesto que la convicción que tiene el paciente de que se le está tratando de manera efectiva a menudo conduce a una reducción de su estrés, lo cual redundaría en una mejora en su estado de salud. Adicionalmente, podrían haber existido otros factores, por lo general desconocidos, que influyeran sobre la reducción en los niveles de colesterol. Quizás el que la temperatura hubiera sido excepcionalmente cálida (o fría) podría haber hecho que los voluntarios estu-

vieran fuera de casa más o menos tiempo de lo habitual, lo que podría ser un factor determinante. Así pues, se ve que el experimento consistente en suministrar el medicamento a todos los voluntarios no está bien diseñado para generar datos a partir de los cuales se puedan sacar conclusiones acerca de la eficacia del medicamento.

Un experimento mejor intentaría neutralizar las posibles causas que afectan al nivel de colesterol, con excepción del medicamento. Una forma aceptada de conseguir esto consiste en dividir a los voluntarios en dos grupos: uno de ellos recibe el medicamento, mientras que el otro grupo recibe una pastilla (conocida como *placebo*) con la misma apariencia y sabor que el medicamento pero que no tiene ningún efecto fisiológico. Los voluntarios no deberían saber si se les está suministrando el medicamento o el placebo, y realmente sería mejor que tampoco lo supiera el personal médico que supervisa el experimento, para que sus propias actitudes no jueguen papel alguno. Adicionalmente, es deseable que la división de voluntarios en dos grupos se haga de tal forma que ninguno de los grupos se vea favorecido en el sentido de que incluya a los “mejores” pacientes. Para conseguir esto, el procedimiento generalmente más aceptado consiste en que la división de voluntarios sea “aleatoria”; se entiende por este término que la división se haga de tal forma que todas las elecciones posibles de personas que compongan el grupo que recibe el medicamento sean igualmente probables. Al grupo que no recibe tratamiento alguno (los voluntarios que reciben el placebo) se le denomina grupo *de control*.

Una vez finalizado el experimento, se describirán los datos. Por ejemplo, se presentarían los niveles de colesterol de cada voluntario antes y después del experimento, y el experimentador anotaría para cada voluntario si éste ha recibido el medicamento o el placebo. Adicionalmente, se determinarían los valores sumariales, tales como la reducción media de colesterol de los miembros del grupo de control y de los miembros del grupo tratado con el medicamento.

Definición

La parte de la Estadística relacionada con la descripción y la clasificación de los datos se conoce con el nombre de *Estadística descriptiva*.

1.2.2 Estadística inferencial y modelos de probabilidad

Cuando se ha completado el experimento, y una vez que se han descrito y clasificado los datos, deberíamos ser capaces de sacar conclusiones sobre la eficacia del medicamento. Por ejemplo, ¿se puede concluir que es efectivo como reductor de los niveles de colesterol en la sangre?

Definición

La parte de la Estadística relacionada con la extracción de conclusiones a partir de los datos se conoce con el nombre de *Estadística inferencial*.

Para poder sacar conclusiones a partir de los datos se ha de tener en cuenta el azar. Supongamos que la reducción media de colesterol es mayor para el grupo que recibió el medicamento que para el grupo de control. ¿Se puede concluir que ese resultado se debe al

Tabla 1.3 Tabla de mortalidad de Graunt

Edad de muerte	Muertes por cada 100 nacimientos
0–6	36
6–16	24
16–26	15
26–36	9
36–46	6
46–56	4
56–66	3
66–76	2
≥76	1

Nota: Las clases se acercan al valor de la derecha, pero no lo incluyen. Por ejemplo, 0-6 incluye las edades de 0 a 5 años.

empezó a registrar cifras de mortalidad en 1667; y en 1730 registrar las edades de muerte era una práctica común en toda Europa.

El término *Estadística*, que se utilizó hasta el siglo XVIII como una abreviatura de la ciencia descriptiva de los Estados, se identificó cada vez más, en el siglo XIX, con las cifras cuantitativas. Hacia 1830, en Francia e Inglaterra, el término ya fue usado de forma general como sinónimo de la *ciencia numérica* de la sociedad. Este cambio de significado se debió a que, desde 1800, los gobiernos de Europa occidental y de Estados Unidos comenzaron a recopilar y publicar sistemáticamente una gran cantidad de registros de censos y de otros tipos de tablas.

Aunque a lo largo del siglo XIX la teoría de la probabilidad había sido desarrollada por matemáticos tales como Jacob Bernoulli, Karl Friedrich Gauss y Pierre Simon Laplace, su aplicación al estudio de hechos estadísticos fue casi inexistente, ya que la mayor parte de los estadísticos sociales de la época se contentaban con dejar que los datos hablaran por sí mismos. En particular, en esa época los estadísticos no estaban interesados en sacar inferencias a partir de individuos, más bien se centraban en la sociedad en su totalidad. Por consiguiente, no estaban preocupados por el muestreo sino que intentaban obtener censos de la población al completo. Como resultado, la inferencia probabilística sobre la población a partir de muestras era prácticamente desconocida en las estadísticas sociales del siglo XIX.

