$$\text{Sesgo} = \frac{\text{Media-Moda}}{\text{Desviación típica}}$$

Ahora bien, esta fórmula requiere el cálculo de tres índices, por lo que se utiliza una fórmula más sencilla en base a los momentos de segundo y tercer orden, y que no es otra que el coeficiente B_1:

$$\text{Coeficiente de sesgo} = B_1 = \frac{m_3}{\sqrt{m_2^3}}$$

Si la curva está sesgada a la derecha, B_1 tendrá un valor positivo, mientras que si el sesgo es negativo, B_1 ofrecerá un valor negativo. Por ser una magnitud relativa, B_1 expresa la cantidad relativa de asimetría y puede ser utilizada para comparar distribuciones que contienen diferentes unidades de medición.

En cuanto a B_2, se utiliza como coeficiente de curtosis o medida del grado de apuntamiento de una distribución. Los valores pequeños de B_2 representan una curva platicúrtica (más baja que la curva normal), mientras que valores altos de B_2 indican una distribución leptocúrtica o apuntada. La curva normal tiene un valor de B_2 igual a tres. A continuación vamos a ocuparnos de este último tipo de distribución.

3.5. La distribución normal

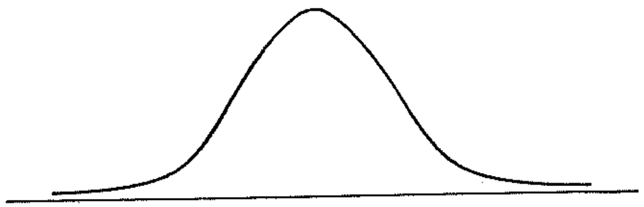
Vamos a tratar ahora un tipo especial de distribución de frecuencias, la curva normal, que es muy importante en el análisis estadístico. Tal distribución resulta útil no sólo porque un gran número de distribuciones de frecuencias presentan formas aproximadamente normales, sino también por la significatividad teórica de la curva normal en el campo de la estadística inferencial. Ahora no vamos a ocuparnos de este último aspecto, limitándonos a exponer las propiedades de la curva normal en relación a la desviación típica *.

Antes de continuar adelante conviene que distingamos entre *distribuciones de frecuencias finitas* y *distribuciones de frecuencias infinitas*. Las distribuciones que hemos visto hasta ahora siempre se han referido a un número finito de casos. Sin embargo, resulta útil, desde un punto de vista matemático, pensar en términos de distribuciones basadas en un número infinito de casos. Tales distribuciones vendrán representadas por curvas cuyos extremos se van acercando suavemente al eje X, pero sin cruzarse con él, y que, además, pueden expresarse por medio de ecuaciones matemáticas relativamente simples. La distribución normal es una curva de este tipo. Veamos algunas de sus características.

* Al estudiar las pruebas de decisión estadística y la teoría de las muestras en próximos capítulos, se hará evidente la utilidad de la distribución normal en la estadística inferencial. El objetivo de la presente sección es el de mostrar las propiedades de la curva normal y el uso de las tablas basadas en ella.

3.5.1. *La curva normal*

La curva normal responde al tipo de curva perfectamente simétrica, basada en un número infinito de casos, por lo que sólo puede ser tratada de forma aproximada cuando se opera con datos reales. Tiene una forma acampanada, tal como se observa a continuación:



Forma general de una curva normal

Por tratarse de una curva simétrica y unimodal, coinciden la media, la moda y la mediana. La ecuación matemática de la curva normal es relativamente simple, en la que intervienen la desviación típica s y las desviaciones de las puntuaciones con respecto a la media $|X-\overline{X}|$, de la forma siguientes:

$$Y=\frac{1}{s\sqrt{2\pi}}\, exp \left| -\frac{(X-\overline{X})^2}{2s^2} \right| \qquad [3.24]$$

en donde Y representa la altura de la curva para cualquier valor dado de X, valor de la puntuación en la abscisa; *exp* representa la base e de los logaritmos naturales, elevada a la potencia indicada entre paréntesis, y π es el número pi. No resulta necesario memorizar esta fórmula, sino recordar simplemente que en su composición intervienen la media y la desviación típica. Además, en la práctica nunca se utiliza la fórmula [3.24], ya que para operar con ella se utilizan unas tablas que dan directamente el área que queda por debajo de la curva normal para determinados intervalos. Esta tabla se ha podido construir basándose en una importante propiedad de la curva normal, y es que, con independencia de los valores particulares que tomen la media y la desviación típica de una curva normal cualquiera, habrá siempre un área constante (o proporción de casos) entre la media y una ordenada que se encuentre situada a una distancia dada con respecto a la media en términos de unidades de desviación típica.

En términos estadísticos, resulta conveniente considerar una curva normal cuyas puntuaciones se expresen en puntuaciones típicas —puntuaciones z— en lugar de sus unidades originales (pesetas, años, etc.). Es lo que se llama una curva normal tipificada, y al venir la variable X expresada en unidades de desviación, $z=|X-\overline{X}|/s$, la ecuación [3.24] queda sustituida por la forma llamada tipificada:

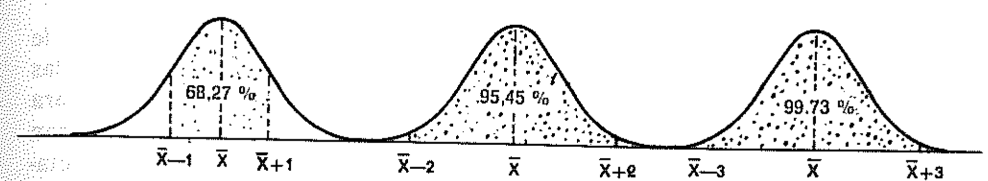
$$Y=\frac{1}{\sqrt{2\pi}}e^{-1/2\,z^2}$$

En este caso se dice que la curva se distribuye normalmente con media cero y varianza uno.

