

Joaquín Tomás-Sábado

# Fundamentos de bioestadística y análisis de datos para enfermería

Escola Universitària d'Infermeria i de Fisioteràpia «Gimbernat»

Escola d'Infermeria

Universitat Autònoma de Barcelona  
Servei de Publicacions  
Bellaterra, 2009

Primera edición: diciembre de 2009

Edición e impresión:  
Servei de Publicacions  
Universitat Autònoma de Barcelona  
Edifici A. 08193 Bellaterra (Barcelona). Spain  
sp@uab.cat  
<http://publicacions.uab.cat/>

Impreso en España. Printed in Spain

Depósito legal: B. 47.850-2009  
ISBN 978-84-490-2616-4

# Índice

PRESENTACIÓN .....	9
TEMA 1. INTRODUCCIÓN CONCEPTUAL E HISTÓRICA .....	13
1.1. Estadística. Concepto y definición .....	13
1.2. Concepto de <i>bioestadística</i> .....	14
1.3. Áreas de la estadística .....	14
1.4. Estadística descriptiva (deducción) .....	14
1.5. Estadística inferencial o analítica (inducción) .....	14
1.6. Estadística. Desarrollo histórico .....	14
1.6.1. John Graunt (1620-1674) .....	15
1.6.2. Johann Carl Friedrich Gauss (1777-1855) .....	16
1.6.3. Jacques Quetelet (1796-1874) .....	16
1.6.4. Francis Galton (1822-1911) .....	17
1.6.5. Florence Nightingale (1820-1910) .....	18
1.6.6. William Sealey Gosset (1876-1937) .....	19
1.6.7. Ronald Fisher (1890-1962) .....	19
TEMA 2. POBLACIÓN Y MUESTRA .....	21
2.1. Conceptos preliminares .....	21
2.2. Individuo .....	21
2.3. Universo .....	21
2.4. Población .....	21
2.5. Muestra .....	22
2.6. Muestreo .....	22
2.7. Muestreo aleatorio .....	23
2.7.1. Muestreo aleatorio simple .....	23
2.7.2. Muestreo aleatorio sistemático .....	23
2.7.3. Muestreo aleatorio estratificado .....	23
2.7.4. Muestreo aleatorio por conglomerados .....	24
2.8. Muestreo no aleatorio .....	24
2.8.1. Muestreo no aleatorio accidental o incidental .....	24
2.8.2. Muestreo no aleatorio intencionado .....	24
2.8.3. Muestreo no aleatorio por cuotas .....	24
TEMA 3. VARIABLES ESTADÍSTICAS Y ESCALAS DE MEDIDA .....	25
3.1. Variables .....	25
3.2. Clasificación de las variables .....	25
3.3. Variables cuantitativas .....	25
3.3.1. Variables cuantitativas continuas .....	25
3.3.2. Variables cuantitativas discretas .....	26

3.4.	Variables cualitativas .....	26
3.4.1.	Variables cualitativas dicotómicas.....	26
3.4.2.	Variables cualitativas politómicas .....	26
3.5.	Variables ordinales o cuasicuantitativas .....	27
3.6.	Variables independientes .....	27
3.7.	Variables dependientes.....	27
3.8.	Variables de confusión.....	27
3.9.	Variables aleatorias y variables controladas .....	27
3.10.	Escalas de medida.....	28
3.11.	Escala nominal.....	28
3.12.	Escala ordinal.....	28
3.13.	Escala de intervalo.....	28
3.14.	Escala de proporción o razón.....	29
TEMA 4.	CLASIFICACIÓN DE LOS DATOS: TABULACIÓN Y REPRESENTACIÓN GRÁFICA	31
4.1.	Tabulación.....	31
4.2.	Frecuencias y tablas de frecuencias .....	32
4.3.	Tipos de frecuencias .....	32
4.3.1.	Frecuencias absolutas ( $f_i$ ) .....	32
4.3.2.	Frecuencias relativas ( $h_i$ ) .....	32
4.3.3.	Porcentajes ( $p_i$ ) .....	32
4.3.4.	Frecuencias absolutas acumuladas ( $F_i$ ) .....	33
4.3.5.	Frecuencias relativas acumuladas ( $H_i$ ) .....	33
4.3.6.	Porcentajes acumulados ( $P_i$ ).....	33
4.4.	Representaciones gráficas.....	34
4.5.	Gráficos: ventajas e inconvenientes.....	34
4.6.	Construcción de gráficos .....	34
4.7.	Gráficos para variables cualitativas .....	35
4.7.1.	Diagrama de barras.....	35
4.7.2.	Ciclogramas o sectores circulares .....	37
4.7.3.	Otros gráficos para variables cualitativas.....	37
4.8.	Gráficos para variables cuantitativas continuas.....	38
4.8.1.	Histogramas .....	38
4.8.2.	Gráficos de caja ( <i>boxplot</i> ) .....	38
4.8.3.	Polígonos de frecuencias .....	39
4.8.4.	Curvas de frecuencias acumuladas.....	40
TEMA 5.	MEDIDAS DE TENDENCIA CENTRAL .....	41
5.1.	La media aritmética .....	41
5.2.	Propiedades de la media aritmética .....	41
5.3.	Cálculo de la media aritmética .....	42
5.4.	Cálculo de la media aritmética con los datos agrupados por frecuencias.....	42
5.5.	Media ponderada .....	43
5.6.	Otras medias .....	44
5.6.1.	Media geométrica .....	44
5.6.2.	Media cuadrática .....	45
5.6.3.	Media armónica .....	46
5.7.	La mediana.....	47
5.8.	Cálculo de la mediana.....	47
5.9.	Cálculo de la mediana con los datos agrupados por frecuencias.....	47
5.10.	La moda .....	48

