



Sobre Probabilidad e Inferencia Estadística

Jesús Montanero Fernández

Presentamos aquí un par de capítulos cuyo propósito es aclarar el uso del Cálculo de Probabilidades en la Inferencia Estadística. Puede servir de complemento al manual de Bioestadística que puede encontrarse en este mismo sitio, dado que en el mismo este tema se trata muy brevemente.

Capítulo 1

Probabilidad

La Teoría de la Probabilidad, que introduciremos en el presente capítulo, constituye una disciplina con autonomía respecto a la Estadística. De hecho, los inicios y motivaciones de ambas materias fueron absolutamente dispares: mientras que la primera surge del estudio de los juegos de azar, la segunda emana de la necesidad de clasificación e interpretación de datos referentes a poblaciones. La fusión de ambas especialidades se produce avanzado el siglo XIX, como consecuencia de diversos estudios acerca de la evolución de las especies. Intentaremos ilustrar más adelante el porqué de la conexión entre ambas materias.

En cuanto a la Probabilidad hemos de decir que, si bien sus comienzos pueden presentar cierto tinte de frivolidad, su campo de aplicación se ha ido extendiendo paulatinamente al describirse multitud de fenómenos, a parte de los consabidos juegos de azar, que se ajustan a lo que entendemos por fenómenos aleatorios. No obstante, existen diversas opiniones respecto a este hecho, algunas ciertamente radicales, pues el concepto de azar es objeto de polémica. En la primera sección del capítulo intentaremos precisamente profundizar en dicho concepto. Ya advertimos en la introducción que la mayor parte del capítulo puede pecar de excesivo formalismo, de ahí que se recomiende el lector interesado en la Probabilidad y Estadística como mera herramienta para el análisis de datos una lectura rápida, que no obstante puede ser suficiente para afrontar los capítulos siguientes. En todo caso aconsejamos tener bien presente al menos el párrafo en el recuadro de la sección 3.1, que supone en cierta medida una desmitificación del concepto de probabilidad.

1.1. Fenómeno aleatorio

En esta sección intentaremos delimitar qué entendemos por fenómeno aleatorio y fabricaremos el modelo matemático que lo formaliza.

1.1.1. ¿Sabe alguien qué es el azar?

Solemos decir que un fenómeno es determinista cuando podemos predecir su resultado. Por contra, existen multitud de fenómenos cuyo desenlace no puede preverse pues ofrecen múltiples posibilidades. En ese caso, se denomina **suceso** a cualquiera de las posibles situaciones que en principio puedan acaecer tras la ejecución del experimento. Vamos a centrar nuestra atención en aquellos fenómenos no deterministas que verifica la siguiente propiedad:

- (i) Pueden repetirse tantas veces como se quiera y aparentemente en idénticas circunstancias sin que el resultado de una ejecución pueda evidenciar una deformación o variación respecto a las mismas.

Tal podría ser el caso, por poner un ejemplo, de una serie de lanzamientos de una misma moneda. Efectivamente, no podemos predecir si el resultado de cada lanzamiento será cara o cruz, pero podemos aceptar que todos los lanzamientos se efectúan en igualdad de condiciones sin que el hecho de que un lanzamiento resulte cruz altere dicha perspectiva en los lanzamientos sucesivos.

En tal caso, diremos que un fenómeno de tipo (i) es aleatorio si puede explicarse mediante un modelo matemático en el que se asigna a cada suceso una medida precisa y cuantitativa de su grado de posibilidad denominada **probabilidad**. Podemos convenir que sea un número en el intervalo $[0, 1]$, de manera que un 0 significa que el suceso es **imposible** y un 1 significa que es **seguro**. De acuerdo con la condición (i), debe ser idéntica para toda la serie de ejecuciones. Además, la propia serie da lugar a sucesos compuestos (por ejemplo, tras dos lanzamientos de una moneda podemos hablar de los sucesos cara-cara, cara-cruz, cruz-cara o cruz-cruz). Teniendo en cuenta de nuevo la condición (i), la medida del grado de posibilidad de un suceso compuesto debe obtenerse de manera multiplicativa (es decir, la medida de la posibilidad de obtener cara-cruz se obtiene multiplicando la de cara por la de cruz). En este modelo matemático puede demostrarse, mediante un resultado denominado **Ley de los Grandes Números**¹ que, en una serie infinita de repeticiones, es **seguro** que la proporción de resultados favorables a un suceso converja a la medida del grado de posibilidad que le hemos asignado.

Por lo dicho anteriormente concluimos que, en un fenómeno aleatorio, la probabilidad que lo explica debe interpretarse para cualquier suceso como el límite al que tienden las proporciones de resultados favorables al mismo en una secuencia infinita de repeticiones. En consecuencia y desde un punto de vista empírico, consideraremos aleatorios (es decir, que siguen la Ley del azar) los fenómenos reales que verifican la propiedad (i) junto con esta otra:

- (ii) Para cualquier suceso considerado, las proporciones de resultados favorables al mismo tienden a estabilizarse tras un gran número de repeticiones.

La propiedad (ii) puede expresarse así: si A denota un suceso y $\hat{P}_n(A)$ la proporción de resultados favorables al mismo tras n repeticiones del experimento, es decir, la frecuencia relativa, existe un número $P(A)$ tal que

$$(ii) \quad \boxed{\lim_{n \rightarrow \infty} \hat{P}_n(A) = P(A)}$$

A continuación nos podemos plantear una serie de preguntas de carácter más filosófico que puramente matemático: ¿existen realmente fenómenos aleatorios? ¿Se deduce de la condición (i) la existencia de una probabilidad que rige el fenómeno y, en consecuencia, la condición (ii)?

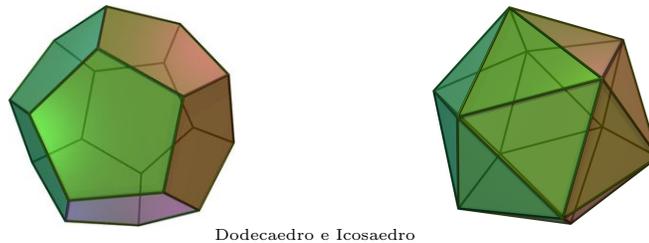
Respecto a la primera cuestión nos atrevemos a responder positivamente. Ahora bien, ¿se basa dicha respuesta en premisas racionales o es de carácter empírico? Son muchas preguntas que intentaremos responder parcialmente distinguiendo tres categorías de fenómenos².

¹Para más información consultar Billingsley, *Probability and Measure*, Willey (1986).

²En Gnedenko, *Theory of Probability*, Chelsey Publishing Company (1967), se puede encontrar un análisis más profundo de esta problemática.

Fenómenos a priori aleatorios

Nos referimos a aquéllos que responden a una clara *simetría*. Es el caso de una ruleta (círculo), una lotería (esferas), el lanzamiento de una moneda, de un dado convencional, es decir, un cubo, o de cualquier otro sólido platónico: tetraedro, octaedro, dodecaedro o icosaedro regulares. Cabe incluso conjeturar si en el fondo de todo fenómeno aleatorio existe una razón relacionada con la simetría, es decir, que lo que comúnmente denominamos *azar* no sea sino la consecuencia de una simetría más o menos compleja y más o menos evidente.



Dodecaedro e Icosaedro

En todo caso, en fenómenos de este tipo no parece que haya mucho inconveniente en asumir que pueden repetirse tantas veces como se quiera en igualdad de condiciones sin que el resultado de una ejecución condicione el de las restantes. Podemos aceptar pues la propiedad (i). En estas circunstancias, la propia geometría nos conduce al concepto de *equiprobabilidad*, pues no parece tampoco difícil convencerse de que, por ejemplo, los 6 lados de un cubo perfecto (simétrico) tienen un mismo grado de posibilidad de quedar arriba una vez lanzado el cubo (dado).

Nótese que es la propia simetría la que nos ha permitido asumir la propiedad (i). Efectivamente, si consideramos que la simetría en el fenómeno es extensible a la repetición del experimento, nada hace pensar que, en 10 ejecuciones del lanzamiento del dado, alguna de las 6^{10} posibles series resultantes tenga mayor grado de posibilidad de ocurrir que el resto. En particular, el experimento puede repetirse sin que el resultado de un lanzamiento condicione el de otro distinto.

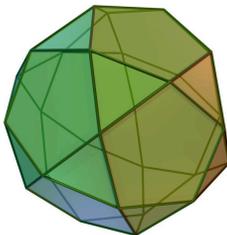
Parece pues claro que estamos en disposición de formular un modelo matemático en el que asignamos a cada suceso una medida *a priori* de su grado verosimilitud. En este caso, a cada lado del cubo se le asigna probabilidad $1/6$. La comprobación empírica de que, tras una larga serie de lanzamientos de un dado, las proporciones de resultados favorables a cada puntuación sea próxima a $1/6$ debe considerarse un claro signo de que el modelo matemático propuesto se adecua satisfactoriamente al fenómeno real estudiado. Y, efectivamente, si el lector está lo suficientemente aburrido podrá comprobar cómo tras lanzar 100, o mejor 1000 veces un dado, la proporción de cincos obtenidos es muy próxima a $1/6$. No descartamos que dicha proporción no converja realmente a $1/6$, o que ni siquiera converja, pero estamos predispuestos a interpretar ese hecho como un defecto de construcción del dado. Es lo que denominaríamos un dado trucado

Fenómenos aleatorios a posteriori

Podemos pensar en otro tipo de fenómeno aleatorio que, al menos en apariencia, no se explica por un argumento de pura simetría. Son comúnmente admitidos como fenómenos aleatorios las variaciones accidentales en un proceso de medición o fabricación. Decimos que son fenómenos aleatorios en tanto en cuanto se dan la propiedades (i) y (ii). Puede que la primera pueda asumirse en virtud de la propia naturaleza del experimento, como ocurre con el lanzamiento de una moneda; sin embargo, se antoja imprescindible contrastar empíricamente la segunda

propiedad (ley de azar), pues su violación dejaría patente la ineptitud del modelo matemático basado en el concepto de probabilidad a la hora de formalizar el fenómeno real.

Nos preguntamos, por ejemplo, si el lanzamiento de un sólido arquimediano, como el icosidodecaedro (sus caras forman 20 triángulos equiláteros y 12 pentágonos regulares) es un fenómeno aleatorio.



Icosidodecaedro

Posiblemente lo sea. No obstante, aunque podamos asumir la condición (i), respecto a la condición (ii) convendría contabilizar el número de veces en las que el poliedro cae sobre un pentágono y comprobar que la proporción de resultados favorables tiende a estabilizarse a medida que repetimos el experimento. Sólo entonces, es decir *a posteriori*, podremos aceptar que el fenómeno es aleatorio y la probabilidad de caer en pentágono será parecida a la frecuencia relativa de la serie.

Aunque sucediera eso no podemos pensar en una probabilidad universal para todos los icosidodecaedros, pues no se puede descartar que las frecuencias relativas converjan a distintos números dependiendo de si el poliedro utilizado es hueco o macizo, o incluso de su volumen, densidad, etc. En todo caso, en fenómenos de este tipo las probabilidades correspondientes a cada suceso no pueden calcularse *a priori*. Sólo podemos obtener una aproximación empírica a las mismas tras una larga serie de repeticiones.

Fenómenos inciertos

Entrando ya plenamente en el campo de la conjetura nos atrevemos afirmar que son los más abundantes (quizás los únicos). Nos referimos a fenómenos como un partido de fútbol, la presencia de una enfermedad, la talla de un recién nacido, etc. No pueden considerarse aleatorios según hemos convenido dado que no verifican la condición (i), pues ni siquiera pueden volver a repetirse en idénticas circunstancias. Por lo tanto y en rigor, no deberíamos hablar de probabilidad en estos casos. Este abuso del término, muy frecuente en el lenguaje habitual, se debe a el principio asumido en muchos ámbitos de que los fenómenos se dividen en dos clases: deterministas y aleatorios. Posiblemente sea más riguroso decir que todo fenómeno descompone en una componente determinista y otra aleatoria.

Efectivamente, este tipo de fenómenos inciertos no pueden repetirse en idénticas condiciones porque existen unas causas o factores concretos que influyen en el resultado y fluctúan en las diversas repeticiones del experimento. La conjunción de dichos factores da lugar a una **componente determinista** a la que posiblemente se sume otra **componente aleatoria** en sentido estricto que sí verifica las condiciones (i) y (ii). De hecho, desde el punto de vista estadístico el diseño de un experimento tiene como objetivo aislar lo mejor posible esa componente aleatoria pura.

Por poner un ejemplo, no podemos afirmar que el lanzamiento de icosidodecaedros sea un fenómeno aleatorio porque es muy posible que las tendencias en los lanzamientos dependan de

factores como el volumen o densidad del objeto utilizado. Sin embargo, si controlamos estos factores, lo cual puede conseguirse utilizando el mismo objeto en toda la serie lanzamientos, tal vez podría considerarse aleatorio a *posteriori*. En ese caso, podríamos diseñar varios experimentos paralelos con icosidodecaedros de distinto volumen o composición, para determinar si estos factores influyen realmente en las probabilidades obtenidas en cada caso.

