



TÉCNICO SUPERIOR
EN DOCUMENTACIÓN
Y ADMINISTRACIÓN
SANITARIAS

Validación y explotación de datos

COORDINADORAS

Manuel Familiar Ramos

Victoria Emilia Climent Sánchez



Autores

Coordinadores

Manuel Familiar Ramos

Licenciado en Farmacia Orientación Ecológica por la Universidad de Valencia. Diploma de Estudios Avanzados del programa de doctorado Biología Vegetal del Departamento de Biología Vegetal de la Universidad de Valencia. Profesor Técnico de FP de Sanidad desde 1991; formación en la especialidad de Documentación Sanitaria desde 1997 en distintos hospitales. En la actualidad, imparte distintos módulos de Procedimientos Sanitarios y Asistenciales en el IES Jorge Juan de Sagunto-Puerto. Sagunto, Valencia

Victoria Emilia Climent Sánchez

Licenciada en Farmacia, Orientación Sanitaria, por la Universidad de Valencia. Grado de Licenciatura en Farmacia. Programa de Doctorado por Química Física en 1991. Máster en Reproducción Humana realizado en el Instituto Valenciano de Estudios en Salud Pública por el Hospital La Fe, Valencia. Profesora Técnico de FP de Sanidad desde 1991, en la especialidad de Documentación Sanitaria desde 2010 en distintos hospitales. En la actualidad imparte distintos módulos de Procedimientos Sanitarios y Asistenciales, en la especialidad de Documentación y Administración Sanitarias, en el CIP del Complejo Educativo de Chestre. Valencia

Dedicatoria

Dedicado a Manoli y a Néstor, por todo su apoyo.

Agradecimientos

Los autores queremos agradecer en primer lugar a la Editorial Arán Ediciones la confianza depositada en nosotros para la confección de estos capítulos y de habernos ilusionado con este proyecto, tanto como ellos lo están.

También queremos agradecer a todos los profesores y profesionales que a lo largo de nuestra vida académica y profesional han contribuido mediante sus enseñanzas a aumentar nuestra formación y al perfeccionamiento de nuestros conocimientos.

Nuestro agradecimiento a Néstor García Durán, arquitecto, por su acertada infografía de algunos capítulos.

A nuestros respectivos cónyuges que durante meses han estado alentándonos, sobre todo en los momentos que más lo necesitábamos.

Y a los compañeros que de alguna manera también han contribuido con sus consejos y ánimos a llevar a buen puerto el libro.

Índice

Capítulo 1

Realización de análisis estadísticos	15
1. Muestreo. Métodos de observación de una población. Extracción de una muestra	16
2. Tablas de distribución de datos cualitativos. Gráfico de barras. Tablas de doble entrada.....	39
3. Distribuciones de frecuencias.....	48
4. Medidas de tendencia central y de posición. Media muestral. Mediana. Moda. Cuartiles y percentiles.....	60
5. Medidas de variabilidad o dispersión	75
6. Los diagramas de caja (<i>boxplots</i>).....	82
7. Herramientas informáticas para el tratamiento estadístico de datos clínicos.....	85
8. Análisis descriptivo. Informes estadísticos.....	86

Capítulo 2

Cálculo de los valores de probabilidad	99
1. Distribuciones de variables continuas.....	100
2. Distribución normal. Propiedades de la curva normal. Distribución normal tipificada. Valores típicos y desviación típica	103
3. Cálculo de probabilidad. Límites 2σ	110

Capítulo 3

Análisis de relaciones. Estadística bidimensional	129
1. Relación entre variables cuantitativas	130
2. Relaciones entre variables categóricas	141
3. Magnitud de la asociación para tablas de 2 x 2	149

Capítulo 4

Introducción a la epidemiología	159
1. Introducción a la epidemiología	160
2. Medidas en epidemiología	160
3. Medidas epidemiológicas de frecuencia: prevalencia e incidencia	162
4. Medidas de asociación	165
5. Medidas de impacto potencial	168
6. Tipos de estudios epidemiológicos	170

Capítulo 5

Construcción de un sistema de información sanitario	185
1. Los sistemas de información sanitarios	186
2. Descripción de los elementos de un sistema de información clínico-asistencial	193
3. Aplicaciones específicas para el registro y explotación de datos clínico-asistenciales	196
4. Control de calidad de los datos sanitarios	199
5. Aspectos legales de los sistemas de información	201

Capítulo 6

Manejo de la estructura, utilidades y validaciones del Conjunto Mínimo Básico de Datos	215
1. Definición. Composición y características del Conjunto Mínimo Básico de Datos	216
2. Comité técnico del CMBD del Consejo Interterritorial	221
3. Usos y aplicaciones del CMBD. Su integración como sistema de información para la gestión clínico-asistencial y administrativa	224
4. Explotación y validaciones	226
5. Envíos del CMBD. Periodicidad. Destinatarios	234