TEMA 6. MEDIDAS DE POSICIÓN: CUANTILES .....	51
6.1. Cuartiles .....	51
6.2. Deciles .....	51
6.3. Centiles o percentiles .....	52
6.4. Equivalencias .....	52
6.5. Cálculo de los cuartiles .....	52
6.6. Cálculo de los deciles .....	53
6.7. Cálculo de los centiles .....	53
6.8. Cálculo de los cuantiles a partir de los porcentajes acumulados .....	54
TEMA 7. MEDIDAS DE DISPERSIÓN .....	55
7.1. Amplitud o rango de la distribución .....	55
7.2. Amplitud intercuartil .....	55
7.3. Desviación cuartil .....	56
7.4. Desviación media.....	56
7.5. La varianza.....	57
7.6. Propiedades de la varianza.....	57
7.7. Cálculo de la varianza.....	58
7.8. Cálculo de la varianza con los datos agrupados por frecuencias .....	59
7.9. Desviación típica o desviación estándar .....	60
7.10. Propiedades de la desviación típica .....	60
7.11. Coeficiente de variación .....	60
TEMA 8. MEDIDAS DE FORMA .....	61
8.1. Coeficiente de asimetría .....	61
8.2. Medidas de tendencia central y asimetría .....	61
8.3. Curtosis o apuntamiento .....	62
8.4. Coeficiente de curtosis .....	62
TEMA 9. CÁLCULO DE PROBABILIDADES .....	63
9.1. Experimentos y sucesos aleatorios .....	63
9.2. Concepto de <i>probabilidad</i> .....	63
9.3. Conceptos básicos .....	63
9.3.1. Espacio muestral .....	63
9.3.2. Suceso .....	64
9.4. Propiedades de la probabilidad .....	64
9.5. Teoremas básicos de la probabilidad .....	66
9.5.1. Teorema de la probabilidad total o de la suma .....	66
9.5.2. Teorema de la probabilidad compuesta o del producto .....	67
TEMA 10. DISTRIBUCIÓN DE LA PROBABILIDAD PARA VARIABLES CONTINUAS:	
LA DISTRIBUCIÓN NORMAL .....	69
10.1. La ley normal .....	69
10.2. La curva normal tipificada .....	70
10.3. Aplicaciones de la curva normal .....	70
TEMA 11. DISTRIBUCIÓN DE LA PROBABILIDAD PARA VARIABLES DISCRETAS:	
LA DISTRIBUCIÓN BINOMIAL .....	77
11.1. La ley binomial .....	77
11.2. Definición de <i>distribución binomial</i> .....	77
11.3. Aproximación de la distribución binomial a la distribución normal .....	78

TEMA 12. INTRODUCCIÓN A LA ESTADÍSTICA INFERENCIAL .....	81
12.1. La inferencia estadística .....	81
12.2. Estimación puntual e intervalos de confianza .....	81
12.3. Pruebas estadísticas inferenciales .....	81
12.4. Contrastes de hipótesis y significación.....	82
12.5. Hipótesis nula .....	82
12.6. Hipótesis alternativa .....	82
12.7. Nivel de significación .....	82
12.8. El $p$ -valor .....	83
12.9. Tipos de errores .....	83
12.10. Región crítica y región de aceptación.....	84
12.11. Contrastes de hipótesis bilaterales .....	84
12.12. Contrastes de hipótesis unilaterales .....	84
12.13. Intervalo de confianza para una media .....	85
12.14. Uso de intervalos de confianza para verificar hipótesis .....	86
12.15. Pruebas de contraste de hipótesis paramétricas y no paramétricas.....	86
TEMA 13. PRUEBAS PARAMÉTRICAS .....	89
13.1. Prueba $t$ de Student .....	89
13.1.1. Prueba $t$ de Student para una muestra.....	89
13.1.2. Prueba $t$ de Student para dos muestras relacionadas (datos apareados) .....	90
13.1.3. Prueba $t$ de Student para muestras independientes .....	91
13.2. Análisis de la varianza de un factor (ANOVA) .....	92
13.3. ANOVA de medidas repetidas .....	93
13.4. Coeficiente de correlación de Pearson.....	94
TEMA 14. PRUEBAS NO PARAMÉTRICAS .....	97
14.1. Prueba de Kolmogorov-Smirnov para comprobar la normalidad.....	97
14.2. Prueba $\chi^2$ cuadrado de Pearson ( $\chi^2$ ) .....	97
14.3. El coeficiente Phi .....	99
14.4. Prueba $V$ de Cramer.....	99
14.5. Prueba para dos muestras independientes: la prueba $U$ de Mann-Whitney.	99
14.6. Prueba para dos muestras relacionadas: la prueba de Wilcoxon .....	100
14.7. Prueba para tres o más muestras independientes: el ANOVA de Kruskal-Wallis .....	101
14.8. Prueba para tres o más muestras relacionadas: la prueba de Friedman .....	102
14.9. Coeficiente de concordancia $W$ de Kendall.....	103
14.10. Coeficiente de correlación de Spearman .....	104
PREGUNTAS VERDADERO-FALSO .....	105
PREGUNTAS DE OPCIÓN MÚLTIPLE .....	121
BIBLIOGRAFÍA .....	145

## Presentación

La presentación de este libro supone para mí una gran satisfacción, principalmente porque me ofrece la posibilidad de compartir todo lo que ha rodeado su origen. Es por esto que quiero dirigirme especialmente a los estudiantes de la Escuela Universitaria de Enfermería Gimbernat, para hacerles partícipes de su *backstage*.