La descomposición de los fenómenos en componentes deterministas y aleatorias viene a ser una solución ecléctica entre dos posturas radicalmente enfrentadas: por un lado, la postura determinista que entiende que se habla de simetría o equiprobabilidad en aquellas circunstancias en las que renunciamos por completo a controlar las causas del resultado; es decir, que la clasificación de fenómenos en deterministas y no deterministas no obedece a la naturaleza de los mismos sino a nuestra capacidad o actitud a la hora de explicarlos. Desde ese punto de vista, el azar no sería más que una especie de saco donde se refugian las causas que no podemos o no queremos controlar. Cabría entonces esperar que el progreso científico fuera menoscabando los dominios del azar. No obstante, parece haber sucedido lo contrario. Efectivamente y aunque la siguiente afirmación siga siendo objeto de polémica, la Mecánica Cuántica introduce con éxito empírico el concepto de azar para explicar la observación de lo muy pequeño, lo cual parece tumbar los principios deterministas³. Es más, algunos científicos piensan que se puede ir más allá de manera que sea ese azar puro constituya la explicación última de todo fenómeno, aunque sea macroscópico, de manera que incluso aquello que damos por seguro sea sólo muy probable? Ésa sería la postura contraria.

1.1.2. El modelo de probabilidad

Al margen de estas disquisiciones infructuosas desde un punto de vista práctico convendrá saber con mayor detalle en qué consiste el modelo matemático que formaliza un fenómeno considerado aleatorio, con razón o sin ella, de manera meditada o precipitada (eso deja de importarnos en el momento que nos situamos en un plano puramente formal). Primeramente, debemos distinguir el modelo que corresponde a una única ejecución del experimento del que corresponde a una serie de n repeticiones verificando (i). Al primero lo denominaremos aquí modelo de probabilidad original y al segundo, modelo de probabilidad producto. Advertimos ahora y volveremos a hacerlo al final de la sección que estas distinciones no responden en esencia a aspectos formales sino sólo didácticos, y que en la práctica podremos hablar de un único modelo de probabilidad, a secas.

Modelo original

Pensemos como ejemplo en el lanzamiento de un dado simétrico. Lo primero que debemos tener en cuenta es el conjunto pormenorizado de los posibles resultados en que puede desembocar el experimento. Dicho conjunto, que se denota por la letra Ω , se denominará **espacio original**. En el caso del dado distinguiremos seis posibilidades, tantas como caras tiene el cubo, es decir:

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

Si entendemos por **suceso** cualquier circunstancia susceptible de ocurrir tras la ejecución del experimento, debemos definirlo formalmente como cualquier subconjunto de Ω . Por ejemplo,

³El lector interesado puede acceder a una visión más profunda del tema en Martínez Muñoz, *El azar en la mecánica cuántica: de Bohr a Bell*, Crítica (Revista Hispanoamericana de Filosofía), vol XXIII (69), 1991, pp. 137-154

que ocurra el suceso $\{2, 4, 6\}$ significa que el resultado del lanzamiento sea **par**. En general, decimos que se verifica un suceso cuando el resultado del experimento es cualquiera de los elementos que lo componen. El propio espacio Ω es un suceso que, por lo tanto, ocurre siempre, de ahí que se denomine **suceso seguro**. Por contra, el espacio vacío \emptyset no ocurre nunca pues suponemos que el experimento aporta siempre un resultado, de ahí que se denomine **suceso imposible**. Es un elemento necesario en el álgebra de sucesos. Los elementos de Ω son a su vez sucesos, con la particularidad de que no pueden descomponerse en otros más simples. Se denominan **sucesos elementales**.

El conjunto de los sucesos está dotado de un algebra que nos permite unirlos, intersecarlos y complementarlos. Concretamente, dados dos sucesos A y B , se verificará la unión $A \cup B$ cuando se verifique A o B (o ambos); se verificará la intersección $A \cap B$ cuando se verifiquen simultáneamente A y B , y el complementario \bar{A} cuando no se verifique A . Decimos que dos sucesos son incompatibles o disjuntos cuando $A \cap B = \emptyset$.

Una vez configurado el espacio inicial y, en consecuencia, el conjunto de los posibles sucesos, debemos asignar a cada uno de ellos su **probabilidad**, que será un número en el intervalo $[0, 1]$ que asigne un 1 al suceso seguro y con las características propias de una medida, es decir, que si A y B son incompatibles entonces

$$P(A \cup B) = P(A) + P(B)$$

La probabilidad de cualquier suceso es igual por lo tanto a la suma de las probabilidades de los sucesos elementales que lo componen. En el caso de que la aleatoriedad del fenómeno responda a una simetría perfecta, como es el caso del dado, los sucesos elementales serán equiprobables. Por lo tanto, cuando se da una simetría perfecta, la probabilidad de un suceso cualquiera será igual al número de sucesos elementales que lo componen dividido por el número total de sucesos elementales, es decir, será el cociente entre el número de casos favorables al suceso y el número de casos posibles. Así, por ejemplo, la probabilidad de que el resultado de un lanzamiento sea par es $3/6$.

Hemos visto que existe compatibilidad entre la unión disjunta de sucesos y la suma de probabilidades. ¿Es también cierto que la intersección de sucesos se traduce en el producto de sus probabilidades? En general no. Por ejemplo, en el caso del lanzamiento de un dado, la intersección de los sucesos **par** e **impar** es el conjunto vacío, luego su probabilidad es nula. Sin embargo, la probabilidad de **par** multiplicada por la probabilidad de **impar** es igual a $1/4$. Decimos que dos sucesos A y B son **independientes** cuando sí se verifica que $P(A \cap B) = P(A) \times P(B)$. En caso contrario se dice que son dependientes.

Por ejemplo, son independientes los sucesos **múltiplo de 2** y **múltiplo de 3**. Efectivamente, el primero está compuesto por $\{2, 4, 6\}$ siendo su probabilidad $1/2$; el segundo está compuesto por $\{3, 6\}$ siendo su probabilidad $1/3$; la intersección de ambos sucesos es el suceso elemental $\{6\}$, cuya probabilidad puede obtenerse multiplicando $1/2$ por $1/3$. Un ejemplo más ilustrativo del concepto de independencia podemos encontrarlo en el lanzamiento de dos dados que veremos a continuación.

Modelo producto

El modelo producto de orden n pretende explicar globalmente el fenómeno aleatorio, pues viene a formalizar n ejecuciones del experimento aleatorio. Un ejemplo muy sencillo puede ser dos lanzamientos consecutivos de un dado o, equivalentemente, el lanzamiento simultáneo de dos dados. El espacio Ω^n de las posibles series de resultados se denomina **espacio muestral**.

En nuestro ejemplo tendríamos el espacio $\Omega^2 = \{(1, 1), (1, 2), \dots, (6, 6)\}$ con un total de 36 elementos denominados series o muestras aleatorias. El hecho de que las repeticiones verifiquen (i) se formaliza construyendo la probabilidad P^n sobre este espacio como producto n veces de la probabilidad original. Efectivamente, si, por ejemplo, lanzamos dos veces un dado, podemos obtener un total de 36 series o muestras aleatoria de tamaño 2 diferentes, y por pura simetría hemos de asignar a cada cual idéntica probabilidad, es decir, $1/36$. Nótese entonces que la probabilidad P^2 en el espacio muestral se obtiene de forma multiplicativa a partir de la probabilidad P en el espacio original. Por ejemplo:

$$\begin{aligned} P^2(\text{dado}[1]=5, \text{dado}[2]=3) &= P(\text{dado}[1]=5) \times P(\text{dado}[2]=3) \\ \frac{1}{36} &= \frac{1}{6} \times \frac{1}{6} \end{aligned}$$

Otro ejemplo:

$$\begin{aligned} P^2(\text{dado}[1]=\text{par}, \text{dado}[2]=\text{par}) &= P(\text{dado}[1]=\text{par}) \times P(\text{dado}[2]=\text{par}) \\ \frac{9}{36} &= \frac{3}{6} \times \frac{3}{6} \end{aligned}$$

En definitiva, al construir la probabilidad P^2 como producto de una misma probabilidad asumimos implícitamente que los sucesos relativos al resultado del primer dado son independientes de los sucesos relativos al segundo. De esta manera se formaliza la condición (i).

Otro ejemplo más: consideremos 5 lanzamientos de una moneda simétrica. El espacio original es $\Omega = \{C, X\}$, teniendo ambos sucesos elementales probabilidad $1/2$. El espacio muestral es

$$\Omega^5 = \{CCCCC, CCCCX, CCCXC, CXXXX, \dots, XXXXX\}$$

con un total de $2^5 = 32$ series o muestras aleatorias equiprobables, es decir, la probabilidad de cada uno de ellos es $(1/2)^5 = 1/32$

Repetimos que, a pesar de haber distinguido dos tipos diferentes de modelos probabilísticos no existe distinción formal entre ellos pues comparten los dos elementos esenciales: un conjunto de posibilidades y una función de probabilidad sobre el mismo. Denominamos modelo de probabilidad, a secas, a este marco teórico común. No debe pues preocuparnos si el modelo que tenemos entre manos es original o se deriva de otro como producto. De hecho y para simplificar la notación, hablaremos en todo caso de una probabilidad P , sea cual sea el tipo de espacio sobre el que se define.

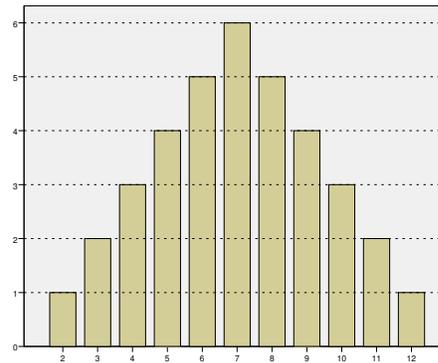
1.2. Distribución de probabilidad

En la práctica, el estudio de un determinado fenómeno aleatorio no consiste en la descripción exhaustiva de los sucesos elementales derivados del mismo, sino en el análisis de uno o varios caracteres cuantitativos considerados. La medición X de cualquier carácter sobre cada suceso elemental constituye lo que denominamos **variable aleatoria**. Por lo tanto, si nuestro estudio se centra en un determinado carácter lo que nos importa realmente es determinar su **distribución de probabilidad**, lo cual significa conocer qué valores puede tomar la variable y con qué probabilidad en cada caso. Se denomina también distribución teórica para distinguir la de la distribución de frecuencias estudiada en Estadística Descriptiva.

1.2.1. Función de probabilidad

Retomemos el ejemplo del lanzamiento de dos dados. Sabemos que en determinados juegos de azar no importa exactamente cuál ha sido el resultado de cada uno de los dados sino la suma X de ambas puntuaciones. Ése es un sencillo ejemplo de variable aleatoria, que puede tomar 11 valores diferentes, concretamente $x_1 = 2$, $x_2 = 3$, $x_3 = 4, \dots, x_{11} = 12$. Si suponemos una simetría perfecta, podemos determinar su distribución de probabilidad contabilizando el número de casos favorables a cada resultado de la variable dividido por el número total de casos que presenta el espacio, es decir, 36:

| x_i | $P(X = x_i)$ |
|-------|--------------|
| 2 | 1/36 |
| 3 | 2/36 |
| 4 | 3/36 |
| 5 | 4/36 |
| 6 | 5/36 |
| 7 | 6/36 |
| 8 | 5/36 |
| 9 | 4/36 |
| 10 | 3/36 |
| 11 | 2/36 |
| 12 | 1/36 |



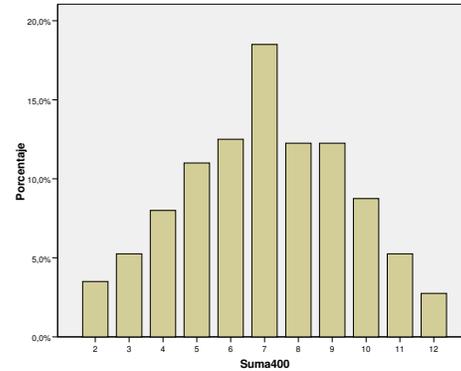
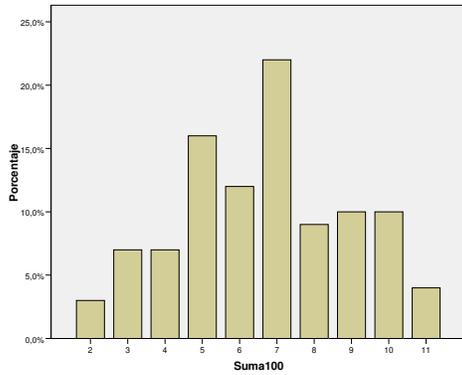
La función que asigna a cada posible valor su probabilidad se denomina **función de probabilidad** y caracteriza la distribución de la variable. Viene dada por la tabla anterior y guarda una gran similitud con la tabla de frecuencias relativas estudiada en el primer capítulo. Al igual que aquélla se ilustra mediante el denominado diagrama de barras para frecuencias relativas, ésta da lugar al diagrama que vemos a su derecha.

La diferencia entre este gráfico y el diagrama de barras estriba en que en este caso, en lugar de computarse la frecuencia relativa \hat{p}_i para cada valor x_i de una muestra de tamaño n , se computa la probabilidad p_i de que la variable X tome el valor x_i . Sin embargo, la Ley del Azar establece una clara relación entre ambas. Concretamente, si lanzamos ambos dados un total de n veces obtendremos una muestra aleatoria de tamaño n , lo cual supone n datos⁴ comprendidos entre 2 y 12 correspondientes a sendas sumas. Podemos contabilizar las correspondientes frecuencias relativas \hat{p}_i . La Ley del Azar viene a decir que si n es grande, entonces

$$(ii) \quad \boxed{\hat{p}_i \simeq p_i}$$

Efectivamente, simulamos mediante un programa estadístico 100 lanzamientos de dos dados; posteriormente, simulamos 400 lanzamientos de dos dados. Obtenemos los correspondientes diagramas de barras:

⁴En lo sucesivo cometeremos el abuso de identificar la muestra en sí con los datos que proporciona al medir la variable estudiada.