No fue hasta finales de este siglo cuando los estadísticos empezaron a preocuparse por inferir conclusiones a partir de los datos numéricos. El movimiento comenzó con los trabajos de Francis Galton sobre el análisis de la influencia de la herencia a través de la utilización de técnicas que actualmente se conocen como análisis de regresión y correlación (véase el capítulo 12), que alcanzaron su mayor auge con los trabajos de Karl Pearson. Éste, que desarrolló los contrastes de bondad de ajuste (véase el capítulo 13), fue el primer director del laboratorio Galton, fundado por Galton en 1904. Allí, Pearson lideró un programa de investigación con el objetivo de desarrollar nuevos métodos en los que la Estadística se utilizaba con fines inferenciales. Su laboratorio potenció que investigadores provenientes de distintas áreas de la ciencia y la industria aprendieran los métodos estadísticos que podían tener aplicación en sus campos. Uno de los primeros estudiantes que

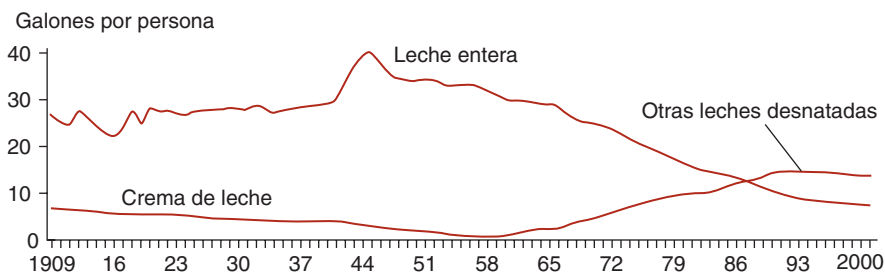
acudió a su laboratorio fue W. S. Gosset, químico de formación, que mostró su devoción por Pearson y publicó sus propios trabajos bajo el seudónimo de *Student*. Existe una famosa leyenda que mantiene que Gosset temía publicar bajo su propio nombre por miedo a que, en la fábrica de cerveza Guinness, sus jefes descubrieran que uno de sus químicos realizaba investigaciones estadísticas. Gosset es famoso por su desarrollo de los contrastes de la t (véase el capítulo 9).

Dos de las áreas más importantes de aplicación de la estadística a principios del siglo xx fueron la biología y la agricultura; todo ello gracias al interés de Pearson y de otros investigadores de su laboratorio, y también gracias a los logros del científico inglés Ronald A. Fisher. La teoría de la inferencia desarrollada por estos investigadores pioneros –y otros, entre los que se encuentran Egon Pearson, hijo de Karl Pearson, y el estadístico matemático polaco Jerzy Neyman– fue lo suficientemente general como para tratar una amplia gama de problemas cuantitativos prácticos. Como resultado, tras los primeros años del siglo xx, aumentó muy rápidamente el número de personas, procedentes de la ciencia, los negocios y la administración, que empezaron a considerar la estadística como una herramienta capaz de suministrar soluciones cuantitativas a una gran variedad de problemas científicos y prácticos.

En la actualidad, podemos encontrar las ideas estadísticas en muchos ámbitos. La Estadística descriptiva puede verse en periódicos y revistas. La *Inferencia Estadística* se ha hecho indispensable en las investigaciones médicas y de salud pública, en la investigación de mercados y en el control de calidad, en la educación, la contabilidad, la economía, en la predicción meteorológica o de las votaciones, y en muestreos, deportes, seguros, en juegos de apuestas y en cualquier tipo de investigación que pretenda ser científica. Hoy en día, la Estadística se ha convertido realmente en una pieza integrante de nuestra herencia intelectual.

Problemas

1. Este problema tiene relación con la tabla 1.1.
 - (a) ¿En qué año hubo la mayor diferencia entre el número medio de cursos finalizados por los estudiantes que comenzaron su escolarización antes o después?
 - (b) ¿Existieron más años en los que el promedio de cursos completados por el grupo de estudiantes que comenzaron más jóvenes fue mayor que el del grupo de estudiantes que comenzaron tarde o sucedió al contrario?
2. El siguiente gráfico muestra los consumos de leche en Estados Unidos desde 1909 hasta 2000. ¿Qué conclusión general se puede extraer?