El libro que tenéis en las manos nace de la inquietud del profesor Tomás-Sábado por ir más allá de la docencia cotidiana. En él encontramos la suma de varios factores que hacen de este manual una herramienta imprescindible no sólo para nuestros estudiantes, sino para todo aquel que quiera introducirse de una forma sencilla en esta área de conocimiento. Estos factores no son otros que los *años de experiencia docente* enfocada hacia las necesidades reales de los estudiantes en una materia ciertamente complicada en los inicios de sus estudios, la motivación personal por la mejora continua y por seguir ofreciendo refuerzo en el proceso de aprendizaje y, muy especialmente, la *ilusión* por el trabajo bien hecho.

La aparición de un libro es siempre una buena noticia para todos, pero, en este caso, especialmente para los estudiantes, puesto que, sin duda, este libro será una guía de referencia y de apoyo metodológico en su proceso de aprendizaje.

Quisiera hacer un último agradecimiento, muy sinceramente, al profesor Tomás-Sábado, por su implicación en la mejora continua de nuestra institución

MONTSERRAT ANTONÍN  
Coordinadora de Titulación  
Escuela Universitaria de Enfermería Gimbernat

*Perseveranter omnia  
consequitur*

## Tema 1. Introducción conceptual e histórica

### 1.1. Estadística. Concepto y definición

La estadística puede definirse como la disciplina científica dedicada al tratamiento de la información que contiene series de datos que proceden de la observación de fenómenos colectivos (demográficos, económicos, sanitarios, etc.), en los que intervienen factores de variación que hacen necesario formular modelos probabilísticos para poder llegar a conclusiones o predicciones bajo un determinado nivel de probabilidad. En general, los procedimientos estadísticos se aplican a la recopilación, organización, presentación, análisis e interpretación de datos numéricos con el fin de realizar una toma de decisión más efectiva.

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

En definitiva, la estadística es un método de análisis cuantitativo de los colectivos que permite interpretar información cuya propiedad fundamental es la variabilidad de los datos. Así, la estadística facilita el estudio de una característica del colectivo que puede expresarse numéricamente, bien porque es medible por naturaleza, bien porque de alguna manera puede expresarse numéricamente. Para su aplicación, la estadística se basa en la teoría de probabilidades y en el cálculo infinitesimal.

El profesor y economista alemán Godofredo Achenwall (1719-1772), docente de la Universidad de Göttingen, introdujo en 1749 el término *estadística* (*statistik*) para denominar a lo que hasta entonces se conocía como *aritmética política*. La palabra deriva de *staat*, que significa «estado» o «gobierno», y Achenwall la aplicó al «conocimiento profundo de la situación respectiva y comparativa de cada estado», de modo que estructuró los métodos estadísticos orientados a investigar, medir y comparar las riquezas de las naciones. Achenwall es autor de obras sobre la historia de los estados europeos, basadas en derecho y economía política, tales como *Elementos de estadística de los principales estados de Europa* y *Principios de economía política*.

Gracias a las investigaciones realizadas en Alemania, John Sinclair (1754-1835) fue el primero en introducir el término *estadística* en Inglaterra, con su trabajo *Statistical Account of Scotland*. Sin embargo, mientras en Alemania la estadística se venía utilizando básicamente como instrumento para medir la fortaleza del Estado, Sinclair proponía su utilización como generadora de información interna para encontrar carencias y proponer mejoras para el país.

A comienzos del siglo XIX, el término *estadística* adoptó un significado más general, orientado a la obtención, clasificación y tratamiento de cualquier conjunto de datos colectivos cuantitativos.

## 1.2. Concepto de *bioestadística*

Se denomina *bioestadística* la aplicación particular de la estadística a las ciencias biológicas y de la salud. Es decir, la obtención y el análisis de datos biológicos o de salud mediante la utilización de métodos estadísticos. Por ejemplo, la bioestadística se puede usar para ayudar a comprender las posibles causas de un cáncer o con qué frecuencia se presenta una enfermedad en un determinado grupo de personas. En algunos ámbitos también se denomina *biometría*

## 1.3. Áreas de la estadística

La estadística ofrece métodos para analizar series de datos de modo descriptivo o inferencial. Según esto, podemos distinguir entre *estadística descriptiva* y *estadística inferencial*.

## 1.4. Estadística descriptiva (deducción)

La estadística descriptiva pretende describir, analizar y representar las características que existen en un conjunto de datos, obtenidos a partir de una población o de una muestra. Comprende la tabulación, la presentación y la descripción de los datos empíricos, a fin de hacerlos más manejables y comprenderlos e interpretarlos mejor.

Cuando un valor se ha obtenido a partir de una muestra, hablamos de *estadístico*, mientras que un *parámetro* es un valor obtenido a partir de una población. Así, la media de estancias hospitalarias en todos los hospitales de España es un parámetro, mientras que la media de estancias hospitalarias en una muestra de hospitales españoles es un estadístico.

Generalmente, los estadísticos se simbolizan con letras latinas ( $\bar{x}$ ,  $S$ ) y los parámetros, con letras griegas ( $\mu$ ,  $\sigma$ ).