Se aprecia pues una confirmación empírica de la Ley del Azar, pues podemos observar cómo a medida que aumenta el número de ejecuciones del experimento, las proporciones de resultados favorables a cada posible valor de la variable (frecuencias relativas) se aproximan a las probabilidades teóricas. Esto viene a confirmar la aptitud del modelo matemático asumido.

Además de la función de probabilidad, debemos mencionar otra función que también caracteriza la distribución de probabilidad de una variable aleatoria X . De hecho, se denomina **función de distribución**, y se define como aquella que asigna a cada valor x la probabilidad

$$F(x) = P(X \leq x)$$

Recibe este nombre porque, a pesar de resultar menos intuitiva que la función de probabilidad y al contrario de ésta, puede definirse para cualquier tipo de variable, ya sea discreta y continua, por lo que es ideal para caracterizar cualquier distribución. La relación entre esta función y el diagrama de barras para frecuencias relativas acumuladas es idéntica a la que se da entre la función de probabilidad y el diagrama de frecuencias relativas.

1.2.2. Parámetros probabilísticos. Ley de Grandes Números

Así pues, hemos establecido ya la conexión clave entre la Estadística Descriptiva y la Probabilidad: la repetición n veces de un fenómeno aleatorio da lugar a una muestra aleatoria de n datos cuyas frecuencias relativas se van identificando progresivamente con las correspondientes probabilidades. Estamos entonces en condiciones de redefinir todos los valores típicos estudiados en Estadística Descriptiva en términos probabilísticos, en particular la media aritmética y varianza. Recordemos que definíamos en (??) la media aritmética de una muestra con valores x_1, \dots, x_k mediante

$$\bar{x} = \sum_i x_i \hat{p}_i.$$

De esta forma, se define la **Esperanza** o media de una variable aleatoria X con valores posibles x_1, \dots, x_k mediante

$$E[X] = \sum_i x_i p_i$$

Se trata pues del centro de gravedad que se obtiene ponderando los datos por su probabilidades. Este parámetro suele denotarse por la letra griega μ . En el caso del lanzamiento de dos dados es claro que $\mu = 7$. También es claro que, a medida que el número de repeticiones del experimento aumenta, el valor de la media aritmética \bar{x} se aproxima al de μ . Por ejemplo, en la muestra que

se obtiene tras lanzar 100 veces el para de dados se obtiene como media aritmética de la suma de puntuaciones $\bar{x} = 6,670$. Sin embargo, tras 400 repeticiones se obtiene $\bar{x} = 7,008$. Ello es signo del cumplimiento de la Ley del Azar. En nuestro modelo matemático esta convergencia debe verificarse necesariamente en virtud de la denominada Ley de los Grandes Números. Lo expresamos así:

$$(ii) \quad \boxed{\lim_{n \rightarrow \infty} \bar{x} = \mu}$$

De manera análoga podemos redefinir la varianza en términos de probabilidad. Recordemos que definíamos en (??) la varianza muestral mediante

$$s^2 = \sum_i (x_i - \bar{x})^2 \cdot \hat{p}_i.$$

Así pues, se define la varianza probabilística $\text{var}[X]$, que también se denota por la letra griega σ^2 , mediante

$$\text{var}[X] = \sum_i (x_i - \mu)^2 \cdot p_i$$

Su raíz cuadrada es la desviación típica probabilística y se denota por σ . Se verifica pues que, en las condiciones mencionadas, $\lim_{n \rightarrow \infty} s = \sigma$. Lo mismo puede decirse en definitiva del resto de parámetros estudiados, cosa que omitimos.

1.2.3. Ejemplo: distribución binomial

Supongamos que el color de ojos (distinguímos únicamente entre claros y oscuros) depende únicamente de una gen cuyo alelo dominante **A** determina un color oscuro y cuyo alelo recesivo **a** determina un color claro. Consideremos una pareja donde ambos individuos son hererocigóticos **Aa**. El color de los ojos de un descendiente depende de qué alelos se combinen, con lo que el número de posibilidades es 4:

$$\Omega = \{AA, Aa, aA, aa\}$$

Si asumimos la simetría en lo que respecta a este gen tanto en el proceso de meiosis como en el de la fecundación, podemos suponer que todas las posibilidades son equiprobables y que el color de ojos de un descendiente no condiciona el que pueda tener otro. La probabilidad de que un descendiente tenga ojos claros es pues 1/4. Supongamos que la pareja tiene 5 descendientes, lo cual nos conduce al complicado espacio muestral Ω^5 . No obstante, sólo estamos interesados en este caso en conocer cuántos descendientes poseerán ojos claros.

Así pues, la variable aleatoria considerada X es el número de descendientes con ojos claros, que puede tomarlos valores 0,1,2,3,4 ó 5. Basta conocer la función de probabilidad para caracterizar la distribución de esta variable. Nos preguntamos, por ejemplo, cuál es la probabilidad de tener exactamente 2 descendientes con los ojos claros. Ese suceso puede verificarse de muchas formas, por ejemplo si tenemos la secuencia o muestra aleatoria **CC000**. La probabilidad de dicha combinación puede calcularse de dos formas: dividiendo el número de casos favorables en el espacio muestral Ω^5 por el número total de casos posibles, en este caso 27/1024; más fácil es calcularla multiplicando las probabilidades de ir obteniendo cada suceso de la secuencia, es decir,

$$\frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{27}{1024}$$

Pero hemos de tener en cuenta que no es ésta la única combinación que aporta dos descendientes con ojos claros, pues cualquier alteración del orden da lugar a otra igualmente válida e igualmente probable, por la conmutatividad del producto: C0C00, 0C00C, etc. La pregunta es: ¿cuántas combinaciones con 2C entre 5 posibilidades pueden darse? La respuesta es clara si tenemos nociones básicas de combinatoria: $\binom{5}{2}$, es decir,

$$P(X = 2) = \binom{5}{2} \cdot \frac{27}{1024} = 0,26$$

Si este modelo matemático explica el fenómeno considerado, ¿qué debería suceder en la práctica? Pues que, dada una gran cantidad de parejas en esas condiciones y con 5 descendientes, aproximadamente el 26% de las mismas debe tener dos descendientes con ojos claros. Generalizando los cálculos podemos decir que

$$P(X = j) = \binom{5}{j} \cdot \left(\frac{1}{4}\right)^j \cdot \left(\frac{3}{4}\right)^{5-j}, \quad j = 0, 1, 2, 3, 4, 5$$

Hemos construido pues la distribución de probabilidad. Podemos generalizar aún más los cálculos de la siguiente forma: si una variable X contabiliza el número de veces que se verifica cierto suceso, que ocurre con una probabilidad p , tras n repeticiones independientes del experimento, la probabilidad de que X tome un valor $j = 0, 1, \dots, n$ es la siguiente:

$$P(X = j) = \binom{n}{j} \cdot p^j \cdot (1 - p)^{n-j}$$

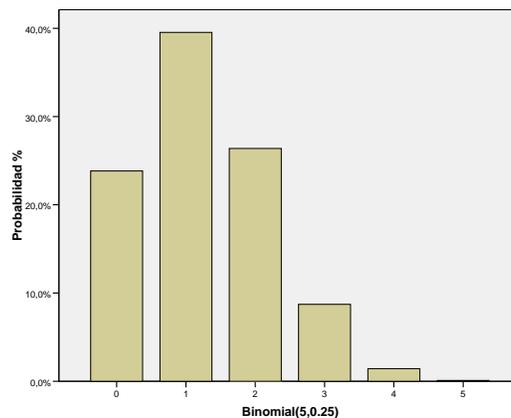
En ese caso, se dice que la variable X sigue un tipo o modelo de distribución Binomial de parámetros n y p , denotándose

$$X \sim B(n, p)$$

Si es así, tanto la media como la desviación típica pueden obtenerse directamente y sin demasiada dificultad (al menos en el caso de la media) conocidos n y p . Concretamente

$$\mu = np, \quad \sigma = \sqrt{np(1 - p)}$$

Así pues, en el ejemplo anterior puede decirse que el número de hijos de ojos claros sigue un modelo de distribución $B(5, 1/4)$. Aplicando las fórmulas anteriores obtenemos $\mu = 1,25$ y $\sigma = 0,56$. Representamos a continuación su función de probabilidad



1.2.4. Distribuciones continuas. Distribución Normal

Lo dicho hasta ahora no es válido para cualquier tipo de variable aleatoria sino sólo para aquellas que pueden tomar una cantidad finita o al menos enumerable (ordenable) de valores, dando lugar a lo que denominamos **distribuciones discretas**. Por contra, las variables que pueden tomar cualquier valor en un intervalo (nótese que éstos no pueden enumerarse) darán lugar a las **distribuciones continuas**.

Como ejemplo podemos considerar un disco que presenta una marca en un punto de su perímetro y que gira en un viejo tocadiscos. Nos preguntamos en qué ángulo exacto de la circunferencia (medido en radianes) quedará la marca cuando el disco se detenga. La medida de dicho ángulo es una variable aleatoria X con valores en el intervalo $[0, 2\pi)$. Podemos calcular diversas probabilidades por simetría: por ejemplo, la probabilidad de que la marca quede en el primer cuadrante es

$$\frac{\pi/2}{2\pi} = \frac{1}{4}$$

Sin embargo, podemos razonar fácilmente que la probabilidad de que X tome un valor exacto dentro del intervalo considerado es tan pequeña como se quiera, es decir, nula, lo cual podría resultar paradójico si se piensa que la marca debe detenerse en algún sitio concreto. Lo cierto es que nosotros no apreciamos un punto de parada sino un intervalo, que será más pequeño cuanto mayor sea nuestra precisión, y ese intervalo puede tener una probabilidad muy escasa pero nunca será nula. Esta paradoja es consecuencia de una pequeña pero insalvable discordancia entre la realidad percibida por los sentidos y el modelo matemático que la idealiza.

A la hora de formalizar este tipo de situaciones nos encontramos pues con los problemas inherentes a las mediciones sobre un **continuo**, por lo que se precisa una cierta familiaridad con las técnicas de integración y el lenguaje infinitesimal. Aquí no tiene sentido hablar de la función de probabilidad y los parámetros de la distribución no pueden definirse como vimos anteriormente. No obstante, sí que podemos establecer relaciones entre un incremento de la variable Δx y el correspondiente incremento de la probabilidad ΔP . En este caso concreto y por tratarse de una simetría pura, la relación entre ambos, $\Delta P/\Delta x$, es constante y vale $1/2\pi$. Sea o no constante, lo que nos interesa para medir probabilidades es el límite del cociente incremental en cada punto x del intervalo. La función definida en el lenguaje infinitesimal mediante

$$f(x) = \frac{dP}{dx}$$

se denomina **función de densidad**. En ese caso, tenemos que

$$dP = f(x) dx$$

Así pues, la probabilidad de que X pertenezca a un intervalo $[x_1, x_2]$ se obtiene integrando

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} dP = \int_{x_1}^{x_2} f(x) dx$$

La función de densidad caracteriza la distribución de la variable pues nos permite obtener la probabilidad de cualquier intervalo mediante el cálculo del área subyacente a la misma entre sus límites. Como dijimos en el caso discreto, se puede definir también la función de distribución mediante

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

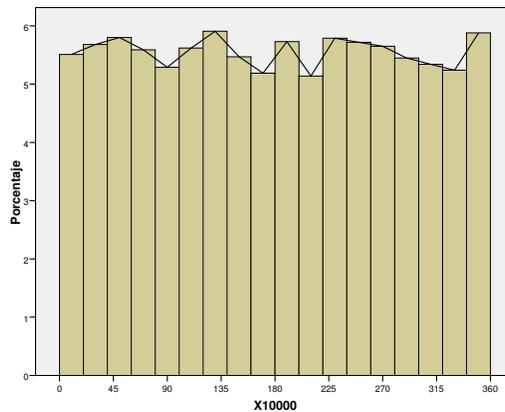
Caracteriza igualmente el modelo de probabilidad. Por otra parte, la densidad permite calcular los diferentes parámetros probabilísticos. Por ejemplo, si la media se definía mediante $\mu = \sum_i x_i p_i$ en el caso discreto, en el caso continuo se define mediante

$$\mu = \int X dP = \int x f(x) dx$$

De manera análoga puede definirse la varianza.

En el ejemplo del disco, la función de densidad será la función constante $1/2\pi$ definida entre 0 y 2π y la media de la distribución es $\pi = 180$ grados. También es la mediana. Cuando la función de densidad (o la de probabilidad en el caso discreto) es constante se dice entonces que la distribución es **uniforme**.

Si la función de probabilidad de una distribución discreta guardaba una estrecha relación con el diagrama de barras de frecuencias relativas, la de densidad se vincula claramente al histograma. Efectivamente, simulamos mediante un programa estadístico el fenómeno anterior (midiendo el ángulo en grados) en 10.000 ocasiones y representemos mediante un histograma de frecuencias relativas los 10.000 valores entre 0 y 360 grados obtenidos.



Podemos observar que los diferentes intervalos considerados aportan rectángulos de áreas muy similares. Recordemos que las áreas de éstos son proporcionales a las frecuencias relativas de cada intervalo y éstas, por la Ley del Azar, deben ser parecidas a las que determina la distribución uniforme. Este efecto es más acusado cuantas más veces se repite el experimento. La media aritmética de las 10.000 mediciones es, por cierto, $\bar{x} = 179,88$ grados, muy cercana a $\mu = 180$.