Un gráfico de esta curva normal tipificada se muestra en la figura 2, indicándose en el mismo gráfico las áreas incluidas entre $s=-1$ y $+1$, $s=-2$ y $+2$, $s=-3$ y $+3$, que son, respectivamente, el 68,27, 95,45 y 99,73 por 100 del área total, que, como se recordará, vale uno.

FIGURA 2

Areas bajo la curva normal



Dicho de otra forma, alejandonos una unidad, dos unidades o tres unidades de desviación típica con respecto a la media se encuentra el 68,27, el 95,45 y el 99,73 por 100, respectivamente, del área total.

Esta propiedad de la curva normal ofrece una interpretación de la desviación típica y un método para visualizar su significado. Y es que son muchas las distribuciones de frecuencias que son lo suficientemente parecidas a la distribución normal como para que en ellas se den también las anteriores relaciones entre áreas y desviaciones típicas. Incluso en el caso de distribuciones de ingresos económicos, o distribuciones de la talla y del peso de la población, que son ligeramente asimetricas en la dirección de los valores altos, habitualmente se puede encontrar que los dos tercios de los casos se encuentran dentro de una unidad de desviación típica con respecto a la media.

Los valores numéricos de cualquier curva normal pueden transformarse de tal modo que una sola tabla puede ser utilizada para evaluar la proporción de casos que queda dentro de un determinado intervalo. Supongamos, por ejemplo, que tenemos una curva normal de media 60 y desviación típica 15, y que deseamos conocer la proporción de casos que queda dentro del intervalo 60 a 85. Para ello, calculamos en primer lugar el número de unidades de desviación típica que separa a 85 de 60, y lo hacemos mediante la fórmula:

$$z=\frac{X-\overline{X}}{s}=\frac{85-60}{15}=1{,}66$$

El valor de $z=1{,}66$ indica que la ordenada se encuentra a 1,66 unidades de desviación típica con respecto a la media. Para saber la proporción de casos que queda dentro de dicho intervalo recurriremos a la tabla B del apéndice, en la que aparecen las áreas que quedan por debajo de la curva normal, para diferentes valores de z. Los valores de z aparecen en la columna de la izquierda y en la fila superior. Los dos primeros dígitos de z se obtienen leyendo a lo largo de la columna de la izquierda, y el tercer dígito leyendo en la fila superior. Las cifras que forman el interior de la tabla indican la proporción del área entre la media (que vale 0) y la ordenada correspondiente a z. En el ejemplo anterior, con $z=1{,}66$, el área que queda dentro de tales límites vale 0,4515. Si el valor de z hubiera sido 1,6, el área correspondiente hubiera sido 0,4452. Es decir, que aproximadamente el 45 por 100 de los casos queda dentro del intervalo 60 a 85 en la distribución normal de media 60 y desviación típica 15.

Aunque hemos dicho anteriormente que muchas distribuciones de frecuencias se asemejan a la distribución normal, son más todavía las que se alejan del modelo normal. En tal caso, no se pueden utilizar para estas distribuciones las propiedades de la desviación típica que se han visto al estudiar la curva normal. De ahí que para describir correctamente tales distribuciones habrá que recurrir a otras medidas de tendencia central, forma y variación.

3.6. TERMINOLOGÍA

Se recomienda la memorización y comprensión del significado de cada uno de los términos y conceptos siguientes:

— Posición o tendencia central de una distribución.
— Moda.
— Mediana.
— Media aritmética.
— Media geométrica.
— Media armónica.
— Media cuadrática.
— Variación o dispersión de una distribución.
— Recorrido o rango.
— Recorrido intercuartílico, recorrido semiintercuartílico.
— Desviación media.
— Desviación típica.
— Varianza.
— Coeficiente de variabilidad.
— Puntuaciones normalizadas o típicas.

— Simetría/asimetría de una distribución. Sesgo.
— Curtosis.
— Momentos de orden n.
— Distribución normal. Curva normal.

EJERCICIOS

1. Calcular la moda, mediana y media en la distribución de frecuencias del ejercicio 4 del capítulo 2.

2. Calcular la moda, mediana y media en la distribución de frecuencias del ejercicio 5 del capítulo 2.

3. En una encuesta de opinión pública la población se autoubicó en una escala ideológica izquierda-derecha (recorrido 1-10) tal como aparece en la siguiente distribución. Calcular la media y la mediana.

Escala izquierda-derecha	f_i
1- 2	13
3- 4	17
5- 6	42
7- 8	17
9-10	11

4. Las calificaciones de un estudiante en los cuatro exámenes parciales del curso fueron 5, 7, 6, 8. Si los pesos asignados a cada examen son 1, 2, 2, 1, ¿cuál es la nota final del curso? ¿Cuál sería si todos los pesos fuesen iguales?

5. El salario medio percibido por los empleados de una empresa es 80.000 pesetas. El salario medio de un hombre en dicha empresa es 85.000 pesetas y el de las mujeres 78.000 pesetas. Determinar el porcentaje de hombres y mujeres que trabajan en la empresa.

6. Calcular el recorrido, el rango intercuartílico, la desviación media, la varianza y la desviación típica en la distribución de frecuencias del ejercicio 4 del capítulo 2.

7. Calcular el recorrido, el rango intercuartílico, la desviación media, la varianza y la desviación típica en la distribución de frecuencias del ejercicio 5 del capítulo 2.

