

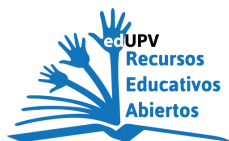
# Estadística aplicada al ámbito sanitario

Raffaele Vitale | Marta Hermenegildo Caudevilla  
Alberto J. Ferrer Riquelme



Raffaele Vitale  
Marta Hermenegildo Caudevilla  
Alberto J. Ferrer Riquelme

# **Estadística aplicada al ámbito sanitario**



[http://tiny.cc/edUPV\\_rea](http://tiny.cc/edUPV_rea)

Colección *Académica*

Para referenciar esta publicación utilice la siguiente cita:

Vitale, Raffaele; Hermenegildo Caudevilla, Marta; Ferrer Riquelme, Alberto J. (2022). *Estadística aplicada al ámbito sanitario*. Valencia: edUPV

Autoría

© Raffaele Vitale

Marta Hermenegildo Caudevilla

Alberto J. Ferrer Riquelme

Editado por edUPV, 2022

[www.lalibreria.upv.es](http://www.lalibreria.upv.es) / Ref.: 6368\_01\_01\_01

ISBN: 978-84-1396-024-1

DOI: <https://doi.org/10.4995/REA.2022.636801>

Si el lector detecta algún error en el libro o bien quiere contactar con el autor, puede enviar un correo a [edicion@editorial.upv.es](mailto:edicion@editorial.upv.es)



Estadística aplicada al ámbito sanitario / edUPV

Se permite la reutilización y redistribución de los contenidos siempre que se reconozca la autoría y se cite con la información bibliográfica completa. No se permite el uso comercial ni la generación de obras derivadas

# Autores

## RAFFAELE VITALE

Doctor en Estadística y Optimización por la Universitat Politècnica de València. Está especializado en el análisis de datos de naturaleza química, biológica y biomédica. Profesor titular de la Universidad de Lille, en Francia, donde combina su experiencia en estadística con su sólida formación en ciencias de la vida, desarrollando tareas docentes e investigadoras. Su trabajo se centra principalmente en el diseño y aplicación de metodologías y algoritmos para el procesamiento de imágenes hiperespectrales y de microscopía óptica. Colabora activamente con entidades e instituciones públicas y privadas en Italia, Países Bajos, Brasil, Noruega, Polonia y España.

## MARTA HERMENEGILDO CAUDEVILLA

Doctora en Farmacia por la Universitat de València y farmacéutica especialista en Farmacia Hospitalaria. Trabaja en el Servicio de Farmacia del Hospital Universitario Dr. Peset de Valencia donde coordina la Unidad de Atención Farmacéutica a Pacientes Externos. Ha sido jefa de servicio de Gestión y Coordinación de la Investigación Sanitaria de la Conselleria de Sanidad. Ha publicado numerosos artículos en revistas de ámbito nacional e internacional, así como capítulos de libros y monografías relacionadas con la farmacia hospitalaria y con la gestión de la investigación sanitaria.

## ALBERTO J. FERRER RIQUELME

Catedrático de universidad del Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, y director del grupo de Ingeniería Estadística Multivariante de la Universitat Politècnica de València. Especializado en el desarrollo y aplicación de técnicas estadísticas multivariantes y su integración con herramientas de *machine learning*, mantiene una intensa actividad docente, investigadora y de transferencia tecnológica en técnicas de análisis de datos y mejora de procesos en diversos ámbitos como la industria, la tecnología y el sector de la salud.

# Resumen

Esta monografía proporciona un resumen de las herramientas básicas de estadística descriptiva e inferencia estadística, con un apartado dedicado al análisis de supervivencia. Se presentan 3 ejemplos de aplicación de las técnicas con datos reales del ámbito clínico usando el programa estadístico SPSS. Está pensada para usuarios que quieran recordar de forma sencilla y rápida los conceptos básicos de estadística descriptiva y contrastes de hipótesis, así como tener una guía de qué técnica aplicar en función del problema a resolver. El capítulo dedicado al análisis de supervivencia está especialmente destinado a las personas interesadas en disponer de una guía práctica de cómo realizar este análisis con SPSS. Los ejemplos ilustrados son del ámbito de las Ciencias de la Salud, lo que puede ser especialmente útil, aunque no exclusivo, para todos los interesados de este ámbito.