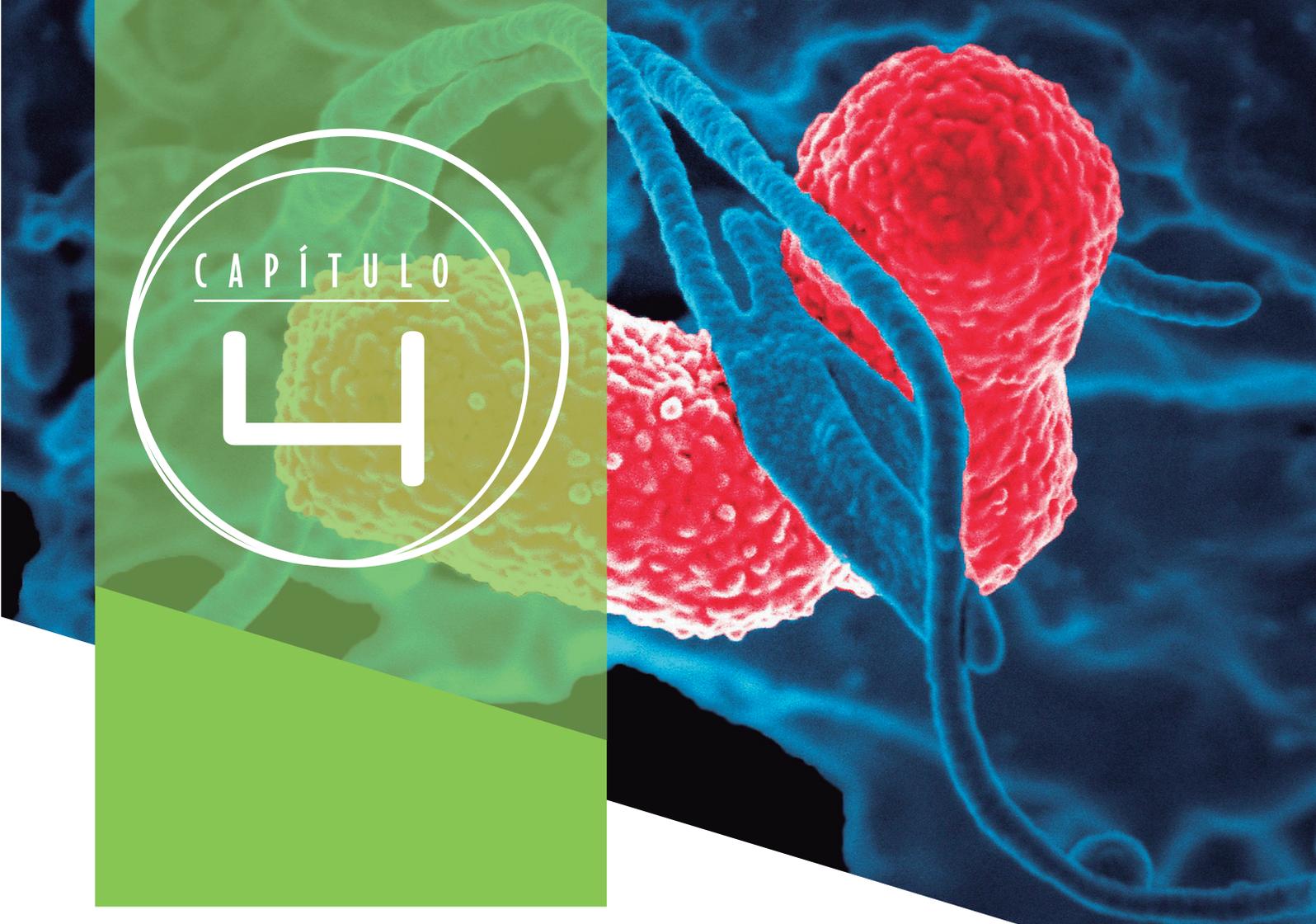
Capítulo 7

Identificación de los sistemas de clasificación de pacientes	243
1. Origen. Medicare y medicaid	244
2. Fundamentos, características y necesidad de los sistemas de clasificación de pacientes	244
3. Información necesaria para su elaboración	246

4. Modelos de clasificación de pacientes: <i>All Patient Diagnosis Relation Groups</i> (APGRD)	247
5. Estructura, jerarquía y algoritmos de clasificación. APGRD con información atípica o inválida	251
6. El concepto de <i>case mix</i> o casuística hospitalaria. Peso de los GRD	258
7. Aplicación de los sistemas de clasificación de pacientes en la gestión sanitaria	261

Capítulo 8

Calidad de la asistencia sanitaria	273
1. Historia y definición de la calidad en los servicios sanitarios. Necesidad de evaluar la calidad	273
2. El ciclo de mejora continua de la calidad	277
3. Gestión por procesos de la calidad asistencial	278
4. Medición de la calidad. Indicadores	281
5. La seguridad del paciente	288
6. Modelos para gestionar la calidad total	291
7. Acreditación y certificación	304
8. La investigación como garantía de calidad asistencial	305
Soluciones “Evalúate tú mismo”	320