8. Si la media de una distribución normal es 70 y su desviación típica 8:

   a) ¿Qué proporción de casos se encuentra entre 70 y 85?

b) ¿Qué proporción de casos se encuentra entre 80 y 93?
c) ¿Qué proporción de casos es menor de 65?
d) ¿Cuántas unidades de desviación típica a ambos lados de la media hay que recorrer para obtener dos colas que contengan cada una de ellas el 3 por 100 del área total? ¿Y el 10 por 100?
e) ¿Qué puntuación tiene el 5 por 100 de los casos por encima de ella? (es decir, localizar el percentil 95).

9. Supóngase que una curva normal tiene una media de 50 y que el 7 por 100 de los casos tiene puntuaciones por encima de 70. ¿Cuál es la desviación típica?

BIBLIOGRAFIA

ALCAIDE INCHAUSTI, Angel: *Estadística aplicada a las Ciencias Sociales*, Madrid, Pirámide, 1976.
AMÓN, Jesús: *Estadística descriptiva para psicólogos*, Madrid, 1973.
BLALOCK, Hubert M.: *Social Statistics*, New York, McGraw-Hill, 1960.
DÍEZ NICOLÁS, H., y J. R. TORREGROSA: «Aplicación de la Escala de Cantril en España: Resultados de un estudio preliminar», *Revista Española de la Opinión Pública*, 10, 1967, págs. 77-100.
JIMÉNEZ BLANCO, José, et al.: *La conciencia regional en España*, Madrid, C.I.S., 1977.
LOETHER, H. J., y D. G. MCTAVISH: *Descriptive Statistics for Sociologists*, Boston, Allyn and Bacon, 1974.

Capítulo 4

ESTADISTICA INFERENCIAL: PROBABILIDADES Y TIPOS DE MUESTREO

4.1. INTRODUCCIÓN A LA ESTADÍSTICA INFERENCIAL

En los capítulos precedentes se han expuesto diversas técnicas para obtener y presentar de una forma resumida la información estadística, con el fin de facilitar la interpretación y el análisis de los datos. Este tipo de trabajo estadístico es lo que se denomina *estadística descriptiva*, ya que en realidad se utiliza para describir un grupo de individuos, características o *items* observados. Las medidas resumen que hemos estudiado en la estadística descriptiva no se pueden utilizar para obtener generalizaciones que sean aplicables a individuos, características o *items* que no hayan sido observados.

Muchas veces, sin embargo, el investigador está interesado en ampliar la indagación estadística, más allá del estudio de los objetos estudiados, a otras poblaciones de mayor alcance. Esta es precisamente la tarea que se realiza con la *estadística inferencial*, cuyo objetivo es la obtención de generalizaciones estadísticas sobre una población determinada, a partir del estudio de las características de una muestra extraída de dicha población o universo.

La sociología, como ciencia, aspira a establecer principios científicos y a predecir la conducta social. Precisamente, el sociólogo utiliza las técnicas que le brinda la estadística inferencial para realizar predicciones sobre el comportamiento de poblaciones determinadas, a partir del estudio directo de muestras pertenecientes a tales poblaciones. Veamos a través de un ejemplo el diferente uso que realiza el sociólogo de la estadística descriptiva y de la estadística diferencial.

Como es sabido, los demógrafos y los sociólogos de la población han desarrollado un esquema conceptual para describir los cambios demográficos que se producen en los países al pasar por diversos estadios de industrialización y urbanización. Tres son los tipos que se utilizan para describir los países: 1) de alto crecimiento potencial; 2) de crecimiento transicional, y 3) de decadencia incipiente. Tales tipos se definen en términos de tasas de natalidad, de mortalidad y de crecimiento vegetativo. Warren S. Thompson (1959), que desarrolló originalmente esta tipología, tiende a tratar los tres tipos como un esquema clasificatorio para los países del mundo. Pero hay otros autores que prefieren tratar los tres tipos como integrantes de una teoría de la población, llamada teoría de la transición.

Ahora bien, si se aspira a tratar los tres tipos como una teoría tendrá que someterse a la prueba de la verificabilidad. El valor de una teoría científica radica en su capacidad para predecir más allá de los datos que sirvieron de base para formularla. Si se desea que el esquema clasificatorio desarrollado por Thompson sirva como una teoría predictiva, será preciso realizar un análisis inferencial para comprobar su capacidad predictiva. Así, por ejemplo, se suele formular la hipótesis de que las estructuras políticas dominantes en los países varían según su estadio de crecimiento demográfico. Para contrastar dicha hipótesis se necesitará desarrollar indicadores fiables de estructura política, extraer una muestra de países del mundo, clasificarlos según su estadio de desarrollo, medir sus correspondientes estructuras políticas y, a través del uso del análisis estadístico, verificar el tipo de relación existente entre estructura política y crecimiento demográfico. Caso de que los resultados fueran positivos, se podrán generalizar al resto de los países del mundo.

En realidad, el campo de la estadística descriptiva no difiere en sus técnicas del campo de la estadística inferencial. La diferencia entre ambos campos de la estadística estriba en la manera de utilizar tales técnicas. Si las técnicas se utilizan tan sólo para resumir datos, se dice entonces que se trata de técnicas descriptivas. Si se utilizan para estimar parámetros de una población a partir de los cálculos realizados con los datos de una muestra, entonces se trata de técnicas inferenciales. Aquí aparece una dimensión terminológica que conviene tener siempre presente. Cuando nos refiramos a las características de una población hablaremos de *parámetros*, mientras que si nos referimos a características de la muestra tendremos *indicadores estadísticos* o, simplemente, *estadísticos*. Con el fin de diferenciar con toda claridad ambos tipos de características, se utilizan signos diferentes. Las letras griegas se utilizan habitualmente para referirse a las características de la población, mientras que las letras del abecedario latino se emplean con las características muestrales. Así, la letra griega mu (μ) representa la media aritmética de la población, mientras que la letra latina $\overline{X}$ denota la media aritmética de la muestra. Igualmente, la desviación típica de la población se representa por la letra griega sigma (σ), y la desviación típica de la muestra por la letra s. Los parámetros, que son valores fijos de la población, suelen desconocerse. Los estadísticos, que varían de muestra a muestra, se utilizan para estimar los parámetros. El proceso de estimación, eje de la estadística inferencial, se basa en la teoría de las probabilidades y en la teoría del muestreo.