El modelo de distribución continua más importante, por razones que veremos a continuación, es sin duda la **distribución normal**. Se dice que una variable aleatoria X sigue un modelo de distribución normal de parámetros μ y σ cuando su función de densidad es la denominada curva normal:

$$f(x) = (\sigma\sqrt{2\pi})^{-1} \cdot e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$$

Se denota $X \sim N(\mu, \sigma)$. En ese caso, puede demostrarse que μ y σ son, respectivamente, su media y desviación típica, de ahí la notación utilizada. Las probabilidades de los distintos intervalos se obtienen calculando las áreas subyacentes a la curva. De esta forma, puede comprobarse, por ejemplo, que la probabilidad de que la variable esté en el intervalo $(\mu - \sigma, \mu + \sigma)$ es 0.68, y en el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$ es 0.95. ¿Nos suena esto de algo?

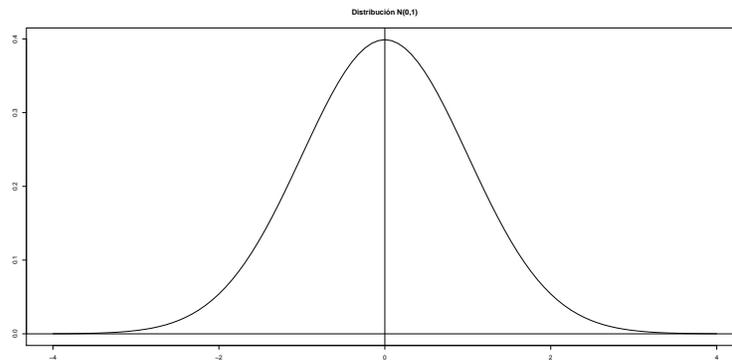
Desde el punto de vista gráfico, la media μ nos da el eje de simetría y la desviación típica indica, por supuesto, el grado de condensación. El área total subyacente a la curva es 1, como corresponde a una función de densidad. Se verifica en general, por las propiedades de la media y la desviación típica, que si X es una variable aleatoria de media μ y desviación típica σ , la variable

$$Z = \frac{X - \mu}{\sigma}$$

posee media 0 y desviación típica 1. Este proceso al que con frecuencia se someten las variables, sigan o no un modelo de distribución normal, se denomina **tipificación** o estandarización. Puede demostrarse que, si X sigue un modelo de distribución normal, también lo sigue cualquier transformación afín de la misma y en particular su tipificación Z . Por lo tanto,

$$Z \sim N(0, 1)$$

Este último modelo de distribución se denomina normal estándar.



La tipificación nos permite, entre otras cosas, calcular probabilidades correspondientes a cualquier normal a partir de la distribución normal estándar.

1.2.5. Distribuciones muestrales

En la primera sección del capítulo distinguimos entre el modelo de probabilidad asociado a una única ejecución del experimento aleatorio y el asociado a n ejecuciones del mismo. El segundo se denomina producto y está compuesto por el espacio de las muestras aleatorias de tamaño n y la probabilidad producto que rige el grado de verosimilitud de las mismas.

Se denomina **variable aleatoria muestral** a cualquier variable sobre el espacio producto de las muestras aleatorias. De la distribución de dicha variable decimos que es una distribución muestral. Se trata pues de un caso particular del concepto de distribución estudiado anteriormente. Concretamente, es una variable muestral la suma de las puntuaciones obtenidas por los dos dados, pues cada lanzamiento de dos dados puede considerarse una muestra aleatoria de tamaño $n = 2$ del fenómeno aleatorio consistente en el lanzamiento de uno. Su distribución, estudiada ya con detalle, es por lo tanto una distribución muestral. Si dividimos por 2 la suma de las puntuaciones estaremos hablando de la media aritmética de las mismas.

Por lo tanto y en general, la media aritmética valorada en el espacio de las muestras aleatorias de tamaño n es una variable muestral. Queremos decir con esto que no la entendemos como un simple número sino que puede variar de una muestra aleatoria a otra. Su distribución muestral determina entonces qué valores puede tomar y con qué probabilidades. Lo mismo

puede decirse de la varianza y de todos los valores típicos estudiados en los capítulos 1 y 2. Así pues, desde este punto de vista más amplio, los parámetros descriptivos pueden entenderse como variables muestrales con sus correspondientes distribuciones, en cuyo caso se denotarán mediante las letras mayúsculas \bar{X} , S^2 , \bar{X} , etc. Una vez obtenida una muestra aleatoria concreta, la variable muestral aportará el número correspondiente que se denota en minúscula: su media \bar{x} , su desviación típica s , su coeficiente de asimetría, su coeficiente de correlación si se trata de dos variables, etc.

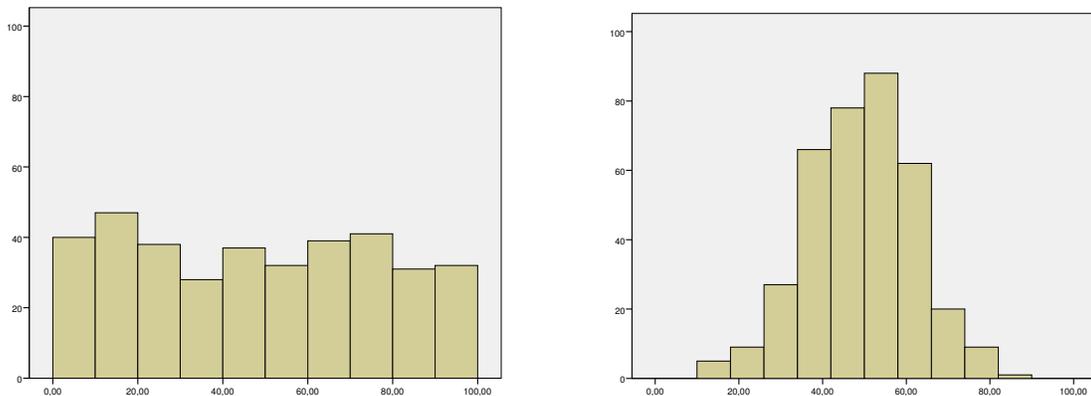
Dada una variable X de media μ y varianza σ^2 , puede demostrarse que la esperanza y varianza de la media muestral definida sobre las muestrales aleatorias de tamaño n son las siguientes:

$$\mathbb{E}[\bar{X}] = \mu, \quad \text{var}[\bar{X}] = \frac{\sigma^2}{n}$$

Es decir, el valor medio esperado para \bar{X} es la propia media probabilística de X pero su varianza es inversamente proporcional al tamaño muestral considerado. Por lo tanto, dado que la varianza expresa el grado de dispersión de los valores respecto a su media, se verifica que, si n es suficientemente grande, la probabilidad de que la media aritmética de una muestra aleatoria de tamaño n se aleje de la media probabilística será muy pequeña. Esto se parece desde luego a la condición

$$(ii) \quad \boxed{\lim_{n \rightarrow \infty} \bar{x} = \mu}$$

Veamos otro ejemplo de distribución muestral: se estudia la media aritmética de cinco números entre 1 y 99 extraídos mediante un sorteo de lotería con reemplazamiento. Estamos pues hablando del modelo producto que corresponde a $n = 5$ repeticiones del fenómeno aleatorio consistente en extraer una bola entre 99 posibles. El valor de la bola sigue una distribución discreta uniforme con media $\mu = 50$ y desviación típica $\sigma = 28,6$. Así pues, la media muestral tendrá media 50 y desviación típica $\sigma/\sqrt{5} = 12,8$. Para ilustrar la diferencia entre ambas distribuciones vamos a imaginar que el sorteo se repite todos los días de un año, es decir, 365 veces. Vamos a anotar, por un lado, el resultado de la primera bola extraída cada día, que decimos sigue una distribución uniforme. Por otro lado, anotamos la media aritmética de las cinco bolas extraídas cada día. Los resultados simulados mediante un programa estadístico dan lugar a los siguientes histogramas:

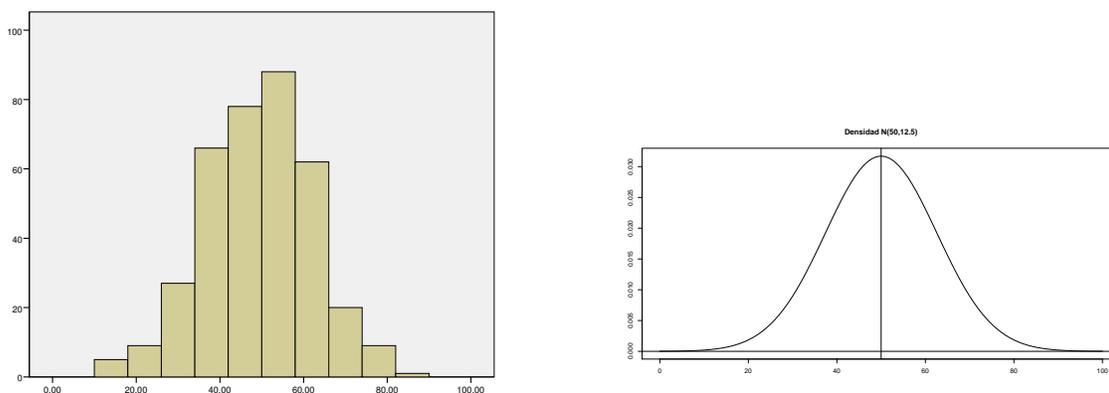


Comprobamos cómo efectivamente los datos correspondientes a la primera bola ofrecen una histograma relativamente plano, como corresponde a una distribución uniforme, cosa que realmente ocurre con las bolas restantes, pero no con la media aritmética de las cinco bolas, pues

ésta se distribuye también de manera simétrica en torno a la media 50, pero más concentradamente. La explicación heurística puede ser la siguiente: en primer lugar, la distribución ha de ser simétrica respecto a 50, pues, por la simetría del fenómeno, nada nos hace pensar que los números mayores que 50 son más probables que los menores; en segundo lugar, el hecho de que se condensan más en torno a 50 se debe a que todas las posibles series o muestras aleatorias son equiprobables, pero la mayoría aporta una media aritmética próxima a 50, pues para obtener una media aritmética extrema es necesario que todas las bolas lo sean. No queremos decir que la serie o muestra aleatoria (1, 1, 1, 1, 1) sea menos probable que la serie (49, 51, 47, 62, 36). Lo que sucede es que, por pura combinatoria, son más numerosas las series cuya media aritmética se aproxima a 50. Si se permite la expresión, los caminos al centro son más variados que los caminos a los extremos.

1.2.6. Teorema Central del Límite

En el histograma de la derecha del ejemplo anterior se perfila una curva que a estas alturas debe ser ya familiar:



Se trata en efecto de la denominada **curva normal**, concretamente hablamos de la curva

$$N(50, 12.8)$$

que es la que corresponde según al media y varianza de \bar{X} . Es decir, la distribución de la media aritmética de las 5 bolas se aproxima a una distribución continua normal de media 50 y desviación típica 12,8. Esta aproximación a la distribución normal es más precisa cuanto mayor sea el tamaño de la muestras aleatorias. Realmente, puede demostrarse, y en eso consiste en Teorema Central del Límite, que esto sucede con carácter general. Es decir, que para muestras aleatorias suficientemente grandes, la media muestral de una variable X con media μ y varianza σ^2 sigue aproximadamente un modelo de distribución $N(\mu, \sigma/\sqrt{n})$. Tipificando obtenemos pues

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{\text{aprox}}{\sim} N(0, 1)$$

Si el tamaño de muestra es grande, dado que $\lim_{n \rightarrow \infty} s = \sigma$, podemos sustituir en la anterior expresión la desviación típica probabilística σ por la muestral S , es decir,

$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{\text{aprox}}{\sim} N(0, 1)} \quad (1.1)$$

Puede demostrarse también que si la distribución original de la variable es normal, la distribución muestral de la media aritmética será también exactamente normal. Por lo tanto, al tipificar obtendremos una $N(0, 1)$ exacta. Si sustituimos entonces σ por la desviación típica muestral S obtendremos una distribución muy similar que comentaremos a continuación: la distribución t-Student.

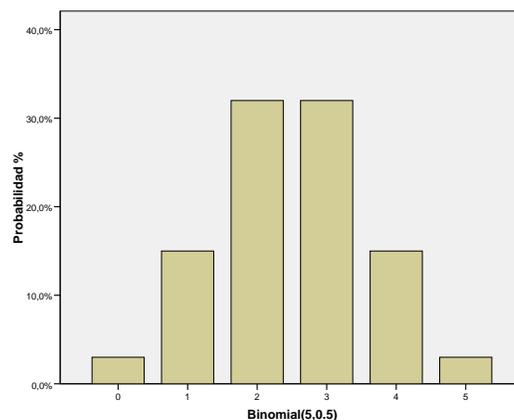
El resultado anterior, importantísimo, otorgará a la distribución normal un papel **central** en la Inferencia Estadística. Además, podría explicar por qué se observa con cierta frecuencia en la naturaleza. Cabe conjeturar, como apuntábamos en el primer capítulo, que cuando una variable aleatoria sigue una distribución aproximadamente normal se trata internamente del resultado de sumar una serie de variables o factores independientes. El caso es que esta distribución fue ya caracterizada por Gauss y Laplace al estudiar una variable que puede considerarse aleatoria: el error en la medición de parámetros astronómicos. De ahí que reciba comúnmente el nombre de campana de Gauss.