## 1.5. Estadística inferencial o analítica (inducción)

La estadística inferencial o analítica es la que, apoyándose en el cálculo de probabilidades y a partir de los datos obtenidos de una muestra, trata de sacar conclusiones acerca de las características de una población.

## 1.6. Estadística. Desarrollo histórico

Puede afirmarse que el uso de los procedimientos estadísticos es tan antiguo como la humanidad. En efecto, desde que se constituyeron las primeras sociedades humanas, fue necesario desarrollar algún tipo de sistema, como representaciones gráficas o símbolos de distinta índole en pieles, rocas, trozos de madera o paredes de cuevas, para contar el número de personas, animales, alimentos o enseres.

A partir del análisis del material arqueológico se sabe que 3.000 años aC se usaban en Babilonia tablas de arcilla donde se recopilaban datos sobre la producción agrícola

y resultados de actividades comerciales. Por la misma época, los antiguos egipcios, además de recopilar datos relativos a la población y la riqueza del país, desarrollaron sofisticados sistemas de cálculo y medida, obligados por la necesidad de redistribución de las tierras después de las periódicas inundaciones provocadas por el Nilo.

Aunque se conoce que los chinos y los griegos realizaban, hace más de cuarenta siglos, censos de poblaciones con fines tributarios, sociales y militares, fue durante el Imperio romano cuando se utilizaron con mayor rigor todos los recursos de la estadística para recoger una gran cantidad de datos sobre la población, la superficie y la renta de todos los territorios bajo su control. Los romanos efectuaban cada cinco años un censo de población que incluía nacimientos, defunciones y matrimonios, además del recuento de ganado y de las riquezas de las tierras conquistadas. Se da la circunstancia de que el nacimiento de Cristo se produjo precisamente en uno de estos empadronamientos periódicos que se realizaban en todo el territorio bajo la autoridad imperial.

Durante la edad media, la estadística, al igual que la mayoría de las ciencias, sufrió una considerable regresión, tanto en su uso como en su desarrollo. No obstante, se tienen noticias de algunos censos minuciosos realizados en Europa, como los ordenados por Pipino el Breve y Carlomagno en los años 758 y 762, respectivamente. También, en Inglaterra, el rey Guillermo I encargó en 1086 la realización de un censo sobre datos de propiedad, extensión y valor de todas las tierras de su reino, que fueron recopilados en el *Domesday book* o *Libro del gran catastro*, considerado como el primer compendio estadístico de Inglaterra.

La revolución, a nivel de pensamiento, que supuso el Renacimiento, permitió sentar las bases para el advenimiento de la ciencia moderna. Hombres como Leonardo da Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Francis Bacon y René Descartes, hicieron grandes aportaciones al método científico, de manera que, cuando se crearon los estados nacionales y surgió con fuerza el comercio internacional, existía ya un método aplicable a los datos económicos. Paralelamente, la estadística empezó a adquirir una importancia creciente en la tarea de recopilar de manera sistemática datos demográficos, sociales y económicos para los nuevos estados europeos.

Desde entonces hasta nuestros días, han sido muchos los científicos e investigadores que, desde distintas disciplinas, han aportado su contribución al desarrollo de la estadística. Evidentemente, no es posible hacer aquí una relación exhaustiva de sus nombres y su obra, de manera que expondremos únicamente y a grandes rasgos los aspectos relevantes de aquellos personajes cuyas aportaciones son consideradas más significativas y más influyentes en los métodos estadísticos contemporáneos.

### **1.6.1. John Graunt (1620-1674)**

Demógrafo y economista inglés. Demostró la uniformidad y, por tanto, las posibilidades de pronóstico de los fenómenos biológicos, sentando las bases de la estadística científica. A partir del análisis del censo de fallecimientos de Londres observó que había un mayor número de nacimientos masculinos, una alta tasa de mortalidad en los primeros años de vida, una mayor tasa de mortalidad en los ambientes urbanos con respecto a los rurales y una influencia de la estación del año en el número de fallecimientos. Graunt estableció una clasificación de las causas de muerte de acuerdo con los conocimientos de la época. Este primer estudio epidemiológico, publicado bajo el título de *London bills of mortality*, estimó una mortalidad en niños nacidos vivos, menores de seis años, del 36 %. Asimismo, elaboró la primera tabla de vida que se conoce y que especifica las probabilidades de morir y vivir durante el curso de la vida. Graunt es considerado unánimemente el primer demógrafo, el fundador de la bioestadística y el precursor de la epidemiología.

### **1.6.2. Johann Carl Friedrich Gauss (1777-1855)**

Matemático, astrónomo y físico alemán que contribuyó significativamente en muchos campos, incluida la teoría de los números, el análisis matemático, la estadística, la geometría diferencial, la geodesia, el magnetismo y la óptica. Considerado «el príncipe de las matemáticas» y «el matemático más grande desde la antigüedad», Gauss ha tenido una influencia notable en muchos campos de la ciencia y es considerado uno de los matemáticos que más influencia ha tenido alrededor de la historia.

Gauss fue un prodigio. Existen muchas anécdotas acerca de su asombrosa precocidad siendo apenas un niño, e hizo sus primeros grandes descubrimientos cuando era un adolescente. Completó su *Disquisitiones Arithmeticae* a los veintiún años (1798), aunque esta obra no fue publicada hasta 1801. Es un trabajo que fue fundamental para que la teoría de los números se consolidara y que ha moldeado esta área hasta los días presentes. Las principales aportaciones de Gauss a la estadística fueron en la teoría de la estimación: el método de los mínimos cuadrados y, como consecuencia, el llamado *modelo lineal de Gauss*. Estudió también la teoría de los errores y dedujo la curva normal de la probabilidad, conocida como *curva de Gauss*, que todavía se usa en los cálculos estadísticos. Fue bien entrado el siglo XIX cuando la ley normal, así bautizada por Galton, obtuvo una aceptación universal, siendo reconocida como la ley de los errores por excelencia.