4.2. Probabilidad: Nociones básicas y definición

Todos nosotros tenemos algún tipo de noción intuitiva del concepto de probabilidad, aunque no sepamos muy bien cómo definirlo. Si una persona afirma: «Es probable que mañana llueva», no tiene necesidad de explicar a su interlocutor el significado del término «es probable que...», ya que se sobreentiende que se refiere a la posibilidad de que pueda producirse «mañana» el suceso de «la lluvia». En este sentido, el concepto de probabilidad salpica el lenguaje común y la comunicación interpersonal cotidiana.

Ahora bien, si hemos de hablar con mayor precisión acerca del concepto de probabilidad, y especialmente si los matemáticos han de utilizarlo, se hace preciso definirlo con mayor rigor, con lo que surge una aparente contradicción. La *probabilidad matemática*, y las leyes del azar, se refieren tan sólo a sucesos repetidos bajo condiciones determinadas y constantes. Desde este punto de vista objetivista, no tiene sentido hablar de la probabilidad de un suceso concreto, tal como la probabilidad de que llueva mañana, dado que este suceso no es repetitivo. Tampoco se puede afirmar que la probabilidad de acertar una quiniela con resultados plenos es de uno entre un millón. Matemáticamente, «se acertará» o «no se acertará». Eso es todo. La probabilidad matemática tiene muy poco de probable. Tan sólo se admitirá la probabilidad de que, entre un millón de boletos, uno de ellos ofrezca un resultado acertado. La probabilidad matemática u objetiva se refiere al resultado medio de un gran número de apariciones del suceso u ocurrencias.

Pero si se acepta este punto de vista riguroso, el estadístico va a encontrar que a muchos problemas prácticos no es aplicable el concepto de probabilidad. Tan sólo se podrán aplicar las probabilidades a sucesos tales como la tirada de dados, los juegos de azar, los errores de una medición repetida, la producción en masa de un producto y otros sucesos en los que prevalece a largo plazo la variación aleatoria. Sin embargo, no podrá aceptar las afirmaciones probabilísticas de carácter socioeconómico, como, por ejemplo, que el desempleo tenderá a disminuir a lo largo del año, o que probablemente se recuperará la actividad económica en los dos próximos años, ya que para el matemático tales afirmaciones son simplemente correctas o incorrectas.

Desde el punto de vista de la *probabilidad real* o personalista, el término probabilidad se utiliza como una expresión del grado de creencia que una persona tiene de que un suceso vaya o no a ocurrir. Así, cuando alguien afirma que «es probable que vayan a convocarse pronto elecciones generales», está expresando su creencia de que tal suceso vaya a producirse, aunque puede ocurrir que otra persona opine sobre dicho tema de forma diametralmente opuesta. Esto es, que puede asignar una probabilidad «cero» a que se convoquen pronto las elecciones generales.

No obstante, los puntos de vista objetivista y personalista sobre la probabilidad no son tan diferentes como aparentan a primera vista. Porque cuando llega el momento de determinar prácticamente la probabilidad de un determinado suceso sólo existen dos métodos disponibles: el apriorístico y el empírico.

No existe una definición teórica de probabilidad universalmente aceptada. La más utilizada, y que se suele encontrar en la mayoría de los libros de texto, es la llamada *definición clásica de probabilidad:* sea un suceso determinado A, que de un total n casos posibles, todos ellos igualmente posibles, puede presentarse en un número a de los casos y no se presenta en los restantes b casos (siendo $b=n-a$). Entonces, la probabilidad apriorística de aparición del suceso A (llamado también la ocurrencia de A) viene dada por:

$$P(A)=\frac{a}{n}=\frac{\text{casos favorables}}{\text{casos posibles}} \qquad [4.1]$$

Es decir, la *probabilidad a priori* de ocurrencia del suceso A es, por definición, el cociente entre el número de casos favorables y el número de casos posibles. Esta definición supone que todos los casos sean igualmente probables.

Por ejemplo, supongamos que se tiene una urna con 10 bolas: una negra, cuatro blancas y cinco rojas. La probabilidad de que la primera bola extraída al azar sea negra vale $P(N)=\frac{1}{10}$; la probabilidad de que la primera bola sea blanca vale $P(B)=\frac{4}{10}$, y la probabilidad de que la primera bola extraída al azar sea roja vale $P(R)=\frac{5}{10}$.

La probabilidad igual a 1 significa certeza de ocurrencia del suceso. Si las 10 bolas de la urna hubieran sido blancas, la probabilidad de extraer una bola blanca será lógicamente 1. La probabilidad 0 indica, por el contrario, certeza de no ocurrencia del suceso, es decir, se trata de un suceso imposible.

Las probabilidades *a priori* se determinan, pues, en base a la lógica y a la naturaleza del suceso, en lugar de la experiencia o de la experimentación. Pero esta determinación de las probabilidades implica una dificultad lógica de razonamiento circular, ya que, tal como se ha dicho anteriormente, se basan en el supuesto de igual probabilidad o sucesos igualmente probables. Así, pues, la determinación de la probabilidad de los sucesos se basa en el conocimiento previo de las probabilidades de tales sucesos. Esto es fácil de saber en el caso de la tirada de los dados, o en la extracción de una bola de una urna, pero no ocurre así en la mayoría de los fenómenos sociales de interés para el sociólogo.