CAPÍTULO

4

INTRODUCCIÓN A LA EPIDEMIOLOGÍA

*Manuel Familiar Ramos,
Victoria Emilia Climent Sánchez*

Sumario

1. Introducción a la epidemiología
 2. Medidas en epidemiología
 3. Medidas epidemiológicas de frecuencia: prevalencia e incidencia
 4. Medidas de asociación
 5. Medidas de impacto potencial
 6. Tipos de estudios epidemiológicos
- Resumen, glosario, abreviaturas y siglas, ejercicios y test de evaluación

Los **sistemas de información sanitaria** constituyen una herramienta fundamental para la obtención de información sobre el nivel de actividad de los servicios de salud o de los programas sanitarios. En este capítulo, veremos las **características y los componentes** de dicho sistema y su **utilidad** en la gestión en sentido amplio de los servicios sanitarios (tanto en régimen ambulatorio como de estancia hospitalaria).

Las principales **fuentes de datos clínicos** como la historia clínica, la filiación de los pacientes y los registros asistenciales son fundamentales para obtener un óptimo sistema de información sanitaria.

También conocerán los **aspectos legales** de los sistemas de información.



<https://www.boe.es/buscar/doc.php?id=BOE-A-2002-22188>

BOE 274, de 15 de noviembre de 2002



Podemos definir también el sistema de información sanitaria como una combinación de personas, métodos o procedimientos y servicios o equipos destinados a recoger, procesar, analizar y transmitir datos con objeto de generar la información necesaria a la organización para planificar, gestionar y evaluar el funcionamiento de los servicios de salud, así como la investigación y la formación.

I. LOS SISTEMAS DE INFORMACIÓN SANITARIOS

Podemos definir los **sistemas de información (SI)** como aquellos mecanismos por los que una serie de datos son procesados de una cierta manera para que posteriormente puedan proporcionar una información adecuada a la persona indicada y en el tiempo oportuno (Figura 1).

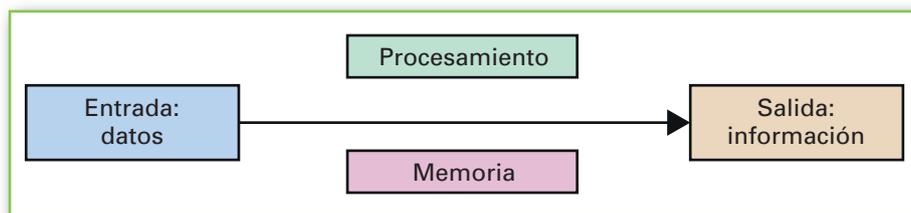


Figura 1. Esquema de un sistema de información.

La Organización Mundial de la Salud (OMS) (1973) define el **sistema de información sanitaria (SIS)** como “el mecanismo para recogida, procesado, análisis y transmisión de la información necesaria para organizar y hacer funcionar los servicios sanitarios, así como para la investigación y docencia”.

Podemos definir también el SIS como una combinación de personas, métodos o procedimientos y servicios o equipos destinados a recoger, procesar, analizar y transmitir datos con objeto de generar la información necesaria a la organización para planificar, gestionar y evaluar el funcionamiento de los **servicios de salud**, así como la investigación y la formación.