Gauss aproximó el problema de la estimación estadística con espíritu empírico, recalcando la cuestión de la estimación no sólo de las probabilidades, sino también de otros parámetros cuantitativos. Descubrió que, para este propósito, el método apropiado era el de la máxima verosimilitud, aunque trató de justificar el método por el principio de la probabilidad. Además, perfeccionó el ajuste sistemático de las fórmulas de regresión simple y regresión múltiple por el método de los mínimos cuadrados.

### **1.6.3. Jacques Quetelet (1796-1874)**

Estadístico y astrónomo belga. Reconocido como uno de los padres de la estadística moderna y fundador de la ciencia social cuantitativa moderna, fue el primero en aplicar la estadística y la teoría de la probabilidad a los fenómenos sociales.

Según Quetelet, la teoría estadística y la investigación pueden utilizarse para determinar si existe una regularidad en las acciones humanas, lo que permitiría formular leyes sociales tan fiables como las que rigen los movimientos de los cuerpos celestes. A partir de esta idea, desarrolló el concepto de *hombre medio*, como modelo social que permite expresar las diferencias entre los individuos en términos de desviación de la norma. Este modelo llevó a Quetelet a formular su «teoría de la oscilación», según la cual, conforme aumentan los contactos sociales y las relaciones interraciales, las diferencias entre los hombres disminuyen en intensidad gracias a un proceso social y cultural de oscilación que dará lugar a un equilibrio cada vez mayor y, finalmente, al equilibrio y a la paz universales.

También alcanzaron gran difusión los trabajos de Quetelet donde aplicaba la estadística a la criminología. En su obra *Física social* defendió la idea de que el delito es un fenómeno social, producido por hechos sociales que son detectables y determinables estadísticamente, de manera que el índice de delitos se puede calcular con anticipación. En este sentido, formuló sus famosas «leyes térmicas», resumidas en tres puntos: 1) en invierno se cometen mayor número de delitos contra la propiedad que en verano, debido a las mayores dificultades que se presentan para vivir; 2) los delitos contra las personas son más abundantes en verano, ya que el calor excita las pasiones humanas, los días son

más largos y la temperatura alta hace a las personas más irritables; y 3) los delitos sexuales se presentan con mayor frecuencia en primavera, época en la que se produce el apareamiento en los animales, cuya fuerza instintiva también afecta al hombre.

#### 1.6.4. Francis Galton (1822-1911)

Puede considerarse que las teorías evolucionistas son las que proporcionaron el impulso definitivo a la ciencia estadística en general y a la bioestadística en particular. La evolución de la inteligencia humana presentaba un gran interés para Sir Francis Galton, quien, influido por el énfasis en las diferencias individuales de la teoría evolucionista de su primo Charles Darwin, se dedicó plenamente a su estudio y a su medida. Galton publicó en 1869 *Hereditary genius: An inquiry into its laws and consequences*, donde llevó a cabo un estudio de la genealogía de las familias de científicos famosos (incluyendo la suya). En este trabajo, Galton planteaba varias cuestiones que aún hoy continúan siendo debatidas. Una de ellas es la referente a la naturaleza de la inteligencia, que Galton identifica con una habilidad natural y considera de naturaleza innata. El conjunto de la obra asume abiertamente la visión científica popular acorde con la filosofía y la política de la Inglaterra de finales del siglo XIX, donde puede ser aplicada la teoría evolucionista, según la cual los hombres blancos instruidos de la clase media inglesa estaban en lo más alto del árbol evolutivo humano. La teoría jerárquica daba un estatus evolutivo inferior a las razas de las colonias, a los irlandeses y a la clase obrera inglesa, hecho que servía como justificación para el mantenimiento de la posición social de la clase dominante.

La pasión por la medida que Galton desarrolló a lo largo de toda su vida resulta paradigmática. Es ampliamente conocido su empeño en contar y medir cualquier cosa que pareciera presentar algún aspecto de regularidad. Entre otras muchas cosas, midió el grado de aburrimiento en las conferencias científicas, trató de determinar el país donde se encontraban las mujeres más hermosas, fue el primero en proponer las huellas dactilares como procedimiento de identificación, quiso medir y predecir el tiempo, inventando el mapa del tiempo y términos como *altas y bajas presiones*, etc. Galton pensaba que las diferencias individuales entre las personas debían ser medidas y clasificadas, y se dedicó con pasión a esta actividad.

En 1883, Galton estableció su laboratorio antropométrico en la South Kensington Exhibition y posteriormente en el South Kensington Museum. Previo pago de tres peniques, las personas asistentes podían obtener una evaluación de una serie de medidas físicas y de sus facultades sensoriales, perceptivas y motoras. En su laboratorio antropométrico evaluó a más de 17.000 personas a lo largo de ocho años de trabajo, consiguiendo acumular una ingente cantidad de datos. Para proceder a su análisis, Galton se dedicó activamente al desarrollo de procedimientos estadísticos.