Por eso, el método empírico para determinar las probabilidades es de uso cada vez mayor. Las *probabilidades empíricas* se basan en el supuesto de que la proporción de aparición de los sucesos observada en el pasado persistirá en el futuro. Como reconoce Boris Parl (1967, pág. 83), las probabilidades empíricas son tan sólo estimaciones de las *probabilidades verdaderas*, pero cuanto mayor sea el número total de casos observados más precisa será la estimación. A través de este método es imposible obtener la probabilidad verdadera de un suceso, ya que ningún observador puede estudiar las tiradas de un dado durante un largo período de tiempo. Ahora bien, apoyándonos en la experiencia previa es posible obtener buenas estimaciones de los sucesos. Así es, por ejemplo, como las compañías aseguradoras estiman las tasas de fallecimiento para establecer los baremos de las pólizas de seguro de vida.

El tema de la teoría de la probabilidad ha provocado, ciertamente, la polémica y la controversia entre los matemáticos. Sobre él se ha escrito mucho, pero, afortunadamente, para establecer las bases de comprensión mínimas de las técnicas estadísticas utilizadas en la investigación sociológica no es preciso que profundicemos en el tratamiento estadístico de la probabilidad. El estudio elemental de algunas propiedades matemáticas de las probabilidades nos va a ser suficiente para poder seguir adelante en nuestra revisión del trabajo estadístico en la sociología empírica.

4.2.1. *Propiedades matemáticas de las probabilidades*

Tal como señala acertadamente Blalock (1960, pág. 102), aunque un estudiante de sociología puede que no necesite nunca calcular probabilidades, es importante que se percate de que, subyaciendo en cada tabla que vaya a utilizar para el contraste y verificación de hipótesis, se encuentran unas pocas y sencillas propiedades de las probabilidades. Por esta razón vamos a exponer a continuación algunas propiedades de las probabilidades empíricas de un suceso.

La primera propiedad ya la hemos visto anteriormente. La probabilidad de un suceso no puede ser mayor de la unidad (certeza total en la ocurrencia del suceso) ni menor de cero. Así, pues:

$$0 \leq P(A) \leq 1 \qquad [4.2]$$

en donde el símbolo $\leq$ significa «menor o igual que».

La segunda propiedad puede considerarse como un caso especial de la *regla de la adición:* si los sucesos A y B son mútuamente excluyentes, la probabilidad de obtener A o B —que se escribe $P(A \text{ o } B)$— es igual a la probabilidad de A más la probabilidad de B; esto es:

$$P(A \text{ o } B)=P(A)+P(B) \qquad [4.3]$$

Cuando decimos que los sucesos deben ser mutuamente excluyentes para que se cumpla [4.3], queremos decir que A y B no pueden ocurrir simultáneamente en el mismo experimento. En otras palabras, no se puede obtener en la misma tirada una cara y una cruz de una sola moneda.

Supongamos que la probabilidad de que un ciudadano español vote por un partido de derechas sea $P(A)=0{,}38$, y que la probabilidad de votar por un partido de izquierdas sea $P(B)=0{,}36$. Se trata, como vemos, de dos sucesos mutuamente excluyentes, ya que si se vota por un partido de derechas, no se puede votar por un partido de izquierdas. Entonces, la probabilidad de que un ciudadano español vote por un partido de izquierdas o por un partido de derechas será, aplicando la fórmula [2.3]:

$$P(A \text{ o } B)=0{,}38+0{,}36=0{,}74$$

La regla de la adición se puede extender al caso de más de dos sucesos. Si A, B, C, ..., K son sucesos mutuamente excluyentes, entonces:

$$P(A \text{ o } B \text{ o } C \ldots \text{ o } K)=P(A)+P(B)+P(C)+\ldots+P(K) \qquad [4.4]$$

Dado que las probabilidades son, en esencia, frecuencias relativas o proporciones, la suma de todos los sucesos posibles de un fenómeno ha de ser la unidad. Así, si en el ejemplo anterior añadimos la probabilidad de votar por un partido radical $P(C)=0{,}06$ (de derechas o izquierdas) y la probabilidad de no votar $P(D)=0{,}20$, a la probabilidad de votar por un partido de derechas $P(A)$ o por un partido de izquierdas $P(B)$, se ha de obtener una suma de 1. Entonces, para este ejemplo:

$$P(A)+P(B)+P(C)+P(D)=1$$

y la probabilidad de que no ocurra un suceso A será igual a la suma de las probabilidades de los restantes (mutuamente excluyentes) sucesos. Si sustraemos $P(A)$ de la unidad, tendremos la probabilidad de no obtener A, ya que:

$$P(A)+P(B)+P(C)+P(D)=1$$

y

$$1-P(A)=P(B)+P(C)+P(D)$$

La propiedad de no votar a un partido de derechas será, en nuestro ejemplo:

$$1-P(A)=0{,}36+0{,}06+0{,}20=0{,}62$$

Si los sucesos no son mutuamente excluyentes, la regla de la adición se formula del siguiente modo:

$$P(A \text{ o } B)=P(A)+P(B)-P(AB) \qquad [4.5]$$

en donde $P(AB)$ representa la probabilidad de obtener simultáneamente A y B.