# Preámbulo

Esta monografía proporciona un resumen de las herramientas básicas de estadística descriptiva e inferencia estadística (contrastes de hipótesis, ANOVA y modelos de regresión lineal y logística), con un apartado dedicado al análisis de supervivencia (método de Kaplan-Meier y modelo de regresión de Cox). Así mismo se presentan 3 ejemplos de aplicación de las técnicas con datos reales del ámbito clínico usando el programa estadístico SPSS. Está pensada para usuarios que, tras haber cursado alguna formación básica en herramientas estadísticas de análisis de datos, quieran recordar de forma sencilla y rápida los conceptos básicos de estadística descriptiva y contrastes de hipótesis, así como tener una guía de qué técnica aplicar en función del problema a resolver. El capítulo dedicado al análisis de supervivencia está especialmente destinado a las personas interesadas en disponer de una guía práctica (con su fundamentación conceptual) de cómo realizar este análisis con SPSS. Aunque todas estas técnicas se aplican en muchos campos de la ciencia y la tecnología, los ejemplos ilustrados son del ámbito de las Ciencias de la Salud, lo que puede ser especialmente útil para todos los alumnos y profesionales de este ámbito.

# Índice

<b>Preámbulo</b> .....	I
<b>1. Conceptos básicos de estadística descriptiva</b> .....	1
1.1 Variables estadísticas: definición y tipos .....	1
1.2 Descriptores de variables .....	2
1.3 Representación gráfica de variables .....	6
<b>2. Inferencia estadística</b> .....	11
2.1 Comparación de variables .....	12
2.1.1 Comparación de variables categóricas .....	13
2.1.2 Comparación de variables numéricas .....	14
2.2 Estudio de la relación entre una variable respuesta y una o varias variables explicativas .....	18
2.2.1 Variable respuesta numérica: regresión lineal simple o múltiple .....	19
2.2.2 Variable respuesta categórica: regresión logística .....	20
2.3 Análisis de supervivencia .....	23
2.3.1 Datos censurados .....	24
2.3.2 Estimación de la curva de supervivencia: método de Kaplan-Meier .....	26

2.3.3 Comparación de funciones de supervivencia: Log Rank test (o prueba de Mantel-Cox) .....	29
2.3.4 Análisis de curvas de supervivencia: modelo de regresión de Cox .....	30
<b>3 Ejemplos prácticos con Software SPSS</b> .....	<b>33</b>
Ejemplo #1: el caso de los antiinflamatorios .....	33
Ejemplo #2: comparación de tiempos de supervivencia de pacientes oncológicos .....	45
Ejemplo #3: relación entre el tipo de cáncer y el tiempo de supervivencia de pacientes oncológicos .....	50
<b>Bibliografía</b> .....	<b>55</b>



# Capítulo 1

## Conceptos básicos de estadística descriptiva

### 1.1. Variables estadísticas: definición y tipos

En el ámbito de la estadística, el término *variable* hace referencia a cualquier característica o valor numérico que se asocia, se mide o se cuenta para los individuos estudiados. La edad, el sexo, el peso, el número de hijos, el tipo de tratamiento recibido o el número de ciclos son ejemplos de variables.

Existen distintos tipos de variables que, en función de su naturaleza y propiedades, se tienen que tratar y analizar de formas diferentes. Las variables **numéricas** toman valores que describen una cantidad medible con un número y, por lo tanto, son variables **cuantitativas**. Las variables numéricas pueden clasificarse además como continuas o discretas: una variable **continua** es una variable numérica que puede tomar cualquier valor en el conjunto de los números reales, como el peso de una persona (65,1 kg; 70,234 kg; ...) o el tiempo de disolución de un fármaco (25,3 s; 0,4217 min; ...). Una variable **discreta** es una variable numérica que puede solamente tomar un valor numérico entero, por ejemplo, el número de ganglios linfáticos afectados en un proceso tumoral o el número de comorbilidades (0, 1, 2, ...).

Por otro lado, las variables **categorías** describen una cualidad o característica y suelen representarse por valores no numéricos (aunque también pueden representarse por valores numéricos que han de tratarse como simples códigos). Las variables categóricas pueden clasificarse como ordinales o nominales. Una variable **ordinal** es una variable categórica que toma valores ordenados o clasificados lógicamente, es

decir, estos valores son representativos de diferencias de rango entre los distintos individuos analizados. Algunos ejemplos de variables categóricas ordinales son el tamaño de un tumor primario (T1, T2, T3 T4) o el nivel de respuesta a la quimioterapia (respuesta completa, respuesta parcial, enfermedad estable, progresión). En cambio, una variable **nominal** es una variable categórica que toma valores que no pueden ordenarse según una secuencia lógica, como puede ser el sexo o el grupo sanguíneo. Los datos contenidos en una variable categórica son datos **cuantitativos**. La Figura 1 resume esta clasificación jerarquizada entre tipos de variables.