3. Los siguientes datos muestran los porcentajes de fumadores adultos en Estados Unidos, clasificados por sexo y nivel educacional, entre los años 1999 y 2002.
- (a) ¿En qué grupos ha existido una reducción sistemática?
- (b) ¿Podría decirse que existe una tendencia general?

Consumo de cigarrillos en Estados Unidos (% de adultos fumadores)

	1999	2000	2001	2002
Total	25,8	24,9	24,9	26,0
Sexo				
Hombre	28,3	26,9	27,1	28,7
Mujer	23,4	23,1	23,0	23,4
Educación				
No graduado en secundaria	39,9	32,4	33,8	35,2
Graduado en secundaria	36,4	31,1	32,1	32,3
Con algunos cursos universitarios	32,5	27,7	26,7	29,0
Graduado universitario	18,2	13,9	13,8	14,5

4. Intentando determinar la eficacia de un medicamento nuevo, un investigador médico ha comenzado con el contraste del medicamento frente a un placebo. Para asegurarse de que los dos grupos de pacientes voluntarios –aquellos que reciben el medicamento y los que reciben el placebo– son lo más parecidos posible, el investigador ha decidido no basarse en el azar sino que, por el contrario, ha analizado detalladamente a los voluntarios y luego él mismo ha elegido los grupos. ¿Es aconsejable este procedimiento? ¿Por qué sí? o ¿por qué no?
5. Explique por qué es importante que un investigador que intenta estudiar la utilidad de un nuevo medicamento no conozca qué pacientes son tratados con el medicamento y cuáles están recibiendo el placebo.
6. Se va a celebrar una votación la semana próxima y se pretende predecir, mediante la selección de una muestra de votantes, si ganará el candidato republicano o el candidato demócrata. ¿Cuál de los siguientes métodos de selección permite obtener una muestra significativa?
- (a) Seleccionar a toda la gente en edad de votar que asiste a un partido de baloncesto universitario.
- (b) Seleccionar a todas las personas en edad de votar que salen de un restaurante de moda de la ciudad.
- (c) Obtener una copia del censo de votantes, elegir 100 nombres aleatoriamente y entrevistarlos.
- (d) Utilizar los resultados de un programa de televisión en el que se pide a los espectadores que llamen por teléfono y comuniquen su elección.
- (e) Elegir nombres de la guía telefónica y llamarles posteriormente.

14. Las cifras de mortalidad de Londres registraban 12 246 muertes en 1658. Suponiendo que una encuesta sobre las parroquias de Londres mostró que, grosso modo, un 2% de la población había fallecido en dicho año, utilice el método de Graunt para estimar la población de Londres en 1658.
15. Suponga que usted es un vendedor de planes de pensiones en 1662, año en el que se publicó el libro de Graunt. Explique cómo habría usado los datos sobre las edades en las que se producían los fallecimientos.
16. Si se basa en la tabla 1.2, ¿cuál de los cinco años de peste parece haber sido el más severo? Explique su razonamiento.
17. Basándose en la tabla de mortalidad de Graunt:
 - (a) ¿Qué proporción de bebés sobrevivió a la edad de 6 años?
 - (b) ¿Qué proporción de bebés sobrevivió a la edad de 46 años?
 - (c) ¿Qué proporción murió entre las edades de 6 y 36 años?
18. ¿Por qué piensa que el estudio de la Estadística es importante en sus áreas de interés?
¿Cómo cree que puede utilizarla en su trabajo futuro?

La cambiante definición de la Estadística

La Estadística tiene el objetivo de realizar una representación fiable de un Estado en una época determinada. (Quetelet, 1849)

La Estadística es la única herramienta mediante la cual se puede conseguir una apertura en la formidable espesura de dificultades que entorpece el camino de aquellos que estudian la Ciencia del hombre. (Galton, 1889)

La Estadística puede considerarse (i) como el estudio de las poblaciones, (ii) como el estudio de las variaciones y (iii) como el estudio de los métodos de reducción de datos. (Fisher, 1925)

La Estadística es la disciplina científica relativa a la recopilación, el análisis y la interpretación de datos obtenidos mediante la observación o la experimentación. Tiene una estructura coherente basada en la *Teoría de la Probabilidad* e incluye muchos procedimientos diferentes que contribuyen a la investigación y el desarrollo en todas las ramas de la Ciencia y la Tecnología. (E. Pearson, 1936)

La Estadística es el nombre de la ciencia que trata de llevar a cabo inferencias bajo situaciones de incertidumbre; para ello, usa los números para averiguar cuestiones relativas a la naturaleza y la experiencia. (Weaver, 1952)

La Estadística se caracteriza en el siglo XX como una herramienta matemática para analizar datos experimentales u observados. (Ross, 2005)

Términos clave

Estadística: Arte de aprender de los datos.

Estadística descriptiva: Parte de la Estadística que trata con la descripción y la clasificación de los datos.