Supongamos que, en una región española, el 75 por 100 de la población ha votado en las últimas elecciones municipales, que el 54 por 100 sea población femenina y que el 40 por 100 de la población sean mujeres que han votado. La probabilidad de que un residente en dicha región sea mujer o haya votado será, aplicando la fórmula [4.5], como sigue:

$$P(A \text{ o } B)=P(A)+P(B)-P(AB)=0{,}75+0{,}54-0{,}40=0{,}89$$

Una tercera propiedad de las probabilidades nos permite obtener la probabilidad de dos o más sucesos que ocurran simultáneamente. La *regla de la multiplicación* se puede formular del siguiente modo: si A y B son dos sucesos cualesquiera, la probabilidad de obtener simultáneamente A y B es igual a la probabilidad de obtener uno de ambos sucesos multiplicada por la probabilidad condicional de obtener el otro suceso una vez ha ocurrido el primer suceso. Es decir:

$$P(AB)=P(A)\,P(B/A)=P(B)\,P(A/B) \qquad [4.6]$$

en donde $P(B/A)$ y $P(A/B)$ representan las probabilidades condicionales. El término *probabilidad condicional* hace referencia a que la probabilidad del suceso A puede depender de la ocurrencia de otro suceso B.

Si la ocurrencia o no ocurrencia del suceso B no afecta a la probabilidad de ocurrencia de A, entonces $P(A/B)=P(A)$, y se dice entonces que A y B son *sucesos independientes;* por el contrario, si la ocurrencia de A depende de la ocurrencia de B, entonces se dice que los sucesos A y B son dependientes.

Veamos un ejemplo en que no existe independencia entre los sucesos. Supongamos que, de una población de mil jóvenes, la distribución numérica de los que manifiestan un carácter conflictivo o no conflictivo y la ideología con la que se identifican es la siguiente:

Rasgo	*Izquierda*	*Derecha*	*Neutro*	*Total*
Conflictivo	150	300	150	600
No conflictivo	300	50	50	400
Total	450	350	200	1.000

Ahora cabe preguntarse: ¿cuál es la probabilidad de que un joven cualquiera, elegido al azar, sea conflictivo e ideológicamente neutro? Dado que hay 150 jóvenes conflictivos y neutros de un total de 1.000, la probabilidad será $\frac{150}{1.000}$ o 0,15. Vamos a ver ahora cómo obtenemos esta misma probabilidad mediante la aplicación de la regla de la multiplicación.

Si A es el suceso de elegir un joven ideológicamente neutro y B el suceso de que el joven sea conflictivo, $P(A)=200/1.000=0,2$, ya que hay 200 jóvenes ideológicamente neutros en una población de 1.000 jóvenes y $P(B)=600/1.000=0,6$, ya que del total son 600 los jóvenes conflictivos. Entre los 600 jóvenes conflictivos hay, por otro lado, 150 que son ideológicamente neutros. Por lo tanto, entre la subpoblación conflictiva, la probabilidad de elegir un joven ideológicamente neutro es $150/600=0,25$. Entre los jóvenes ideológicamente neutros, la probabilidad de elegir uno que sea conflictivo es de $150/200=0,75$. Así, pues, tenemos que:

$$P(A)=0,2;\ P(B)=0,6;\ P(A/B)=0,25 \text{ y } P(B/A)=0,75$$

Aplicando ahora la fórmula [4.6] se obtiene la probabilidad de elegir un joven conflictivo ideológicamente neutro:

$$P(AB)=P(A)\,P(B/A)=(0,2)\,(0,75)=0,15$$
$$=P(B)\,P(A/B)=(0,6)\,(0,25)=0,15$$

Como vemos, con cualquiera de las dos expresiones se llega al mismo resultado.

Veamos otro ejemplo de aplicación de la regla de la multiplicación. Supongamos que el 35 por 100 de los jóvenes de edades comprendidas entre dieciocho y veintiún años se encuentra estudiando. Supongamos también que de esos jóvenes que están estudiando el 25 por 100 se pondrá a trabajar al cumplir los veintiún años, mientras que de los jóvenes de dichas edades que no están estudiando el 10 por 100 volverá a estudiar algún tipo de formación profesional al cumplir igualmente los veintiún años.

Dados estos datos, podríamos preguntarnos, si fuéramos a seleccionar aleatoriamente jóvenes de dieciocho a veintiún años, cuál es la probabilidad de que estuvieran estudiando (llamémosle suceso A) y que al cumplir los veintiún años se pusieran a trabajar (llamémosle suceso B). La probabilidad buscada se calculará, siguiendo la fórmula [4.6], del siguiente modo:

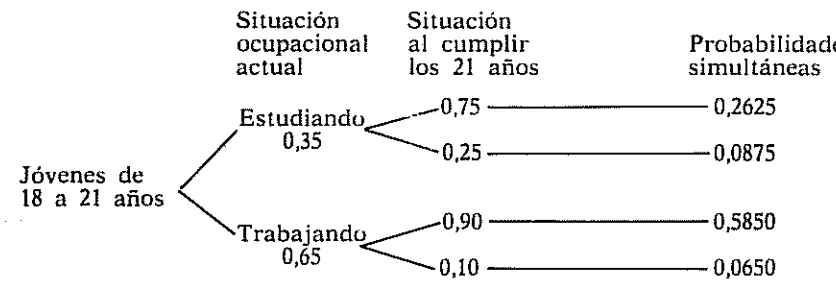
$$P(AB)=P(A)\,P(B/A)=(0,35)\,(0,25)=0,0875$$

También se podría calcular la probabilidad de encontrar jóvenes que no estuvieran estudiando (suceso A') y que vuelvan a reemprender sus estudios al cumplir los veintiún años, mediante la misma fórmula:

$$P(A'B)=P(A')\,P(B/A')=(0,10)\,(0,65)=0,065$$

Esta última probabilidad es un poco más baja que la primera, como consecuencia de que es menos probable que los jóvenes que se pongan a trabajar reemprendan o amplíen sus estudios.