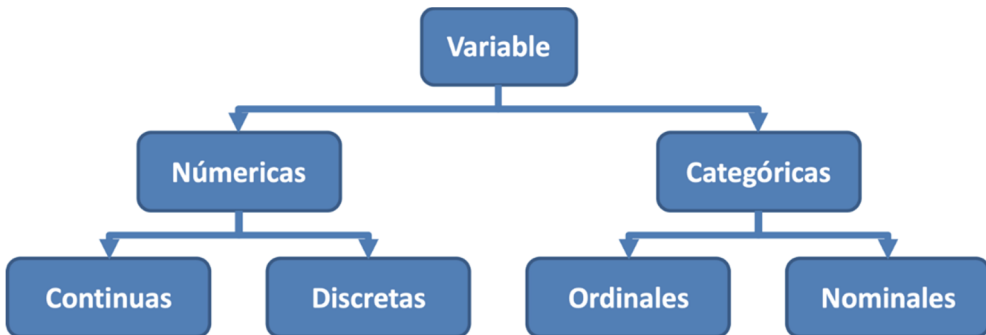


Figura 1 – Clasificación de los distintos tipos de variables

## 1.2. Descriptores de variables

Para resumir y evaluar la información que contienen las variables recogidas para los diferentes individuos incluidos en un estudio, se puede recurrir a varios tipos de medidas estadísticas en función de la naturaleza de dichas variables.

Para variables numéricas y categóricas ordinales, estas medidas se pueden clasificar en dos grupos en base al tipo de información que proporcionan: **medidas de tendencia central** y **medidas de dispersión**.

Las medidas de tendencia central corresponden a los valores centrales de la variable alrededor de los cuales se distribuyen los valores medidos para el conjunto de individuos. Las medidas de dispersión, por otro lado, cuantifican el grado de variación de los valores de la variable alrededor del valor central a lo largo de la muestra investigada. En base al tipo de variable (numérica o categórica ordinal), las medidas de tendencia central y de dispersión a utilizar varían. Algunas de las más utilizadas son:

a) Medidas de tendencia central (posición)

1. **Media aritmética** (para variables numéricas). Es el valor obtenido por la suma de todos los valores  $x_i$  de la variable dividida entre el número total,  $n$ , de valores:  $\bar{x} = \sum_{i=1}^n x_i / n$ . Su cálculo es muy sencillo y se interpreta como el “centro de masas” del conjunto de datos. Así, si la variable analizada se distribuye si-

métricamente alrededor de su valor central (es decir, si la distribución de valores a la derecha y a la izquierda de su valor central es similar), la media aritmética constituye una medida adecuada de tendencia central. Sin embargo, está muy afectada por la asimetría de la distribución de la variable y por valores extremos (anómalos) que la variable puede tomar en algunos individuos.

2. **Mediana**, también llamada **segundo cuartil** (para variables numéricas y variables categóricas ordinales). Es el valor que ocupa la posición central de un conjunto de observaciones ordenadas de menor a mayor (el 50 por ciento de las observaciones son mayores que este valor y el otro 50 por ciento son menores). A diferencia de la media aritmética, la mediana es una medida de posición adecuada para distribuciones asimétricas, y robusta (poco afectada) a la presencia de valores anómalos. Media aritmética y mediana coinciden en distribuciones perfectamente simétricas.
3. **Moda** (para variables numéricas y variables categóricas ordinales). Es el valor más frecuente de la variable. A veces, la moda no es única (puede haber más de una moda si múltiples valores se repiten con la misma frecuencia) y puede no situarse en el centro de la distribución de la variable.

De todas estas medidas de posición, las más usadas son la media aritmética para variables con distribución simétrica y la mediana para variables con distribución asimétrica. Es interesante destacar que la moda también podría calcularse para las variables categóricas nominales, pero no tendría ninguna interpretación como medida de tendencia central, sino simplemente como la característica más frecuente.

#### b) Medidas de dispersión (variabilidad)

1. **Desviación típica** (para variables numéricas). Se calcula 1) sumando los cuadrados de las diferencias de cada valor de la variable en cuestión y su media y 2) calculando la raíz cuadrada del ratio entre esta suma y el número total de valores medidos menos 1:  $s = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / n - 1}$ . Constituye una medida del grado de variación de los valores de una variable alrededor de su media aritmética. Al igual que la media aritmética, está muy afectada por la presencia de valores anómalos que la variable puede tomar y por la asimetría de su distribución (por lo que no es una medida de dispersión adecuada cuando la distribución no es simétrica). Una desviación típica baja indica que la mayor parte de los datos de una muestra tienden a estar agrupados cerca de su media, mientras que una desviación típica alta indica que los datos se extienden sobre un rango de valores más amplio. La desviación típica a veces se expresa elevándola al cuadrado, dando lugar a la **varianza** (otra medida de dispersión).
2. **Rango** (para variables numéricas y variables categóricas ordinales). Corresponde al intervalo entre el valor máximo y mínimo de una variable. Cuanto mayor es el rango, más dispersos serán los datos. Proporciona una idea inmediata

**Para seguir leyendo, inicie el  
proceso de compra, click aquí**