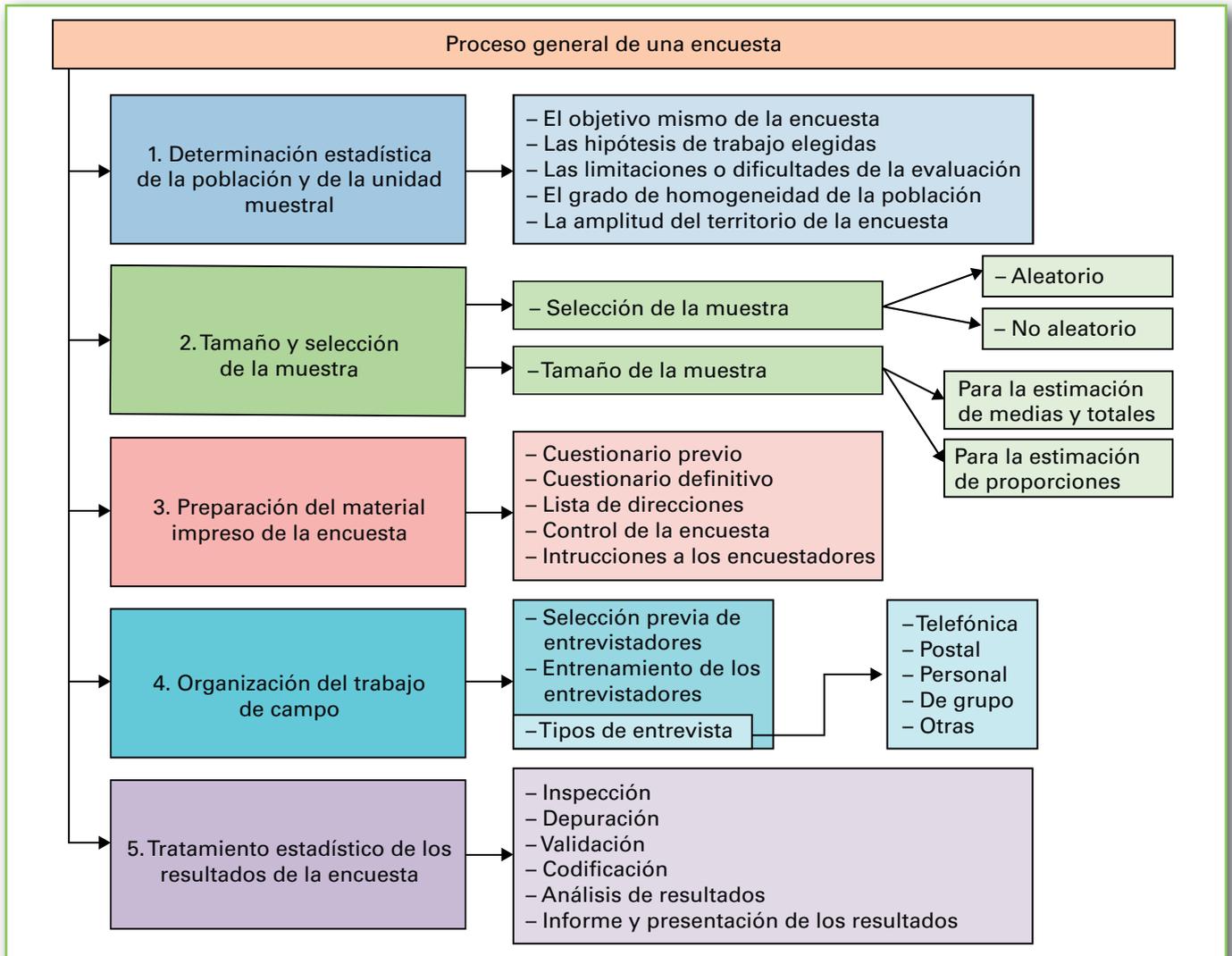


Figura 8. Proceso general de una encuesta.

Se denomina **coeficiente de elevación (K)** al cociente entre el tamaño de la población (N) y el de la muestra (n):

$$K = \frac{\text{Total de población candidata (N)}}{\text{tamaño muestra deseada (n)}}$$

Por ejemplo, si conocemos que el número de viviendas del barrio es de 10.000 (N) y que la muestra será de 400 viviendas (n), es preciso elegir las viviendas concretas en que han de realizarse las entrevistas.

$$K = N / n = 10.000 / 400 = 25$$

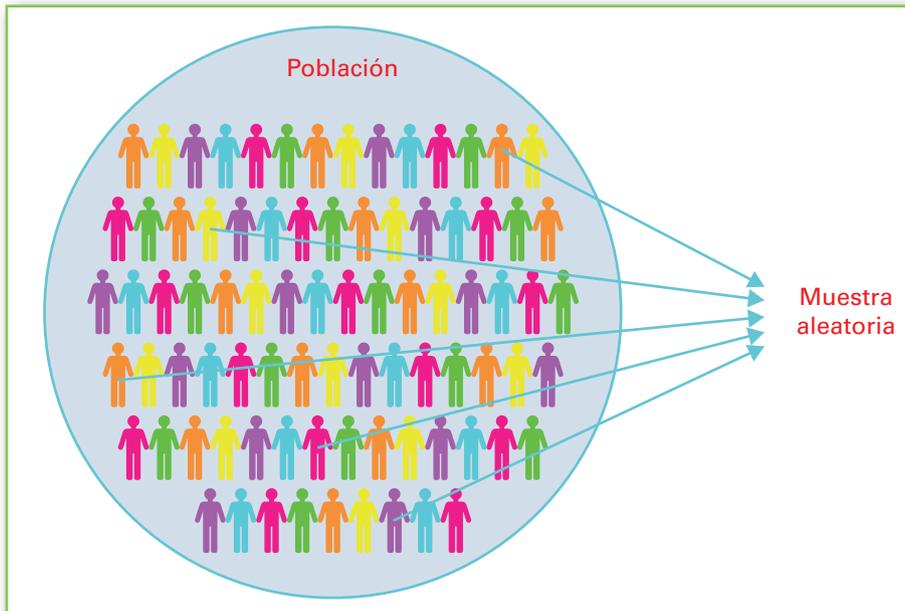


Figura 10. Muestreo probabilístico o aleatorio.

de ser seleccionados para formar la **muestra aleatoria simple** de tamaño n .

Se etiquetan previamente todos los elementos y se prepara un **listado de unidades de muestreo** numerándolas secuencialmente.