Todavía cabe calcular otras dos probabilidades conjuntas. La probabilidad de encontrar jóvenes que estén estudiando y que no se pongan a trabajar al cumplir los veintiún años, que vale $(0,35)\,(0,75)=0,2625$, y la probabilidad de encontrar jóvenes que no estén estudiando y que no reemprendan ulteriormente sus estudios, que vale $(0,65)\,(0,90)=0,585$. Naturalmente, la suma de las cuatro probabilidades simultáneas vale la unidad. Obsérvese también que se puede llegar a calcular las cuatro probabilidades simultáneas multiplicando las probabilidades, siguiendo las ramas que se señalan en la figura:



La regla de la multiplicación general también se puede hacer extensiva a más de dos sucesos. En el caso de tres sucesos, A, B y C, la fórmula de la ocurrencia conjunta de los tres será como sigue:

$$P(ABC)=P(AB)\,P(C/AB)=P(A)\,P(B/A)\,P(C/AB) \qquad [4.7]$$

Cualquiera de las dos expresiones se puede utilizar para calcular el producto de probabilidades.

En la investigación sociológica, cada vez son más importantes los estudios longitudinales y de series temporales, que implican medidas del mismo fenómeno a lo largo del tiempo. Supongamos que disponemos de un modelo que explica cómo se desarrollan los sucesos de un fenómeno dado (por ejemplo, tasas de movilidad intergeneracional, precios, tasas de interés, etc.) en el tiempo. Tal modelo se denomina un *proceso*, y si está regido por leyes de probabilidad se denomina *proceso estocástico*. Siempre que tratamos una secuencia de sucesos a lo largo del tiempo y se calculan las probabilidades de su ocurrencia conjunta, tenemos procesos estocásticos. La palabra estocástico significa que los sucesos son probabilísticos en lugar de determinísticos, es decir, es posible asignar probabilidades a la ocurrencia de tales sucesos.

Una forma especial de proceso estocástico viene dada por las llamadas *cadenas de Markov*, denominadas así en honor del matemático ruso

Markov, que introdujo su concepto en 1907. Las cadenas de Markov pueden considerarse una aplicación de la regla general del producto de probabilidades y de las probabilidades condicionales. En realidad, una cadena de Markov es un proceso al azar que goza de la propiedad de que se puede predecir su futuro a partir del conocimiento del presente, junto con la historia del pasado. Supongamos que $P(E_k/E_i, E_j)$ es la probabilidad condicional de que, en el tiempo $n+2$, el sistema E (una familia, una sociedad, una persona, un organismo) se encuentra en el estado E_k, dado que en los tiempos n y $n+1$ el sistema se encontraba en los estados E_i y E_j, y supongamos también que tenemos las probabilidades condicionales para una secuencia más o menos larga de estados. Pues bien, un proceso es una cadena de Markov si:

$$P(E_k/E_i, E_j)=P(E_k/E_j);\; P(E_e/E_i, E_j, E_k)=P(E_e/E_k);$$

$$P(E_m/E_i, E_j, E_k, E_e)=P(E_m/E_e), \text{ etc.}$$

Como se ha dicho antes, el concepto de proceso estocástico y de cadena de Markov es de uso creciente en sociología, sobre todo en los estudios longitudinales y en aquellos que se basen en probabilidades condicionales.

Cuando dos sucesos son independientes entre sí, es decir, que la ocurrencia del suceso A no depende de lo que le ocurra al suceso B, sabemos que $P(B/A)=P(B)$ y $P(A/B)=P(A)$. En tal caso, la regla general de la multiplicación se simplifica, ya que entonces la probabilidad de la ocurrencia conjunta de sucesos independientes es igual al producto de sus probabilidades por separado. Es decir, para el caso de dos sucesos, A y B:

$$P(AB)=P(A)\,P(B) \qquad [4.8]$$

y en el caso de tres sucesos, A, B y C:

$$P(ABC)=P(A)\,P(B)\,P(C) \qquad [4.9]$$

Como ejemplo de aplicación de la fórmula [4.9], supongamos que estamos interesados en seleccionar al azar matrimonios en los que la esposa haya tenido gemelos, que sean niño y niña, y que lleven menos de diez años casados. En un principio, se trata de tres sucesos que son independientes entre sí, pues no parece que haya ninguna razón especial que los relacione. Si la probabilidad de tener mellizos es 0,01, la probabilidad de que sean niño y niña es 0,35 y la probabilidad de llevar casados menos de diez años es 0,15, la sustitución de estos valores en la fórmula [4.9] nos dará la probabilidad buscada:

$$P(ABC)=(0{,}01)\,(0{,}35)\,(0{,}15)=0{,}000525$$

Se trata, pues, de un suceso que va a ocurrir con poca frecuencia. En realidad, ocurrirá en 5 de cada 10.000 observaciones de matrimonios, según nuestros datos hipotéticos.

Ciertamente, suele ser difícil establecer la independencia de las variables en sociología. Con todo, la fórmula [4.9] es de utilidad en la investigación sociológica, sobre todo en la verificación de ciertas hipótesis estadísticas, como se verá más adelante.

4.2.2. *Combinatoria y probabilidad*

En la determinación de las probabilidades *a priori* se ha supuesto anteriormente que los diversos sucesos eran igualmente probables. Así, al lanzar una moneda, la probabilidad de obtener cara o cruz es la misma, y viene determinada por:

$$\frac{a}{a+b}=\frac{1}{1+1}=\frac{1}{2}$$

En este ejemplo elemental sólo se requiere contar el número de sucesos favorables y posibles. Pero en situaciones en las que hay que tratar con sucesos complejos, en lugar de sucesos elementales, el recuento de los sucesos alternativos puede ser una tarea muy complicada. En tales casos se hace necesario aplicar reglas matemáticas que nos den directamente el número de secuencias en que se pueden distribuir los sucesos favorables y posibles. A continuación vamos a estudiar algunas de las «técnicas de contar» que integran la «combinatoria» más elemental, por su relevancia para la comprensión de los conceptos estadísticos básicos que nos quedan por exponer.