Galton introdujo el análisis estadístico como aspecto fundamental de la toma de decisiones en las cuestiones planteadas por la evaluación psicológica. Entre otras contribuciones, propuso la utilización de la curva normal, que ya Quetelet había aplicado a datos biológicos y sociales, como modelo de la distribución de las puntuaciones de los tests. También introdujo el concepto de *mediana* como medida de tendencia central, considerándola más fiable que la media, al no estar afectada por las puntuaciones extremas. En colaboración con su colega y amigo Karl Pearson (1857-1936), desarrolló el concepto de *regresión a la media* y el coeficiente de correlación producto-momento para el análisis de estos datos. Posteriormente, Pearson continuó el desarrollo del concepto matemático de correlación, agregando los coeficientes parcial y múltiple de *correlación*, la prueba de bondad de ajuste y la prueba ji cuadrado al repertorio de técnicas disponibles.

Otro de sus colaboradores, Charles Spearman (1863-1945), un antiguo oficial del ejército reconvertido en estadístico, desarrolló después procedimientos para el análisis de matrices de correlación más complejas, estableciendo los fundamentos del análisis factorial. También fue Spearman quien sentó las bases de la teoría de la puntuación verdadera, el error de medida y de la fiabilidad, asociados al proceso de medición.

### ***1.6.5. Florence Nightingale (1820-1910)***

En un trabajo de estadística dirigido a estudiantes de enfermería, no podía faltar una referencia a la figura más emblemática de la historia de la profesión. No obstante, puesto que la vida y la obra de Florence Nightingale es, en general, estudiada y analizada en otras materias del plan de estudios, nos limitaremos aquí a reseñar su trabajo en un campo quizás menos conocido por los estudiantes: su notable aportación como experta estadística y pionera de la epidemiología.

Desde muy joven, Florence Nightingale se mostró interesada por el estudio de las matemáticas, especialmente por la estadística, un campo en el que su padre, que trabajaba en el entonces incipiente campo de la epidemiología, era un experto.

Nightingale hizo un amplio uso de los análisis estadísticos en la compilación, el análisis y la presentación de estadísticas de la atención médica y la salud pública. También fue pionera en el uso de los gráficos en la presentación de los informes estadísticos; en este sentido, ideó su diagrama de área polar, o coxcomb, para proporcionar una información gráfica de las cifras de mortalidad durante la guerra de Crimea. Desarrolló técnicas innovadoras de análisis estadístico (como el «ploteo» de incidencias de muerte prevenible entre los militares durante la guerra), con lo cual mostró, finalmente, cómo un fenómeno social podía ser medido objetivamente y analizado matemáticamente. También desarrolló una fórmula modelo de estadística hospitalaria para que los hospitales recolectaran y generaran datos y estadísticas consistentes.

Su interés estuvo siempre dirigido al uso de los métodos estadísticos al servicio de las condiciones de vida y la salud de los ciudadanos. Denunció las condiciones de salubridad de los hospitales británicos, aportando datos que demostraban que entre los enfermos ingresados en hospitales había una tasa de mortalidad del 90 %, mientras que esta tasa descendía al 60 % cuando el enfermo permanecía en su casa. Usando sus estadísticas, ilustró la necesidad de una reforma sanitaria en todos los hospitales militares, consiguiendo que se crease la Royal Commission on the Health of the Army.

En las etapas posteriores de su vida hizo un amplio estudio estadístico en el que aplicó los métodos de la sanidad pública a las zonas rurales de la India y fue la figura principal de la introducción de mejoras en la atención médica y el servicio de salud pública en este país.

Florence Nightingale es todavía una referencia importante en la estadística actual. Considerada como una innovadora en la aplicación del método epidemiológico para su uso en las estadísticas de salud pública, estableció también los fundamentos del análisis de datos en relación con la «gestión de la calidad» y las «auditorías de los cuidados de la salud». Su ingente labor y sus aportaciones al desarrollo de las técnicas estadísticas en el ámbito de la salud hicieron que en 1858 fuera invitada a ingresar en la Royal Statistical Society, siendo la primera mujer que formó parte de esta prestigiosa asociación. En 1874 fue también nombrada miembro honoraria de la American Statistical Society.

### **1.6.6. William Sealey Gosset (1876-1937)**

Estadístico y químico inglés. Trabajó gran parte de su vida como químico de la fábrica de cerveza Guinness, llegando a ser director de la factoría de la marca en Londres. Más conocido por *Student*, seudónimo con el que firmaba sus trabajos, al parecer debido a la prohibición expresa de la empresa que sus empleados publicaran cualquier tipo de información.

Gossett es uno de los estadísticos con más influencia en los métodos actuales. Especialmente conocidos son sus trabajos sobre la forma de la distribución  $t$  y el  $t$ -test (conocido como *prueba t de Student*), para muestras pequeñas, una de las pruebas de contraste de hipótesis más utilizadas en la práctica, que desarrolló mientras trabajaba en la aplicación de técnicas de control de calidad en la fabricación de cerveza.

Curiosamente, mantuvo una cordial relación de amistad y colaboración tanto con Pearson como con Fisher, a pesar de que estos dos últimos se profesaban mutuamente una notable aversión.

### **1.6.7. Ronald Fisher (1890-1962)**

Estadístico, biólogo evolutivo y genetista inglés. A él se deben los fundamentos de la estadística actual y muchos de los métodos de inferencia en las pruebas de contraste de hipótesis. Se considera que Student y Fisher iniciaron una nueva era en el estudio de las distribuciones muestrales.

Fisher sentó las bases del diseño de experimentos, con lo que la estadística se incorporó plenamente a la investigación científica. Introdujo el análisis de la varianza, de la covarianza y otras pruebas no paramétricas. Asimismo, fue uno de los principales impulsores de la genética de poblaciones, armonizando la metodología biométrica con la genética mendeliana.