Por otra parte, debemos tener a la población de referencia en una **base de datos informatizada** o creada manualmente. Además, se deben **seleccionar** tantos números aleatorios como elementos tenga la muestra. La selección de los números aleatorios se puede hacer con el ordenador, con calculadora o con tablas de números aleatorios (**Anexo I**) que incluyen números de varias cifras. Puedes encontrar tablas aleatorias de 2 dígitos, 3, 4, 5, etc.



RECUERDA QUE

La unidad de muestreo es el elemento sobre quien se aplica la técnica de selección; por ejemplo personas, centros sanitarios, escuelas, etc.

INFORMACIÓN IMPORTANTE

VENTAJAS Y DESVENTAJAS DE MUESTREO ALEATORIO SIMPLE

Ventajas: es sencillo y de fácil comprensión.

Desventajas: requiere conocer de antemano la lista completa de todas las unidades de muestreo. Si la lista omite partes de la población, se perderán las ventajas del muestreo aleatorio.

Una muestra aleatoria simple da a cada individuo la misma oportunidad de ser escogido (evitando el sesgo por selección) y también da a cada muestra la misma oportunidad de ser escogida.



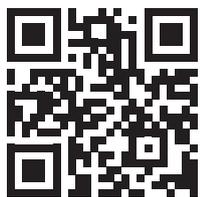
**RECUERDA QUE**

Al usar tablas aleatorias debes situarte en cualquier posición de la tabla y seleccionar “n” números consecutivos que tengan el mismo número de dígitos que el tamaño de tu población; así todos tendrán la misma probabilidad de ser elegidos.



<http://www.alazar.info/generador-de-numeros-aleatorios-sin-repeticion>

Números aleatorios
y números al azar



<https://www.random.org/>

Random

En el ejemplo anterior:

$N = 10.000$ (número de viviendas del barrio).

$n = 400$ viviendas (n.º de elementos a seleccionar para formar la muestra).

Paso 1. Etiquetas. A cada vivienda se le da una etiqueta que las identifique. En este caso, necesitaríamos etiquetas con cuatro dígitos para identificar a las 10.000 viviendas (del 0000, 0001, 0002, 0003... al 9999).

Paso 2. Tabla. Sitúate en cualquier posición de la **tabla del Anexo I** y lee números consecutivos de cuatro dígitos. Supón que elegimos al azar la fila 10 de la tabla de números aleatorio hasta elegir un total de 400 etiquetas, que después los seleccionaremos de la lista.

Muestreo aleatorio sistemático

Cuando el número de elementos de la población es elevado, el procedimiento anterior exige un esfuerzo considerable.

El **muestreo aleatorio sistemático** consiste en que, al conocerse el coeficiente de elevación, obtenemos al azar un número no superior a este coeficiente y este es el primer elemento de la muestra. A continuación, a ese número se le suma el coeficiente de elevación y el número obtenido es el segundo elemento, y así sucesivamente hasta obtener el tamaño de muestra requerido (Figura 11).

1. Calcular coeficiente de elevación:

$$K = N/n$$

2. Extraer la primera unidad al azar entre las k primeras unidades de muestreo y sumar la constante de forma sucesiva hasta completar el tamaño de la muestra.

Seguimos con el mismo ejemplo. Si el número de viviendas del barrio es de 10.000 y el tamaño de muestra es de 400 viviendas, es preciso elegir las viviendas concretas en que han de realizarse las entrevistas.

$$K = N / n = 10.000 / 400 = 25$$

Tomemos de la tabla un número inferior al coeficiente de elevación; por ejemplo el 18; entonces, los 400 elementos de la muestra corresponden a los números:

$$18, 18 + 25 = 43, 43 + 25 = 68... 18 + 399 \times 25 = 9.993$$

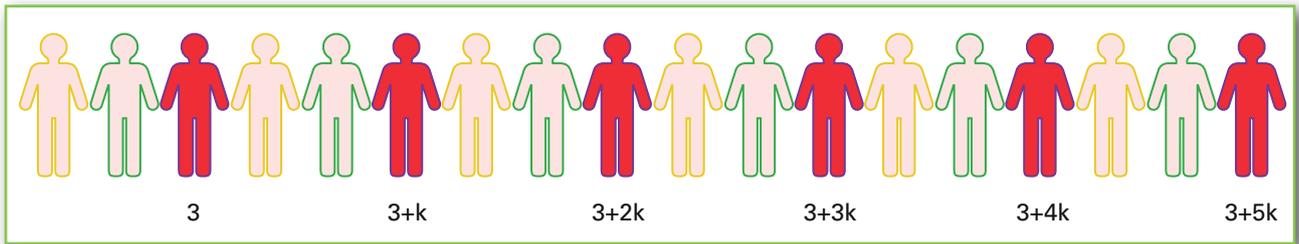


Figura 11. Selección individuos mediante muestreo aleatorio sistemático.

Debe tenerse cuidado con la aplicación de este método, ya que a veces puede llevar a obtener **muestras no representativas**. Esto puede ocurrir siempre que en la lista inicial los elementos no estén incluidos de una manera aleatoria.