El tamaño de muestra n requerido para que la distribución de la media muestral se aproxime satisfactoriamente al modelo normal depende de la distribución original de la variable y, especialmente, de su sesgo. De hecho, cuando la distribución es simétrica, como en el caso del ejemplo, se consigue la aproximación aún con muestras pequeñas. Sin embargo, cuanto mayor es la asimetría más costoso es conseguir que la media muestral se ajuste a un modelo normal. No existe pues una cota universal para el valor de n , aunque con frecuencia se conviene que con $n = 30$ no debemos tener problemas. Otros estadísticos más conservadores exigen muestras aleatorias de al menos 60 datos para tener garantías. Lo más razonable es observar previamente el histograma de la muestra y el coeficiente de asimetría g_1 .

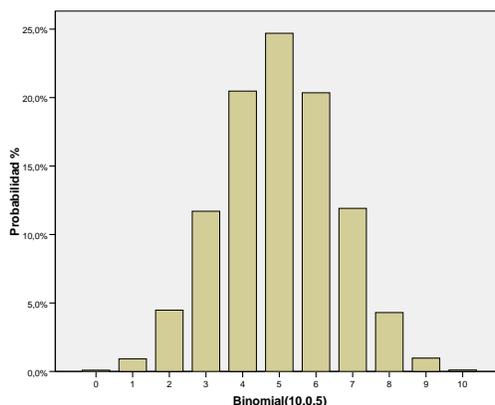
Sin ir más lejos, si una variable X sigue una distribución binomial $X \sim B(n, p)$, es decir, si recoge la suma de resultados favorables a un suceso con probabilidad p tras n ejecuciones del experimento aleatorio, la variable $\frac{1}{n}X$ recoge la media aritmética de una muestra aleatoria de tamaño n para la variable W que asigna un 1 si el resultado es favorable y un 0 si no lo es. En consecuencia, si n es suficientemente grande $\frac{1}{n}X$ seguirá aproximadamente una modelo de distribución normal y, por lo tanto, también será normal X . Dado que su media es np y su varianza $np(1 - p)$, se verifica entonces

$$B(n, p) \stackrel{\text{aprox}}{\approx} N(np, \sqrt{np(1 - p)})$$

El tamaño n requerido para que esta aproximación sea satisfactoria depende, según hemos dicho de la simetría de W y, en definitiva, de p . De hecho, para p próximo a $1/2$ se obtiene una distribución de W muy simétrica y, por lo tanto, una rápida convergencia. Tal es el caso de una distribución $B(5, 1/2)$, que se parece a la distribución $N(0.25, 1.11)$.

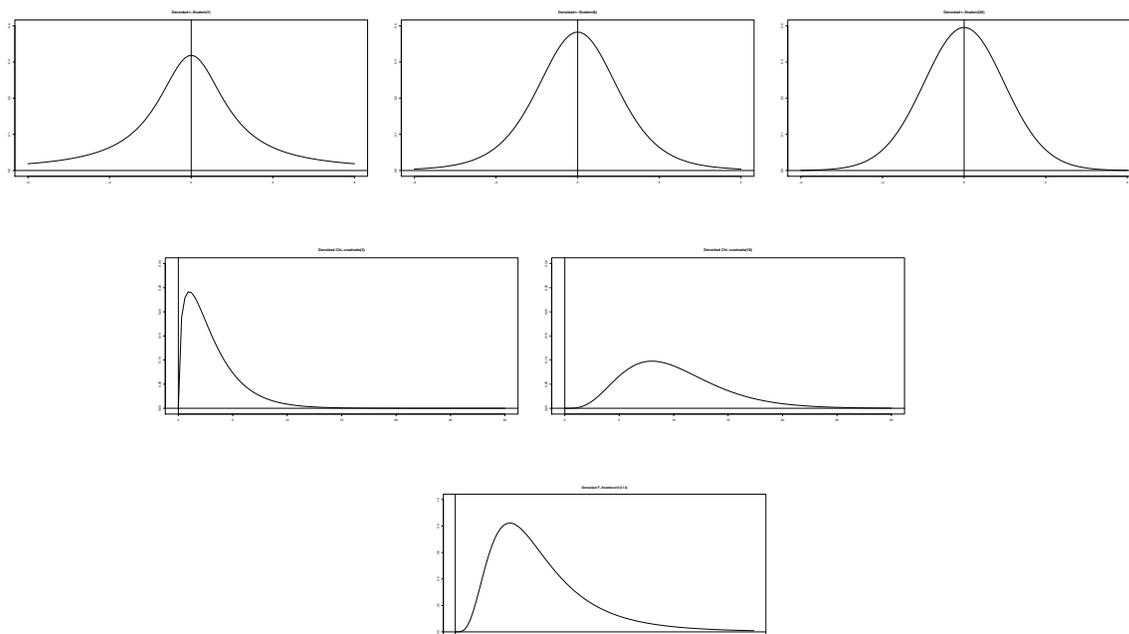


Veamos qué sucede con $B(10, 1/2)$, que debe parecerse a $N(5, 1.58)$.



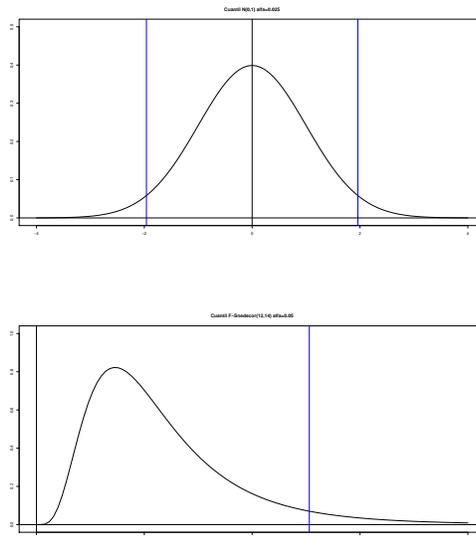
Si queremos aproximar probabilidades de la distribución discreta $B(10, 1/2)$ mediante la distribución continua $N(5, 1.58)$ parece razonable identificar cada valor entero de 0 a 10 con el intervalo de longitud 1 centrado en dicho valor. Es decir, la probabilidad que corresponde, por el ejemplo, al valor 3 según el modelo $B(10, 1/2)$ debe parecerse a la que corresponde al intervalo $(2.5, 3.5)$ según el modelo $N(5, 1.58)$.

Podemos mencionar otras distribuciones continuas que se derivan de la distribución normal: primeramente, la distribución t-Student, que depende de un parámetro entero denominado grado de libertad y es parecida a la distribución $N(0, 1)$. De hecho, a partir de un grado de libertad 30 se consideran idénticas. Segundo, la distribución χ^2 , que depende también de un grado de libertad. Por último, mencionamos la distribución F-Snedecor, que depende de dos grados de libertad. Para hacernos una idea de las distribuciones de las que hablamos mostramos a continuación las funciones de densidad de las t-Student con 1, 6 y 30 grados de libertad, de las χ^2 con 4 y 10 grados de libertad, y de la F-Snedecor con (12,14) grados de libertad:



Estas distribuciones aparecerán en la Inferencia Estadísticas como consecuencia de ciertas operaciones a las que se someterán los datos: así, la distribución t-Student, surge cuando se tipifica una variable normal pero sustituyendo su desviación típica probabilística por la muestral, obteniéndose por lo tanto una distribución similar a la $N(0, 1)$; la distribución χ^2 cuadrado se obtiene a partir de la suma de cuadrados de normales. Recordemos que la suma de cuadrados es la forma habitual de medir errores en Estadística, de ahí que esta distribución esté precisamente asociada a la medición de diversos tipos de errores en sentido amplio. Concretamente, la varianza muestral sigue, salvo una constante, un modelo de distribución $\chi^2(n - 1)$ cuando la distribución original de la variable es normal. Por último, la distribución F-Snedecor surge de la necesidad de dividir, es decir, comparar, errores o varianzas en sentido amplio, es decir, distribuciones χ^2 . Sus grados de libertad se corresponden con los de las χ^2 del numerador y denominador.

Nos interesa especialmente conocer los cuantiles de estas distribuciones así como los de la $N(0, 1)$. Nos referimos a un concepto estudiado ya en Estadística Descriptiva. El cuantil α de una distribución es el valor que deja una probabilidad α a su derecha y $1 - \alpha$ a su izquierda. El caso más importante es el que corresponde a $\alpha = 0,005$. En el caso de las distribuciones simétricas, como $N(0, 1)$ y t-Student, puede ser más interesante el caso 0,025 pues, entre dicho cuantil y su opuesto queda comprendida una probabilidad del 95 %. Mostramos a continuación los cuantiles $z_{0,025} = 1,96$ y $F_{0,005}(12, 14) = 2,53$, correspondientes a las distribuciones $N(0,1)$ y F-Snedecor(12,14). Al término del manual se muestran unas tablas que podemos consultar para encontrar cuantiles de estos tipos d distribuciones.



1.3. Población, Inferencia y Probabilidad

Todo lo dicho hasta ahora en el presente capítulo resultará sin duda apasionante a un ludópata: dados, ruletas, loterías... No obstante, debemos preguntarnos qué utilidad práctica puede tener este estudio para el lector interesado en las Ciencias de la Salud, que es precisamente a quien va dirigido este breve manual. Así pues hemos llegado al punto clave del capítulo y posiblemente de la materia. Es el momento de entender cuál es la conexión entre el Cálculo de Probabilidades, dedicado al análisis de los fenómenos aleatorios y la Estadística entendida

como tratamiento de la Información relativa a poblaciones y variables. Describiremos brevemente cómo interviene en los dos problemas fundamentales de la Inferencia Estadística que abordaremos en el próximo capítulo: Estimación y Contraste de Hipótesis.

1.3.1. Probabilidad y Estimación

En primer lugar, en la introducción definimos Población en sentido amplio como el objeto de nuestro estudio. Aunque suponga una excesiva simplificación, hemos de reconocer que en el caso de las Ciencias de la Salud prevalece la acepción común del término como colectivo de individuos, ya sean personas en general, un colectivo de pacientes, animales o plantas de cierta especie, semillas o espermatozoides. El estudio consistirá concretamente en la descripción de una o varias variables. Por lo tanto, si tuviéramos acceso a las mediciones que aportan o aportaría la población Ω completa, es decir un censo, el estudio se restringiría a lo que hemos denominado Estadística Descriptiva y no se precisaría el concurso del Cálculo de Probabilidades.

Sin embargo y por desgracia, el conocimiento de los valores de toda la población Ω es poco menos que utópico, por lo que deben ser estimadas. En la práctica, aspiramos a estudiar los datos de una muestra de tamaño n extraída de dicha población, la cual se somete a las técnicas propias de la Estadística Descriptiva. La pregunta es ¿en qué medida podemos generalizar o inferir conclusiones relativas a la población Ω a partir de la descripción de una muestra de la misma? Pues resulta que, si la muestra es aleatoria, estamos en condiciones de hacerlo. ¿Qué quiere decir que la muestra sea aleatoria? Pues que los individuos que la componen hayan sido seleccionados mediante un fenómeno aleatorio equivalente a una lotería. Veremos cómo en esas condiciones la descripción de la muestra aporta conclusiones muy concretas respecto a la población total pero que vendrán expresadas lógicamente en términos probabilísticos.

Así, por ejemplo, si estamos estudiando la incidencia de una cualidad C , por ejemplo una enfermedad, que se da en cierta población Ω en una proporción p que queremos determinar, al escoger una muestra aleatoria de tamaño n , ¿cómo calcular la probabilidad de que cada individuo de la misma presente dicha cualidad? Teniendo en cuenta que todos los individuos son sucesos elementales equiprobables del sorteo, debe calcularse dividiendo el número de casos favorables por el número de casos posibles, es decir, el número de individuos de la población que presenta la cualidad C entre el número total e individuos de la población, y eso es precisamente la proporción p . Podríamos denotar $p = P(C)$. Es decir, identificamos proporción en la población con probabilidad en el sorteo de la muestra.

Siguiendo ese mismo razonamiento, si estudiamos una variable cuantitativa X , la media aritmética de la población, que se obtiene como suma de los valores que toma X en la misma ponderados por las frecuencias relativas o proporciones poblacionales (??), coincide con la media probabilística μ correspondiente a la medición de X sobre un individuo seleccionado por sorteo. Lo mismo podríamos decir de la varianza y de cualquier valor típico. Así pues, en este contexto identificamos los parámetros poblacionales con los probabilísticos.

El fenómeno aleatorio que realmente nos interesa no es el sorteo en sí sino la repetición n veces del mismo. De esta forma, la muestra aleatoria de tamaño n es un elemento del modelo aleatorio producto asociado. Los parámetros descriptivos de la muestra, como la media aritmética \bar{X} , la varianza S^2 , las distintas, proporciones \hat{P}_i , etc., no son sino variables muestrales con sus correspondientes distribuciones. Recordemos que la muestra a estudiar es contingente, es decir, ha sido seleccionada de igual forma que podría haber sido seleccionada cualquier otra. De hecho todas son equiprobables. De ahí que la media aritmética y demás parámetros descriptivos deban considerarse variables aleatorias con sus correspondientes distribuciones.