## Tema 2. Población y muestra

### 2.1. Conceptos preliminares

La estadística obtiene y estudia datos sobre diferentes *individuos*, que no tienen que ser necesariamente personas, hombres o mujeres. El conjunto de todos los individuos posibles constituye el *universo*.

En general, no interesan los datos de todos los individuos, de todo un universo, sino que se estudian *poblaciones*. Aún así, obtener y analizar los datos de toda una población suele ser imposible, por lo que, en la práctica, suele seleccionarse una *muestra* de individuos de la población; únicamente en estos individuos estudiaremos los datos que nos interesan.

### 2.2. Individuo

Es cada elemento que lleva asociada una medida, un número de orden o una característica predeterminada.

### 2.3. Universo

Es el conjunto, finito o infinito, de todos los posibles individuos que cumplen ciertas propiedades.

### 2.4. Población

Es el conjunto de todos los individuos que cumplen ciertas propiedades y de quienes deseamos estudiar ciertos datos. Podemos entender que una población abarca todo el conjunto de elementos de los cuales podemos obtener información, entendiendo que todos ellos han de poder ser identificados. La población deberá ser definida sobre la base de las características que la delimitan, que la identifican y que permiten la posterior selección de unos elementos que se puedan entender como representativos (muestra).

Hay que distinguir entre *población diana* o *población objetivo* (aquella población a la que se desea extrapolar los resultados del estudio) y *población accesible* (aquella población cuyos individuos son directamente accesibles al investigador para seleccionar la muestra).

Asimismo, una población puede ser *finita* (por ejemplo, todos los enfermos de los hospitales de Cataluña o los estudiantes de enfermería de España) o *infinita* (los posibles resultados de sucesivas tiradas de una moneda o el conjunto de los números pares).

## 2.5. Muestra

Una muestra es una porción de algo. Si deseamos preguntar a un conjunto de cinco mil personas su opinión sobre un determinado fenómeno, tenemos dos opciones: efectuar las preguntas persona por persona o efectuar las preguntas solamente a una muestra de estas personas, es decir, a un grupo de elementos representativos de ese conjunto.

Evidentemente, si se examina toda la población, mediante un *censo*, podemos conocer exactamente cuál es la distribución de la variable o las variables de interés en esta población. Sin embargo, en la mayoría de las ocasiones los censos resultan inviables, caros y lentos, además de innecesarios. La alternativa al censo es estimar la distribución de la variable en una parte representativa de la población, es decir, en una muestra, lo que tiene la ventaja de ser más rápido y más barato, y si la muestra se ha elegido correctamente, permite obtener una información que aporta una estimación razonable de la situación de la variable en la población.

Así, la muestra es una parte o un subconjunto de la población en el que se observa el fenómeno a estudiar y de donde sacaremos unas conclusiones generalizables a toda la población. En general, se considera que una muestra es *grande* cuando el número de individuos seleccionados es igual o superior a 30, y una muestra es *pequeña* cuando los individuos son menos de 30.

Para que una muestra sea representativa de la población, deberá cumplir unas condiciones básicas:

- Han de delimitarse y definirse claramente las características que conforman la totalidad de la población.
- Ha de haber garantías de que cada elemento de la población tiene las mismas posibilidades de figurar en la muestra. En consecuencia, deberá utilizarse el procedimiento de muestreo adecuado.
- La muestra deberá tener el tamaño adecuado para poder extrapolar los resultados obtenidos al conjunto de la población con garantías de fiabilidad.

## 2.6. Muestreo

El muestreo es el método o procedimiento destinado a obtener una muestra adecuada que reproduzca las características básicas de la población. Existen diferentes criterios de clasificación de los procedimientos de muestreo, aunque, en general, pueden dividirse en dos grandes grupos: métodos aleatorios o probabilísticos y métodos no aleatorios o no probabilísticos.

### Tipos de muestreo

<i>Aleatorio (probabilístico o al azar)</i>	<i>No aleatorio (no probabilístico o no al azar)</i>
Simple	Accidental
Sistemático	Intencionado
Estratificado	Por cuotas
Proporcional	
No proporcional	
Por conglomerados	

## 2.7. Muestreo aleatorio

El muestreo aleatorio, probabilístico o muestreo al azar, parte de una igualdad absoluta de todos los elementos de la población para ser seleccionados, de manera que conocemos la probabilidad que tiene un elemento de ser incluido en la muestra. Así, los resultados obtenidos en esta muestra serán estadísticamente inferibles de toda la población, siempre que el tamaño sea el adecuado y en función del nivel de confianza que hayamos establecido previamente.

### 2.7.1. Muestreo aleatorio simple

Es el método conceptualmente más sencillo. Básicamente, consiste en elegir al azar de una lista todos los individuos que formarán parte de la muestra. Para llevar a cabo un muestreo aleatorio simple se siguen los siguientes pasos:

1. Se confecciona la lista de todos los elementos de la población, asignándoles números consecutivos de 1 a  $N$ , siendo  $N$  el total de elementos de la población. Esto supone que están claramente definidas cuáles son las unidades de base de la muestra, es decir, si se trata de individuos, grupos, familias, etc. Esta unidad de base ha de ser la misma en toda la muestra; no se puede seleccionar un individuo, luego una familia, etc.
2. Se decide el tamaño deseado de la muestra. Llamaremos  $n$  al número de elementos de la muestra. Para conocer el tamaño de la muestra existen tablas ya elaboradas y pruebas estadísticas que la determinan en función del tamaño de la población, del tipo de variables a estudiar y del nivel de confianza deseado.
3. Se extraen al azar los  $n$  elementos que sean necesarios para completar el tamaño de muestra requerido. Para esta extracción se pueden utilizar distintos métodos, como tablas de números aleatorios, sistemas de lotería o cualquier otro sistema que se ajuste a las leyes del azar.