Muestreo aleatorio estratificado

Tanto en el muestreo aleatorio simple como en el muestreo aleatorio sistemático, todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra. Esta característica, que resulta en muchos casos aconsejable, en otras no es interesante (Figura 12).

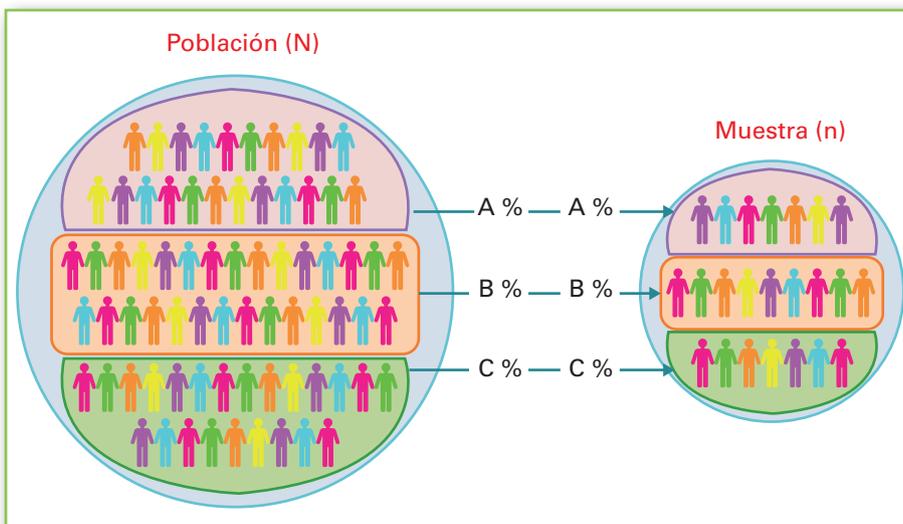


Figura 12. Muestreo estratificado por afijación proporcional.

Pongamos un ejemplo. En nuestro barrio se quiere hacer entrevistas sobre la **calidad de vida**. El total de viviendas del barrio es de 10.000 y el tamaño de muestra a seleccionar debe ser de 400 viviendas. Se sabe que existe relación entre la calidad de vida y el tipo de vivienda; como existen tres tipos de viviendas, A, B, C, de diferente calidad. El reparto de la muestra entre los diferentes estratos se conoce con el nombre de **afijación de la muestra**.

Tablas de doble entrada o tabla 2 x 2. Cada celda recoge las frecuencias o los porcentajes correspondientes a cada par de variables

TABLA 6

		Variable de interés		Totales
		Sí	No	
Exposición al factor	Sí	a	b	a + b
	No	c	d	c + d
Totales		a + c	b + d	a + b + c + d



<https://www.youtube.com/watch?v=dic8v899Ewl>

Riesgo relativo y Odds ratio

» **Analizar si existe alguna relación de asociación o no entre las variables objeto de estudio.** La columna de totales y la fila de totales de una tabla de contingencia dan las **distribuciones marginales** de las dos variables de forma separada. Estas nos informan sobre la relación entre las variables. Los valores absolutos por sí solos no permiten hacer comparaciones, por lo que en muchas ocasiones interesan las **distribuciones condicionales**. Para hallar la distribución condicional, calcula en cada celda el porcentaje de su valor respecto al total, al total de fila o de columna. La comparación de estas distribuciones condicionales es una manera de mostrar la asociación entre variables (Tabla 7).

EJEMPLO:
 En un estudio para conocer si el tabaquismo influye en la aparición de lesiones neoplásicas de pulmón se realizaron necropsias de 2.000 casos en los que se conocen sus antecedentes clínicos y hábitos tabáquicos. Se encontraron 480 lesiones neoplásicas de un total de 1.200 fumadores y 200 lesiones en un total de 800 no fumadores estudiados.



<https://www.youtube.com/watch?reload=9&v=XvPEeQAjTW8>

Test Chi-cuadrado

Tabla de 2 x 2. Frecuencias absolutas observadas en un estudio de investigación sobre la relación de cáncer de pulmón y hábitos tabáquicos

TABLA 7

		Fumador habitual		Totales
		Sí	No	
	Sí	480 (a)	200 (b)	680
	No	720 (c)	600 (d)	1320
Totales		1200	800	2000