¿Y de qué sirve en definitiva que las muestras sean aleatorias? Al satisfacerse la Ley de Azar (ii), debe verificarse una aproximación de los parámetros muestrales a los correspondientes poblaciones o probabilísticos. Es decir,

$$\hat{p}_i \longrightarrow P_i$$

$$\bar{x} \longrightarrow \mu$$

$$s \longrightarrow \sigma$$

$$r^2 \longrightarrow \rho^2$$

Etcétera. Los parámetros muestrales serán pues estimaciones de los análogos poblacionales, y la aproximación a éstos será tanto mejor cuanto mayor sea el tamaño de la muestra. Pero no nos conformaremos con vagas expresiones al respecto. Por ejemplo, veremos en el próximo capítulo cómo el resultado (1.1) de la sección anterior puede servir para acotar de manera probable el error cometido en la aproximación de \bar{x} a μ .

En definitiva, en el contexto de las Ciencias de la Salud debemos inclinarnos a interpretar el concepto de **probabilidad** no como una medida difusa y universal del grado de fe que tenemos en que algo ocurra, sino como una **proporción** respecto al total de la **población**.

1.3.2. Probabilidad y Contraste de Hipótesis

El problema de contraste de hipótesis consiste, como veremos con más detenimiento en el capítulo siguiente, en determinar la validez de un modelo teórico a la hora de explicar una cierta cantidad de observaciones experimentales. Haremos uso del lenguaje probabilístico para plantear y resolver el problema. Efectivamente, este tipo de problemas se afrontará identificando ese modelo teórico con un fenómeno aleatorio, es decir, con un modelo probabilístico ideal y explícito. Así, en el ejemplo 9 del capítulo siguiente, contrastaremos si en una determinada localidad actúan agentes ambientales capaces de influir en el sexo de la población. De no ser así, cabría pensar por simetría, dado que por cada espermatozoide portador del cromosoma X debe existir otro por tanto el cromosoma Y, que el sexo de una serie de n nuevos individuos puede considerarse el resultado de un fenómeno aleatorio equivalente al lanzamiento n veces de una moneda simétrica. Ésa será pues la hipótesis inicial a contrastar, de manera que debemos determinar si las observaciones obtenidas son o no probables según este modelo probabilístico concreto y, en función de eso, tomar una decisión razonable.

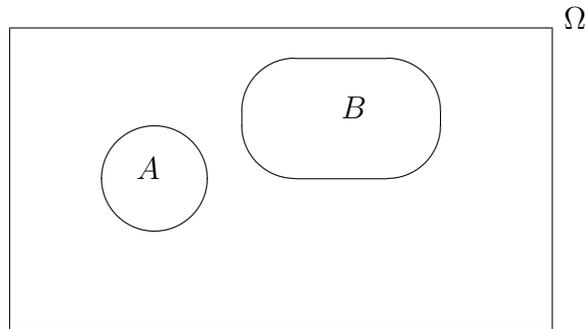
En otras ocasiones, como en el ejemplo 8 del capítulo siguiente, en el que se contrasta si la media poblacional de cierta variable presenta cierto valor concreto μ_0 , no cabe pensar en otro fenómeno aleatorio que el propio muestreo (lotería), de manera que la hipótesis inicial se identifica con un valor medio μ_0 para el modelo probabilístico asociado. En definitiva, los fenómenos aleatorios que pueden interesarnos son la lotería, pues es teóricamente el procedimiento de selección de la muestra, y cualquier otro que pretenda identificar aproximadamente una hipótesis inicial a contrastar.

1.4. Cuestiones propuestas

1. Establecer un paralelismo entre todos los conceptos estudiados en Estadística Descriptiva y los estudiados en este capítulo.

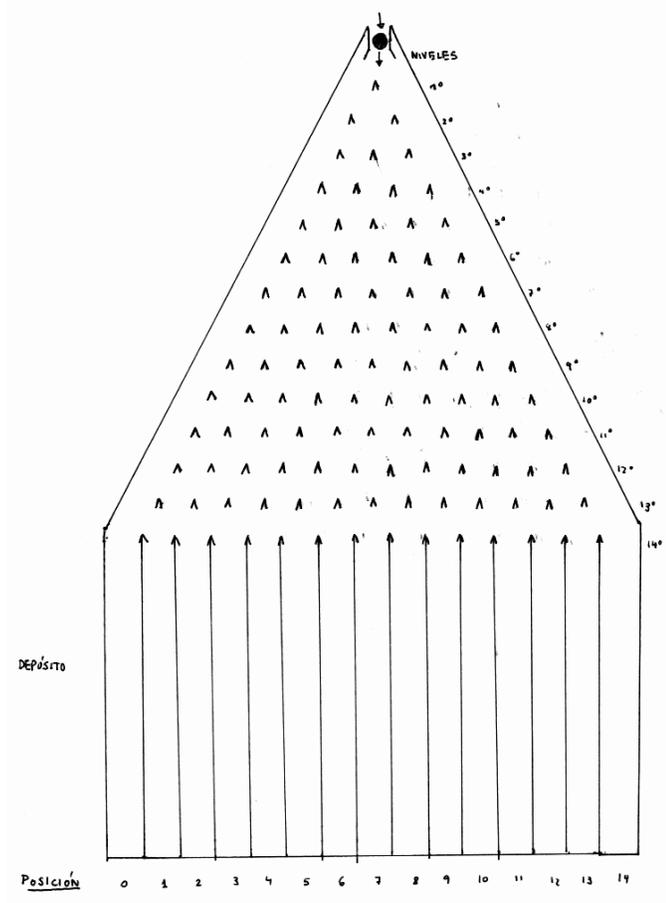
2. Discutir si la medición de colesterolemia puede considerarse un fenómeno aleatorio.
3. Sospechamos que un dado está trucado. ¿Qué podemos hacer para contrastarlo?
4. Se cuenta con un bombo con 99.999 bolas iguales numeradas del 1 al 99.999. Se realiza el experimento consistente en hacer girar el bombo y extraer una bola cualquiera. Comenta las siguientes afirmaciones:
 - No sabemos nada acerca del resultado del experimento.
 - Si el resultado de una extracción es superior a 50.000, el resultado de la siguiente extracción (después del reemplazamiento) ha de ser inferior a 50.000.
 - A medida que vamos extrayendo bolas (con reemplazamiento), los valores de éstas se van aproximando a 50.000.
5. Calcula la probabilidad de obtener dos caras tras cinco lanzamientos de una moneda no trucada.
6. Si una variable aleatoria discreta X sigue una distribución de probabilidad del tipo $B(16, 0.5)$, calcular la probabilidad de que X sea mayor a igual que 14. Lo mismo si el modelo es $B(16, 0.8)$.
7. Se estudia una cierta variable bioquímica X (medida en gramos) sobre una determinada población. Se conoce que el valor de la media es $\mu = 1$ y el de la varianza es $\sigma^2 = 0$. ¿Cómo se interpretan estos datos?
8. Consideremos una variable $Z \sim N(0, 1)$. Calcula mediante las tablas:
 - $P(Z < 0,5)$
 - $P(Z > 0,5)$
 - $P(Z < -0,5)$
 - $P(0,5 < Z < 1)$
 - $P(-1 < Z < 0,5)$
9. Se estudia determinado carácter cuantitativo sobre una población. La correspondiente variable X se distribuye aproximadamente según un modelo Normal, siendo su media 20 y desviación típica 5.
 - Calcula la proporción aproximada de individuos cuyo valor de la variable es inferior a 31.2.
 - Calcula la proporción aproximada de individuos cuyo valor de la variable está comprendido entre 30 y 20.
 - Calcula la proporción aproximada de individuos cuyo valor de la variable es superior a 50.
10. Se tiene dos variables $X \sim N(12, 4)$ e $Y \sim N(12, 2)$. Razonar (sin necesidad de cálculos) si son verdaderas o falsas cada una de las siguientes afirmaciones:
 - $P(X > 11) > P(Y > 11)$

- $P(X \leq 12) = P(Y \geq 12)$
11. Si $Z \sim N(0, 1)$, calcula el número positivo z tal que $P(-z \leq X \leq z) = 0,95$. Entender que ese número es $z_{0,05/2}$.
12. Se tiene un procedimiento para medir glucosa en la sangre. Se comprueba que el método no es exacto, pues el contenido de glucosa medido en una determinada porción de sangre, según el procedimiento, difiere de su verdadero valor. Más aún, distintas mediciones de una misma porción de sangre aportaron distintos resultados, pero verificando la Ley del Azar. Se comprobó además que los distintos valores obtenidos se agrupan formando una Campana de Gauss, por lo que podemos considerar que la variable X =error cometido tras medir glucosa en sangre sigue un modelo de distribución Normal.
- ¿Que hemos de hacer para averiguar lo posible acerca de la media y la desviación típica de dicha variable?
 - Supongamos conocidos los valores de μ y σ . Ordena por orden de conveniencia los siguientes casos:
 - $X \sim N(3, 1)$
 - $X \sim N(0, 4)$
 - $X \sim N(3, 4)$
 - $X \sim N(0, 1)$
13. Consideremos cierta variable bioquímica X que suponemos aleatoria. Se desea saber si se ajusta a un modelo de distribución Normal. Es conocido que el 50% de los valores obtenidos experimentalmente es a 8, que el 20% es superior a 10 y que el 2% son inferiores a 6.
- ¿Contradicen estos datos la normalidad de la variable?
 - ¿Puedes decir algo acerca del coeficiente de asimetría g_1 ?
14. Calcula la esperanza y la varianza de la variable aleatoria X =resultado obtenido tras lanzar un dado no trucado.
15. Calcula la probabilidad de obtener más de 6 caras tras 10 lanzamientos de una moneda no trucada. Calcula la probabilidad aproximada de obtener más de 60 caras tras 100 lanzamientos de una moneda no trucada.
16. Podemos considerar el espacio aleatorio Ω como una sección del plano, y cada suceso, por lo tanto, como un subconjunto de dicha sección. La probabilidad de cada suceso puede interpretarse entonces como la proporción de área que éste ocupa. De esta forma, el diagrama presenta dos sucesos disjuntos. La probabilidad (área) de la unión sera por tanto la suma de las probabilidades (áreas) de cada uno:



¿Cómo se expresarían gráficamente dos sucesos independientes? Recueda que A y B son independientes cuando $P(A \cap B) = P(A) \times P(B)$.

17. Describe la función de probabilidad de la distribución $B(6, 0.8)$.
18. Considera el ingenio que representa la figura:



¿Cuál es la probabilidad de que una bola introducida en la abertura superior termine en la posición 7 del depósito? Si se introducen un total de 200 bolas, que figura se formará en el depósito, una vez hayan caído todas?

19. En numerosas ocasiones hemos afirmado que, si una variable X sigue una distribución normal de media μ y desviación típica σ , la probabilidad de que un valor se encuentre entre $\mu - \sigma$ y $\mu + \sigma$ es, aprox., del 68%. ¿Se trata de un hecho experimental o existe alguna forma de probarlo?

20. Cuando se habla de probabilidad, debemos estar necesariamente refiriéndonos a un fenómeno aleatorio. Por ejemplo; podemos hablar de la probabilidad de obtener cara tras lanzar una moneda, la probabilidad de que la suma de las puntuaciones de dos dados sea 7, etc. Sin embargo, con frecuencia se utilizan expresiones como la siguiente: **la probabilidad de que un individuo varón mayor de 18 años mida más de 1,74m es del 50%**. ¿A qué fenómeno aleatorio nos estamos refiriendo en este caso?

Capítulo 2

Introducción a la Inferencia Estadística

La Estadística, como su propio nombre parece indicar, se concibe en principio para el tratamiento de la información relativa a grandes poblaciones, entendidas éstas como colectivos de individuos. Si bien el término de población puede considerarse hoy en día mucho más amplio, la acepción clásica del mismo es la que prevalece en las Ciencias de la Salud. En todo caso, sucede en la mayoría de las ocasiones que dicha población, entiéndase como se entienda, es demasiado grande o compleja, inabarcable, por lo que su descripción exhaustiva es infactible.

¿Cómo podemos paliar esta incapacidad? Pues, según hemos visto en el capítulo anterior, seleccionando aleatoriamente n individuos de la población, los cuales constituirán una muestra aleatoria de ésta. Nos permitimos el abuso de denominar igualmente por muestra a los datos que aportan esos individuos. Dichos datos serán sometidos a las técnicas descriptivas consideradas en los capítulos 1 y 2 para, posteriormente y en virtud de los métodos que estudiaremos a partir de ahora, inferir o generalizar conclusiones relativas a la población total. Esta nueva fase del estudio se denomina Inferencia Estadística y exige, como hemos dicho, que los componentes de la muestra hayan sido escogidos aleatoriamente. Sólo en esas condiciones estamos capacitados para extrapolar los resultados, pero siempre en términos probabilísticos.

El proceso de selección aleatoria de los integrantes de la muestra se denomina **muestreo aleatorio**. Existen realmente diferentes tipos de muestreos aleatorios, pero nosotros consideraremos únicamente el muestreo aleatorio simple. En el caso de una población en el sentido clásico del término, el muestreo aleatorio simple es equivalente a un sorteo de lotería en el que cada individuo de la población posee la misma probabilidad de ser seleccionado. De ahí que en lo sucesivo identifiquemos la probabilidad de que suceda algo en la población con la proporción de individuos de la misma que verifican ese algo.

El presente capítulo está dedicado a una muy breve explicación de los elementos fundamentales de la Inferencia Estadística, distinguiendo los dos problemas que pretende resolver: el de Estimación y el de Contraste de Hipótesis. En el capítulo siguiente expondremos una clasificación de las técnicas más populares de la Inferencia Estadística, siempre desde la perspectiva de las Ciencias de la Salud.