Un principio es fundamental y previo a las diversas reglas combinatorias. Tal principio se puede enunciar así: si el fenómeno f_1 se puede verificar de n_1 maneras, el fenómeno f_2 de n_2 *maneras*, ..., y el fenómeno f_k de n_k maneras, las distintas maneras como pueden verificarse los k fenómenos vienen dadas por el producto $n_1 \times n_2 \times \ldots \times n_k$. Otro concepto y símbolo importante es el de *«factorial de un número n»*. Se representa por $n!$, y es igual al producto (1) (2) ... (n). Por ejemplo, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

4.2.2.1. *Variaciones y permutaciones*

Las *variaciones* se refieren a los distintos grupos que se pueden formar con m elementos tomados de n en n (siendo $n<m$), con la condición de que dos grupos serán distintos si difieren en el orden o en la natura-

leza de sus elementos. Las variaciones de m elementos tomados de n en n se representa por $V_{m,n}$ y su número viene dado por la fórmula:

$$V_{m,n}=m\,(m-1)\ldots(m-n+1)=\frac{m!}{(m-n)\,!} \qquad [4.10]$$

Si, por ejemplo, tenemos una población de cuatro elementos (a, b, c, d), los distintos grupos de dos elementos que podemos formar serán:

$$V_{4,2}=4\cdot 3=\frac{4!}{2!}=12$$

Estos grupos son: (a,b), (a,c), (a,d), (b,a), (b,c), (b,d), (c,a), (c,b), (c,d), (d,a), (d,b) y (d,c). Obsérvese que estos doce grupos o tienen elementos distintos o bien los tienen en orden distinto.

Cuando pueden darse repeticiones de los elementos, tenemos las *variaciones con repetición*, cuya fórmula es:

$$V_{m,n}=m^n \qquad [4.11]$$

Con los cuatro elementos del ejemplo anterior se pueden formar: $V_{4,2}=4^2=16$ grupos distintos, por lo que se refiere al orden o a la naturaleza de alguno de sus elementos. Tales grupos son (a,a), (a,b), (a,c), (a,d), (b,a), (b,b), (b,c), (b,d), (c,a), (c,b), (c,c), (c,d), (d,a), (d,b), (d,c) y (d,d). Como se puede observar, los grupos difieren en el orden o en la naturaleza de sus elementos, pero, a diferencia de las variaciones simples, hay cuatro grupos con los mismos elementos, es decir, con repetición.

Se tienen *permutaciones* de los elementos cuando los grupos varían tan sólo en el orden de los elementos que los integran. Pueden considerarse como un caso particular de las variaciones en las que $m=n$. Su fórmula es como sigue:

$$P_{n,n}=n! \qquad [4.12]$$

Con los cuatro elementos del ejemplo anterior se pueden formar: $P_{4,4}=4!=24$ grupos que difieren entre sí en el orden de sus elementos. Algunos de estos grupos serían $=(a,b,c,d)$, (a,b,d,c), (a,d,c,b), (a,c,b,d), (b,a,c,d), ... Se observa, pues, que los grupos sólo difieren en el orden de los elementos.

Si en los grupos se repiten algunos de los elementos, se tienen las *permutaciones con repetición*. Se pueden definir del siguiente modo: son los distintos grupos que se pueden formar con n elementos, dentro de los cuales se repiten n_1, n_2, ..., n_k elementos, con la condición de que dos grupos sean distintos si difieren en el orden de los elementos.

La fórmula para obtener el número de permutaciones con repetición es la siguiente:

$$PR_n=\frac{n!}{n_1!\;n_2!\;\ldots\;n_k!} \qquad [4.13]$$

4.2.2.2. *Combinaciones*

En los ejemplos anteriores, los grupos se consideraban distintos si variaban en ellos el orden o la naturaleza de los elementos. Pero podemos estar interesados en obtener grupos que sólo difieran entre sí por la naturaleza de los elementos y no por su orden. Para calcular el número de grupos que se pueden formar de esta manera tendremos que conocer las *combinaciones*, que se definen como los distintos grupos que se pueden formar con m elementos tomados de n en n (siendo $n<m$), con la condición de que dos grupos sean distintos si difieren en la naturaleza de alguno de sus elementos. Su fórmula viene dada por la expresión:

$$C_{m,n}=\binom{m}{n}=\frac{m!}{n!\,(m-n)!}=\frac{(m)\,(m-1)\ldots(m-n+1)}{n!} \qquad [4.14]$$

Para el caso de una población de cuatro elementos, el número de combinaciones que se pueden formar, tomados de dos en dos, es el siguiente:

$$C_{4,2}=\binom{4}{2}=\frac{4!}{2!\,(4-2)!}=\frac{4!}{2!\;2!}=\frac{4\cdot 3}{2}=6$$

Estos grupos son los siguientes: (a,b), (a,c), (a,d), (b,c), (b,d) y (c,d). Se observa, pues, que estos grupos difieren tan sólo en la naturaleza de los elementos.

Veamos ahora algunas propiedades de las combinaciones. La combinación de n elementos tomados de n en n vale la unidad, ya que:

$$C_{n,n}=\binom{n}{n}=\frac{n!}{n!\,(n-n)!}=\frac{n!}{n!\;0!}=\frac{n!}{n!}=1$$

ya que $0!=1$.

También se cumplen las siguientes combinaciones particulares:

$$\binom{n}{0}=1;\quad \binom{n}{1}=n;\quad \binom{n}{2}=n\,(n-1)/2$$

Con esto damos por finalizada la presentación de las reglas combi-