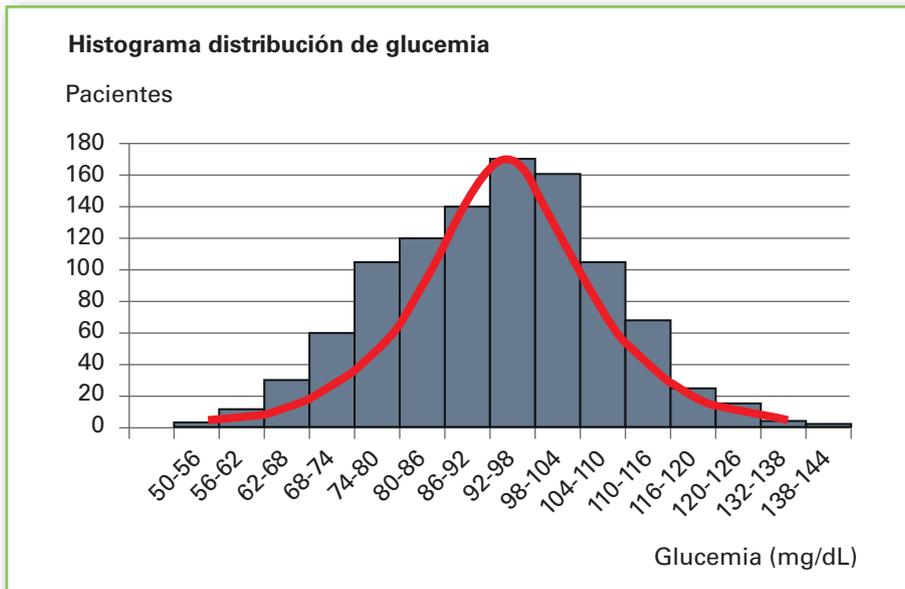


Figura 1. Histograma obtenido a partir de la determinación de glucemia (mg/dL) para 1.020 individuos.

Observa que la distribución es bastante simétrica, sin observaciones atípicas. La curva dibujada es un modelo matemático de la distribución idealizado, cuya función matemática **función de densidad $f(x)$** es la expresión en términos de ecuación matemática de la **curva de Gauss**, también conocida como curva normal, campana de Gauss o distribución normal (Figura 2).



RECUERDA QUE

La variable estadística está relacionada con la recolección de una muestra de datos y la obtención de frecuencias, mientras que la variable aleatoria siempre está vinculada con un experimento aleatorio. Se diferencian principalmente por el uso que se da a las variables aleatorias. Para las variables aleatorias se utilizan letras mayúsculas X, Y... Las letras minúsculas (x, y...) designan valores concretos de cada una de ellas.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ Para } -\infty < x < +\infty$$

- σ :** desviación estándar de la población.
- π :** constante pi (3,141592...).
- e:** constante matemática (2,718281...).
- X:** cualquier valor de la variable aleatoria continua.
- μ :** representa la media de la población.

La integral de $f(x)$ vale la unidad:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

TABLA 9

Clasificación para valorar la evidencia

Nivel	Descripción
Ia	Evidencia derivada de un metaanálisis o bien de diferentes estudios aleatorios
Ib	Evidencia derivada de datos de al menos un estudio aleatorizado
IIa	Evidencia derivada de resultados de al menos un estudio control sin aleatorización
IIb	Evidencia derivada de datos de al menos un estudio de tipo experimental
III	Evidencia derivada de estudios no experimentales como: estudios descriptivos, estudios comparativos, estudios de correlación o estudios caso-control
IV	Evidencia de una serie de casos
V	Opinión de un comité de expertos o bien de la experiencia clínica de una autoridad respetada

TABLA 10

Grados de recomendación de rigor científico

Grados de recomendación	Interpretación
A	Existe buena evidencia para recomendar la intervención clínica de prevención
B	Existe moderada evidencia para recomendar la intervención clínica de prevención
C	La evidencia disponible es conflictiva y no permite hacer recomendaciones a favor o en contra de la intervención clínica preventiva; otros factores pueden determinar la elección
D	Existe moderada evidencia para recomendar en contra de la intervención clínica de prevención
E	Existe buena evidencia para recomendar en contra la intervención clínica de prevención. Existe insuficiente información para recomendar a favor o en contra de la intervención; sin embargo, otros factores pueden mediar en la decisión

La **práctica de la evidencia científica** se usa como un referente de excelencia para guiar las decisiones clínicas e integra la propia experiencia con la de los demás y con las preferencias de los pacientes siendo una manera de mejorar la práctica médica, al disminuir la variabilidad y errores surgidos de la práctica sin evidencia e identifica el mejor estándar, diferenciándolo del resto de alternativas disponibles. El uso de la **evidencia científica** como garantía de la calidad asistencial se realiza teniendo en cuenta una serie de pasos:

- Realizar una pregunta de manera precisa a partir de un problema clínico. La pregunta debe ser relevante para la situación del paciente y ha de derivar a una fuente fiable donde buscar la respuesta (base de datos).
- Localizar las pruebas disponibles en la literatura a través de los motores de búsqueda y bases de datos internacionales.