2.1. Problema de Estimación

Hemos distinguido dos tipos de parámetros o valores típicos: los muestrales o descriptivos, como \bar{x} , s , r , etc, y los probabilísticos, como μ o σ . En el caso de que el fenómeno aleatorio considerado sea el sorteo de una muestra aleatoria simple, sabemos que los parámetros proba-

bilísticos coinciden con los parámetros descriptivos de la población, es decir, que μ es la media aritmética de toda la población, σ^2 es la varianza de toda la población, etc. De ahí que los denominemos a partir de ahora **parámetros poblacionales**.

Estos parámetros se suponen normalmente desconocidos pues la población suele ser inabarcable. Sin embargo, sabemos que los parámetros de la muestra aleatoria convergen a sus análogos poblacionales a medida que el tamaño de la misma tiende a infinito. Esto es lo que da sentido al muestreo aleatorio. El problema de Estimación tiene por objeto estimar o aproximar los parámetros probabilísticos a partir de otros calculados directamente a partir de la muestra. De esa forma, podemos decir por ejemplo que la media aritmética \bar{X} de la muestra es un estimador de la media poblacional μ .

2.1.1. Criterios de Estimación

El problema de estimación es más complejo de lo que puede parecer a simple vista pues debemos establecer primeramente criterios para determinar si un estimador es aceptablemente bueno o si es peor que otro. Así, por ejemplo, puede llamar la atención el hecho de que la varianza muestral se haya definido de la forma

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

cuando lo natural hubiera sido dividir directamente por n . Y esto es así porque S^2 así definida (dividiendo por $n-1$) es un estimador insesgado de σ^2 , lo cual quiere decir que es exacto por término medio.

Los dos criterios más populares a la hora de justificar un estimador son el de **Mínimos Cuadrados** y el de **Máxima Verosimilitud**: el primero pretende minimizar el error cuadrático que se comete por término medio en la estimación, mientras que el segundo escoge el parámetro que hace la observación obtenida lo más verosímil posible. Precisamente, la varianza muestral dividiendo por n es el estimador de máxima verosimilitud de σ^2 cuando la variable considerada sigue un modelo de distribución normal. No obstante, en lo sucesivo no volveremos a hacer hincapié en estos aspectos.

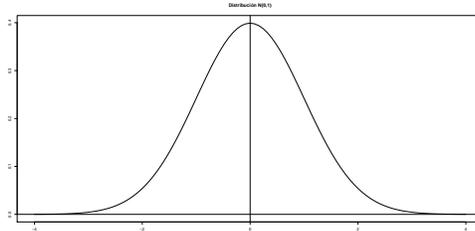
2.1.2. Intervalos de confianza

Existen más parámetros por estimar, como veremos en el capítulo siguiente. Ahora nos centraremos en otro aspecto del problema de estimación. El valor concreto que aporta un estimador en la muestra obtenida se denomina estimación puntual. Así, dada una muestra aleatoria, \bar{x} es una estimación puntual de μ . Por supuesto que dicha estimación está sometida a un error. No podemos esperar que coincida con el valor exacto del parámetro poblacional desconocido que estamos estimando. Sin embargo, nos gustaría precisar un probable margen máximo de error, de manera que podamos determinar un intervalo en el cual se encuentre seguramente el parámetro poblacional. ¿Cómo podemos construir ese intervalo? Veamos un ejemplo.

Se considera cierta variable cuantitativa X sobre una población Ω cuya media es μ . A través de una muestra aleatoria de tamaño n podemos estimar μ mediante su media aritmética \bar{X} (recordamos que la diferencia entre \bar{X} y \bar{x} consiste en que la primera denota la variable muestral y la segunda el valor concreto de dicha variable para la muestra concreta que se estudia). Si el

tamaño de muestra considerado es suficientemente grande (digamos $n > 30$), podemos aplicar el resultado (1.1) del capítulo anterior, de manera que

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$



En ese caso, se verifica entonces aproximadamente que

$$P\left(\frac{|\bar{X} - \mu|}{S/\sqrt{n}} \leq z_{0,05/2}\right) = 0,95$$

Es decir,

$$P\left(|\bar{X} - \mu| \leq z_{0,05/2} \frac{S}{\sqrt{n}}\right) = 0,95$$

Por lo tanto, la probabilidad de que, para una muestra aleatoria de tamaño n , la diferencia entre su media aritmética \bar{x} y la media poblacional μ sea a lo sumo $z_{0,05/2}S/\sqrt{n}$ es del 95%. Dicho de otra forma, en el 95% de las posibles muestras de tamaño n que pueden extraerse de la población, la diferencia entre la media de la muestra y la de la población es a lo sumo $z_{0,05/2}S/\sqrt{n}$. Esa cantidad se denomina margen máximo de error al 95% de confianza y se denota por $E_{\text{máx}}$.

De esta forma, el verdadero aunque desconocido valor de μ quedará dentro del intervalo $\bar{X} \pm z_{0,05/2}S/\sqrt{n}$ en el 95% de las posibles muestras de tamaño n . Dada una muestra concreta de tamaño n , se dice entonces que

$$\bar{x} \pm z_{0,05/2} \frac{s}{\sqrt{n}}$$

es un intervalo de confianza al 95% para la media μ . El valor de $z_{0,05/2}$ es, por cierto, 1.96. Cuando construimos un intervalo de confianza al 95% estamos asumiendo una probabilidad de error o riesgo del 5%. ¿Por qué el 5%? Pues por nada en especial, pero existe un convenio tácito en la Estadística de considerar algo como raro o poco probable cuando la probabilidad de que ocurra sea inferior al 0.05, seguramente por ser una cantidad pequeña y redonda. De ahí que lo más habitual sea construir intervalos al 95% de confianza. No obstante, podemos admitir otras opciones con niveles de riesgo diferentes. En general, si se denota por α la probabilidad de error (en el caso anterior tendríamos $\alpha = 0,05$) el intervalo de confianza a nivel $(1 - \alpha) \times 100\%$ para la media será

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Los valores alternativos más frecuentes para el nivel de riesgo son $\alpha = 0,01, 0,001$. También se asumen en ocasiones riesgos mayores, como $\alpha = 0,10$.

Podemos construir intervalos de confianza para otros parámetros poblacionales como la varianza, el coeficiente de determinación, la pendiente de regresión, etc. No obstante, en la mayoría de las ocasiones será necesario suponer ciertas condiciones en el modelos de distribución considerado.

Ejemplo 8: [Intervalo de confianza para una media]

Se pretende estimar la media μ de la estatura X de las mujeres de entre 16 y 50 años pertenecientes a una amplia población. Para ello se escogió una muestra supuestamente aleatoria de $n = 40$ mujeres, las cuales aportaron una media aritmética de 162.3cm con una desviación típica de 5.2cm.

Así pues ya tenemos una estimación puntual de la media μ : la media aritmética $\bar{x} = 162,3$. El margen máximo de error al 5 % de confianza es

$$E_{\text{máx}} = 1,96 \cdot \frac{5,2}{\sqrt{40}} = 1,6$$

Por lo tanto, el intervalo de confianza al 95 % correspondiente es $162,3 \pm 1,6$. En definitiva, podemos afirmar con una confianza del 95 % que la media de altura de la población se encuentra entre 160.7cm y 163.9cm.

Observemos que, en general, no sólo el intervalo sino el propio margen máximo de error depende de la muestra obtenida. En el caso de que la varianza poblacional σ^2 fuese conocida, el margen máximo de error podría calcularse mediante

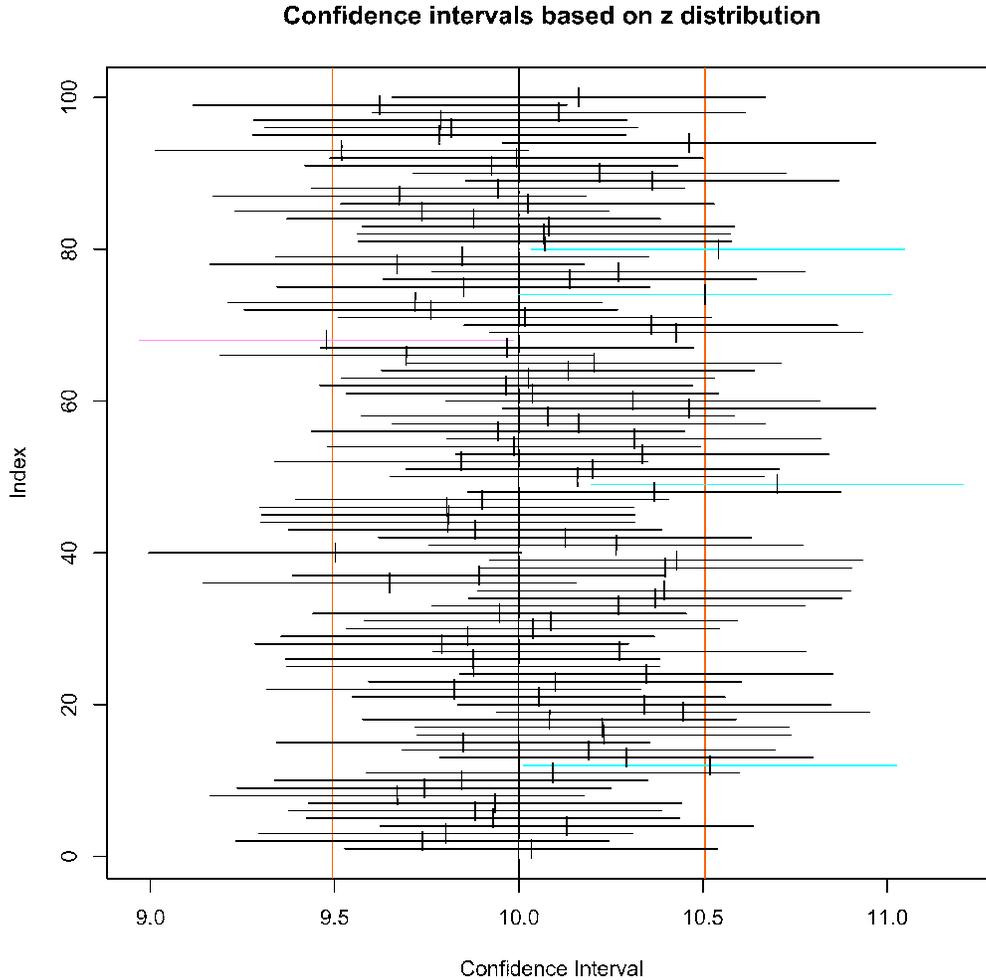
$$E_{\text{máx}} = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \tag{2.1}$$

con lo que dependería de la muestra únicamente a través de su tamaño n . Este hecho será de utilidad en el siguiente capítulo para determinar el tamaño de muestra requerido en función del probable margen máximo de error que estemos dispuesto a asumir en la estimación del parámetro poblacional. También puede ser de utilidad para ilustrar qué entendemos exactamente por intervalo al 95 % de confianza.

Efectivamente, supongamos que la media y desviación típica de la población (parámetros que en la práctica son desconocidos) fueran $\mu = 10$ y $\sigma = 2$. Según la expresión anterior, el margen máximo de error en la estimación de μ con una confianza del 95 % es

$$E_{\text{máx}} = 1,96 \frac{2}{\sqrt{60}} = 0,51$$

Así pues, un intervalo de confianza al 95 % para μ a partir de una muestra de tamaño $n = 60$ será de la forma $\bar{x} \pm 0,51$. A continuación simulamos mediante un programa estadístico la extracción de 100 muestras aleatorias diferentes, cada una de las cuales aporta una media aritmética distinta, pues es una variable muestral. Según lo dicho, debemos esperar que aproximadamente 95 de ellas disten de la verdadera media $\mu = 10$ a lo sumo 0.51 unidades, de manera que sus correspondientes intervalos de confianza contendrán efectivamente a la media de la población. En este caso observamos que en cuatro ocasiones (las líneas horizontales coloreadas) las muestras seleccionadas han diferido de $\mu = 10$ más de 0.51 unidades, de manera que los intervalos de confianza al 95 % contruidos a partir de estas cuatro muestras inducirían a error. El 5 % residual debe pues entenderse como la proporción de muestras cuyos intervalos asociados no contendrán realmente a μ por el hecho de que son extremas. Que sean extremas quiere decir que están compuestas por valores demasiado altos o demasiado bajos, lo cual es poco probable. Pero si tenemos la mala suerte de que ocurra nos haremos una idea equivocada de la media μ de a población.



Si queremos que esta probabilidad de error sea menor, por ejemplo del 1%, basta tomar $\alpha = 0,01$. Ello se traducirá en un mayor margen de error para la estimación y, en consecuencia, en intervalos de confianza más amplios. Es decir, que lo que ganamos en seguridad lo perdemos en precisión. Si queremos mayor seguridad sin disminuir la precisión o viceversa, nos vemos obligados a trabajar con muestras más amplias. De hecho, basta echar un vistazo a la ecuación (2.1) para entender que, a medida que el tamaño de muestra tiende a infinito, el margen máximo de error tiende a 0. Es una forma de decir que

$$\boxed{\lim_{n \rightarrow \infty} \bar{x} = \mu}$$

2.2. Problema de contraste de hipótesis

En esta ocasión no nos ocupa la estimación de un parámetro poblacional, sino evaluar la validez de un determinado modelo teórico para explicar el comportamiento de nuestros datos, denominado hipótesis inicial. La decisión ha de tomarse pues de manera razonable, a partir de la información que presta una muestra aleatoria de tamaño n . Denominamos **test de hipótesis** al algoritmo o proceso matemático preestablecido al que se someten los n datos de la muestra y que desemboca en la toma de decisión. No existe, evidentemente, un único test para cada problema de decisión. Desde un punto de vista teórico, el problema de decisión consiste en

establecer criterios razonables de comparación de tests, y determinar qué test es el mejor, según el criterio establecido.