En la práctica, a menos que se trate de poblaciones pequeñas o de estructura muy simple, es difícil aplicar este tipo de muestreo de forma eficaz.

### 2.7.2. Muestreo aleatorio sistemático

Es una variante sencilla del muestreo aleatorio simple, que parte también de la lista total de la población, pero, en lugar de extraer  $n$  números aleatorios, sólo se extrae uno y, a partir de ahí, se van seleccionando los demás elementos a intervalos fijos (5, 10, 15...), siendo el tamaño de este intervalo el resultado de dividir el tamaño de la población entre el tamaño de la muestra.

### 2.7.3. Muestreo aleatorio estratificado

En ocasiones, especialmente cuando las poblaciones son muy grandes, interesa dividir éstas en subpoblaciones o estratos en virtud de determinadas características (edad, sexo, estado civil...), de manera que al dividir a esta población no homogénea en dos o más estratos, conseguimos que se convierta en homogénea en cada estrato o subgrupo. A continuación se eligen en cada estrato las personas que formarán la muestra por el método aleatorio simple. Este sistema tiene la ventaja de que permite mayor profundización y

mayor precisión en el análisis de cada estrato. Además, permite que en cada estrato se puedan utilizar diferentes sistemas para la selección de los sujetos.

El muestreo estratificado puede ser *proporcional* o *no proporcional*. En el muestreo *proporcional*, el tamaño de la muestra en cada estrato es proporcional al número de individuos que existen en los estratos con relación al total de la población. En el muestreo estratificado *no proporcional*, a cada estrato le corresponde igual número de elementos muestrales.

#### ***2.7.4. Muestreo aleatorio por conglomerados***

En los métodos anteriores, se seleccionan directamente los elementos de la población, es decir, las unidades muestrales son los elementos de la población (individuos). En el muestreo por conglomerados, la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos *conglomerado* (bloques de viviendas, municipios, hospitales...). Este procedimiento consiste en seleccionar aleatoriamente los conglomerados necesarios para alcanzar el tamaño de la muestra, que quedará compuesta por todos los elementos pertenecientes a los conglomerados elegidos.

### **2.8. Muestreo no aleatorio**

En el muestreo no aleatorio o no probabilístico los sujetos se seleccionan siguiendo determinados criterios, de manera que no todos los elementos de la población tienen la misma probabilidad de figurar en la muestra. Este tipo de muestreo sólo es justificable en determinados casos, por ejemplo, en estudios exploratorios, donde un muestreo aleatorio puede resultar difícil o excesivamente costoso.

#### ***2.8.1. Muestreo no aleatorio accidental o incidental***

Se basa en el empleo para la muestra de aquellos sujetos a los que se tiene fácil acceso, con los que puede contarse más fácilmente o con mayor comodidad. Un caso particular que puede incluirse en esta categoría es el de los sujetos voluntarios.

#### ***2.8.2. Muestreo no aleatorio intencionado***

Se caracteriza por la inclusión deliberada en la muestra de aquellos elementos cuyas características son similares a las de la población elegida, en un esfuerzo por obtener muestras supuestamente representativas. Es decir, el investigador selecciona los elementos que a su juicio son representativos, lo que exige un conocimiento previo de la población que se investiga. Este tipo de muestreo es frecuentemente utilizado en estudios cualitativos.

#### ***2.8.3. Muestreo no aleatorio por cuotas***

Es un tipo de muestreo semejante al muestreo aleatorio estratificado, pero sin tener su característica de aleatoriedad. También aquí se divide a la población en estratos o subpoblaciones homogéneas y dentro de cada estrato se fija una «cuota» o número de individuos, que son seleccionados de forma accidental.

## Tema 3. Variables estadísticas y escalas de medida

### 3.1. Variables

Una variable es una característica, una propiedad o un atributo de una persona o un objeto, susceptible de asumir diferentes valores (que pueden medirse) en los diferentes sujetos. El concepto de *variable* se opone al de *constante*, que sería aquel atributo que únicamente puede tomar un valor para todos los sujetos.

### 3.2. Clasificación de las variables

Existen diferentes clasificaciones de las variables, en función del criterio utilizado:

<i>Criterio de medida</i>	<i>Criterio metodológico</i>	<i>Criterio de control</i>
Cuantitativas Continuas Discretas	Independientes	Aleatorias
Cualitativas Dicotómicas Politómicas	Dependientes	Controladas
Ordinales	De confusión	

Criterio de medida

### 3.3. Variables cuantitativas

Son las que pueden medirse numéricamente, es decir, que pueden cuantificarse mediante valores con significado matemático a partir de unas determinadas unidades de medida. Las variables cuantitativas, a su vez, pueden ser continuas o discretas.

#### 3.3.1. Variables cuantitativas continuas

Son aquellas en las cuales, entre dos valores consecutivos dados, podemos encontrar infinitos valores. La dificultad para encontrar todos estos valores intermedios se deberá únicamente a las limitaciones del instrumento de medida. Son ejemplos de variables cuantitativas continuas:

- el peso de una persona en Kg