<https://www.cebm.net/>

CEBM

RESUMEN

- ✓ En este capítulo hemos aprendido que la **estadística bidimensional** estudia la observación simultánea de dos variables X e Y , proporcionando un par de valores (X, Y) para cada individuo.
- ✓ Llamamos X a la **variable explicativa** de los cambios de una segunda variable Y a la que llamamos **variable respuesta**.
- ✓ Hemos aprendido a **examinar los datos de variables bidimensionales cuantitativas**.
- ✓ La **ecuación cuadrático-media** permite predecir el valor de \hat{Y} a partir de cualquier valor de x (regresión lineal).
- ✓ Las **tablas de contingencia** o de doble entrada nos permiten examinar o comparar los datos de dos o más variables categóricas a través de la frecuencia o el porcentaje de cada uno de los cruces de las categorías de las variables involucradas.
- ✓ Mediante las pruebas de **Chi-cuadrado** se puede descubrir si la relación es estadísticamente significativa, es decir, si no se debe al azar. Mediante el coeficiente ϕ y las medidas de asociación e impacto que estudiaremos en epidemiología podremos medir la fuerza de la relación.

G L O S A R I O

Coefficiente de correlación lineal de Pearson: índice que mide la dirección de la relación lineal e intensidad de la agrupación de los datos alrededor de una recta.

Coefficiente de determinación (R^2): medida estadística de la bondad del ajuste o fiabilidad del modelo de regresión.

Correlación: relación que existe entre las dos variables cuantitativas que intervienen en una distribución bidimensional.



EJERCICIOS

- › E1. Busca información sobre las otras familias de GRD.
- › E2. Busca más información sobre los IR-GRD.
- › E3. Haz una lista de servicios que el hospital presta y di de cada uno de ellos si se pueden contar o no.
- › E4. Busca información sobre los distintos SCP.
- › E5. Supongamos que cierta residencia tiene 158 residentes de los cuales hay: De la categoría A: 38; B: 25; C: 13; D: 11; E: 27; F: 25; G: 19.

<i>Case mix del centro</i>			
Categoría	N.º de residentes por categoría	Proporción de cada categoría	Ponderación de cada categoría
A	38		30,92
B	25		43,21
C	13		59,68
D	11		69,88
E	27		89,57
F	25		105,12
G	19		160,21
TOTAL	158		

- a) Calcula el *case mix* del centro.
- b) Calcula el ICMc, sabiendo que el CMs = 53,9710. ¿Qué significa esto?
- c) Supongamos ahora que el CMs = 99,8765. ¿Qué significa esto?

- › E6. Supongamos un hospital con 625 enfermos a los que hemos agrupado en 4 categorías, quedando la tabla siguiente:

<i>Case mix del centro</i>		
Categoría	Nº de enfermos por categoría	Ponderación de cada categoría
A	125	10
B	200	20
C	250	30
D	50	40

- a) Calcula el *case mix* del centro.
- b) Calcula el ICMc, sabiendo que el CMs = 27.
- c) Calcula el ICMc, sabiendo que el CMs = 20.

EVALÚATE TÚ MISMO



1. Si al calcular el coeficiente de correlación de dos variables X e Y $r = -0,20$ significa que:

- a) La pendiente de la recta de regresión es pequeña.
- b) Al ser el coeficiente negativo, la relación entre X e Y es directa y débil.
- c) X e Y están poco relacionadas; al disminuir X, tiende a aumentar Y.
- d) Todas las respuestas anteriores son correctas.

2. La covarianza de dos variables:

- a) Es la raíz cuadrada del coeficiente de correlación.
- b) Indica el sentido de la correlación de las variables y permite medir la variabilidad conjunta.
- c) Nos dice si la correlación entre dos variables es lineal.
- d) Es la media de las varianzas de ambas variables X e Y.

3. Dos variables numéricas son incorreladas, entonces:

- a) $r = 0$.
- b) $S_{xy} = 0$.
- c) La nube de puntos es dispersa, con forma redondeada.
- d) Todas las respuestas anteriores son correctas.

4. Indica con cuáles de los siguiente pares de variables puedes hacer un análisis de regresión lineal:

- a) La presión sanguínea y el grupo sanguíneo.
- b) El grupo sanguíneo y el factor Rh.
- c) El nivel de colesterol y la concentración de lípidos totales.
- d) La patología pulmonar y el hábito tabáquico.

5. Para determinar la intensidad de la correlación lineal entre dos variables:

- a) Calcular el coeficiente r de Pearson.
- b) Deben ser variables cualitativas.
- c) Calcular la covarianza.
- d) Hay que compararlo con el coeficiente r de la tabla.

6. Si la correlación es de $r = 1$:

- a) La covarianza es positiva.
- b) La relación entre ambas variables es muy fuerte y directamente proporcional.
- c) La pendiente de la recta es positiva.
- d) Todas las respuestas anteriores son correctas.

7. El coeficiente de correlación ϕ :

- a) Permite estudiar la intensidad de la relación entre dos variables cualitativas dicotómicas.
- b) Se le llama coeficiente de contingencia de Pearson.
- c) Su valor será cero cuando Chi-cuadrado es cero.
- d) Todas las respuestas anteriores son correctas.



SOLUCIONES EVALÚATE TÚ MISMO



http://www.aranformacion.es/_soluciones/index.asp?ID=41