2.2.1. Planteamiento del problema.

Para entender las dificultades a las que nos enfrentamos y los elementos de un problema de contraste de hipótesis, consideraremos un problema sencillo.

Ejemplo 9:[Contraste bilateral para una probabilidad o proporción]

Se estudia si en una pequeña localidad existen factores de tipo ambiental capaces de modificar la distribución natural del sexo en los recién nacidos. Para ello se tienen en cuenta los 10 nacimientos acaecidos a lo largo del último año. Los resultados son los siguientes: HHVHVHHHHH

Partiremos de la **hipótesis inicial** de que en el pueblo no ocurre nada que pueda influir en el sexo del bebé. Esta hipótesis debe ser juzgada o contrastada mediante la muestra estudiada, es decir, los 10 nacimientos registrados. Analizaremos si estos datos suponen una prueba significativa contra la hipótesis inicial, que se denota por H_0 . Como alternativa a esta hipótesis puede ocurrir que ciertos agentes ambientales favorezcan el nacimiento de hembras o bien el nacimiento de varones.

La primera regla del contraste de hipótesis es que H_0 debe identificarse con una distribución de probabilidad concreta. Sabemos que el sexo del bebé depende de si el espermatozoide que fecunda el óvulo porta el cromosoma X o el Y. En principio, cabe pensar por simetría que la proporción de espermatozoides X es idéntica a la de espermatozoides Y. Supongamos además que ambos tipos de espermatozoides se comportaran igualmente en el proceso de fecundación¹. Entonces, si ningún otro factor externo influye en la fecundación o el posterior desarrollo sexual del embrión y el feto, cabría equiparar la secuencia de $n = 10$ nacimientos con una serie de 10 lanzamientos independientes de una moneda simétrica, por lo que podríamos hablar de una probabilidad $p = 0,50$ de que el bebé sea hembra. Así podemos expresar la hipótesis inicial

$$H_0 : p = 0,50$$

Ya hemos traducido nuestra hipótesis inicial al lenguaje probabilístico, pero dijimos que debíamos identificarla con una distribución de probabilidad, que por definición debe estar asociada a alguna variable numérica, es decir, que nuestra observación debe resumirse en un único número, del que dependerá la decisión final sobre la veracidad de esta hipótesis inicial. En este caso, la observación, es decir, la muestra, consiste en una secuencia de varones y hembras obtenida en un año. A partir de la muestra debemos pues calcular un número que resuma de la mejor manera posible la **información** que la observación contiene respecto al problema planteado. Ese número se denomina **valor experimental** y debe entenderse como el resultado particular de una variable numérica (definida sobre el espacio de todas las posibles muestras de tamaño n) denominada estadístico de contraste. De la capacidad de ese valor experimental para resumir la información dependerá la **potencia** del test que diseñemos a partir del mismo, es decir, la capacidad para detectar la falsedad de la hipótesis inicial en el caso de que dé. Quiere decir esto que si la muestra es pequeña o no somos capaces de resumirla satisfactoriamente con un solo número será difícil probar la posible falsedad de H_0 .

En nuestro caso, parece intuitivo que, para decidir si existe o no mayor tendencia hacia uno de los sexos, basta con fijarnos en el número de hembras en la serie de nacimientos (o,

¹Lo cual es mucho suponer. Ver la cuestión 14.

equivalentemente, el número de varones), variable numérica que se denotará por \mathcal{H} . Ése es nuestro estadístico de contraste. Según estudiamos en el capítulo anterior, que la hipótesis inicial sea cierta equivale a que \mathcal{H} se distribuya según un modelo de probabilidad $B(10, 0.50)$, es decir

$$H_0 : \mathcal{H} \sim B(10, 0.50)$$

La muestra particular obtenida aporta 8 hembras. Ése es pues nuestro valor experimental. Se puede denotar $\mathcal{H} = 8$ o, mejor,

$$\mathcal{H}_{exp} = 8$$

Nótese que el orden en el que han nacido los varones y las hembras no se ha tenido en cuenta en la decisión, cosa que no debe afectar a la calidad del test diseñado, dado que el orden se antoja irrelevante en este problema concreto. La siguiente pregunta es: ¿cómo debe ser el valor experimental para rechazar la hipótesis inicial?

Parece claro que un valor $\mathcal{H} = 5$, es decir, 5 varones y 5 hembras, sería acorde con la hipótesis inicial de que en el pueblo no concurren circunstancias que alteren la simetría en el sexo del bebé. Sin embargo, un gran número de hembras podría evidenciar la falsedad de ese modelo en favor de otro en el cual existiera una mayor tendencia al nacimiento de niñas. Algo idéntico ocurriría con un gran número de varones o, lo que es lo mismo, un pequeño número de hembras, sólo que indicaría una mayor tendencia al nacimiento de niños. Pero ambos casos conducirían por igual a pensar en la falsedad de H_0 . En definitiva, queda claro que un valor experimental demasiado grande o demasiado pequeño, es decir, **extremo**, puede contradecir la hipótesis inicial.

Dado que el test es por definición un algoritmo automático, debemos explicitar previamente a la observación de la muestra concreta qué entendemos exactamente por extremo. ¿Contradice la hipótesis inicial el hecho de obtener 8 hembras? O lo que es lo mismo, ¿podríamos obtener 8 caras tras 10 lanzamientos con una moneda simétrica? Evidentemente, la respuesta es sí. Incluso, teóricamente existe la posibilidad de obtener 100 caras consecutivas con una moneda simétrica. Otra cosa es que ello es muy poco probable por lo que, si ocurriera, una persona razonable se inclinaría a pensar que la moneda tiene tendencia a resultar cara, porque ese modelo teórico explica mejor las observaciones.

Llegamos a sí a la segunda regla fundamental en el contraste de hipótesis y, posiblemente, el axioma principal de la Estadística:

| |
|---|
| Principio de Máxima Verosimilitud: |
|---|

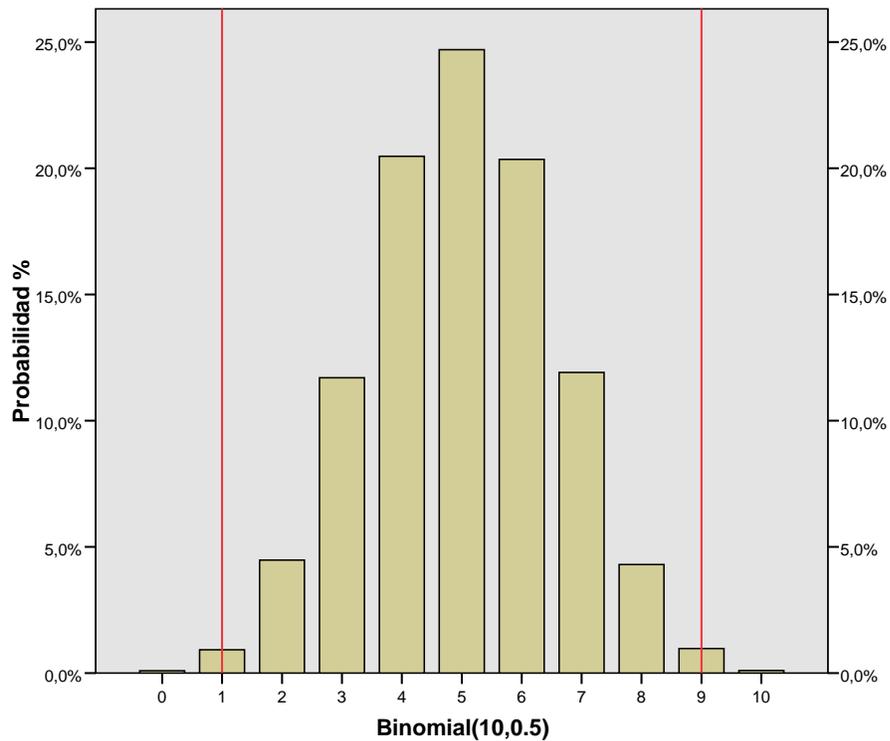
| |
|--|
| Entre dos modelos probabilísticos debemos elegir aquél que haga más verosímil nuestra observación. Es decir, si la observación es rara según un modelo teórico pero verosímil según otro, deberíamos rechazar el primero en favor del segundo. |
|--|

Así pues, siguiendo este axioma, diseñar un test de hipótesis significa determinar *a priori* cuáles de los posibles valores experimentales se considerarán **raros** según el modelo inicial aunque verosímiles para los modelos alternativos, de manera que, si el valor experimental resulta ser finalmente uno de ellos se rechazará la hipótesis inicial H_0 . Este conjunto de valores se denomina **región crítica**, mientras que el resto es la región de aceptación de H_0 .

Pues bien, ya hemos comentado con anterioridad que en Estadística se conviene en considerar un suceso raro cuando su probabilidad es inferior a 0,05. Esta afirmación deja mucho que desear. Basta pensar en una distribución continua donde cualquier valor concreto se verifica con probabilidad 0. Se trata más bien de determinar una región cuyos elementos sumen una

probabilidad de a los sumo 0.05. Teniendo en cuenta lo dicho anteriormente, esa región debe ser extrema, es decir, alejada del valor central 5, porque de esa forma quedan mejor explicados por los modelos alternativos y, además, simétrica respecto a 5 porque no tenemos ninguna razón para privilegiar alguno de los lados. Efectivamente, no parecería razonable, por ejemplo, que obtener 10 hembras condujera a rechazar la hipótesis inicial $p = 0,50$ pero que 10 varones (0 hembras) no lo hiciera, o que 7 hembras condujera a aceptarla pero 7 varones no. Por lo tanto, la región crítica será un conjunto extremo, raro (con probabilidad igual o inferior a 0.05) y, al menos en este caso, simétrico.

Para construirlo nos valdremos del conocimiento de la función de probabilidad de la distribución $B(10,0.50)$ cuyo diagrama de barras es el siguiente



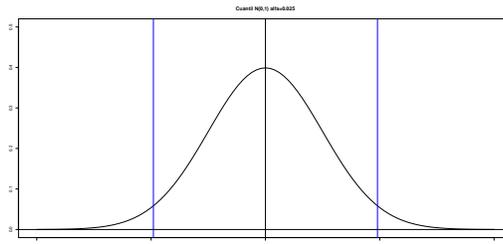
Según este modelo de probabilidad, los valores 0 y 10 pueden obtenerse con probabilidad 0.001. La suma de ambos valores es 0.002 (si 10 pertenece a la región crítica también debe pertenecer 0), es decir, la probabilidad de obtener un resultado tan extremo como 10 es de 0.002, lo cual sería entonces raro y nos llevaría a rechazar el modelo inicial. Por lo tanto, 0 y 10 deben formar parte de la región crítica. No obstante, dado que esta probabilidad es inferior a 0.05 puede que tengamos margen para ampliar la región crítica con otros valores no tan extremos. Si añadimos el par 1 y 9 (el 1 debe conducir a la misma decisión que el 9), dado que cada uno presenta probabilidad 0.01, obtendríamos una región que sumaría una probabilidad de 0.022. Ésa es la probabilidad de de obtener un resultado al menos tan extremo como 9. Es por lo tanto poco verosímil según la hipótesis inicial por lo que, si ocurriera realmente, nos llevaría a rechazar dicha hipótesis. Por lo tanto, 1 y 9 deben estar también incluidos en la región crítica.

¿Podemos ampliarla aún incluyendo los valores 8 y 2? Dado que la probabilidad de obtener 8 es de 0.044 (la de 2 también), obtendríamos una suma acumulada de 0.110. Por lo tanto, obtener un resultado al menos tan extremo como 8 presenta una probabilidad aceptable según

el modelo o hipótesis inicial, por lo que la región constituida por 0,1,2,8,9 y 10 ya no podría considerarse rara para este modelo. Por lo tanto, 8 y 2 no pueden estar incluidos en la región crítica, que queda configurada finalmente por los valores 0,1,9 y 10. Es decir, la muestra contradirá significativamente la hipótesis inicial si el número de hembras es 0,1,9 o 10 (de ahí las líneas rojas que se muestran en el gráfico) o, equivalentemente, si nacen 9 o 10 hembras o bien 9 o 10 varones. En tal caso diremos que el **resultado del test es significativo**, lo cual querrá decir que la observación supone una prueba clara contra H_0 .

En el caso concreto del ejemplo, donde el número de hembras es $\mathcal{H}_{exp} = 8$, aplicando este test obtenemos un resultado no significativo, es decir, la observación no llega a ser lo suficientemente extraña desde el punto de vista de la hipótesis inicial por lo que no logra contradecirla con claridad y, por lo tanto, no permite rechazarla. Por lo tanto, la muestra estudiada no permite concluir que en el pueblo ocurra algo especial que altere la distribución simétrica del sexo en los nacimientos. Ciertamente, se han obtenido bastantes hembras (8 de 10), pero no son suficientes porque pueden achacarse al azar, es decir, aunque no existieran factores que alterasen la simetría sería verosímil que de 10 nacimientos 8 sean hembras, por pura suerte. De haber logrado una hembra más ya no podríamos decir lo mismo y habría supuesto un resultado significativo, con lo que concluiríamos que en el pueblo concurren circunstancias que alteran la simetría original.

En definitiva, un test de hipótesis viene a delimitar los límites del azar a la hora de explicar una observación según un modelo concreto: la hipótesis inicial. En el caso de estadísticos de contrastes que sigan distribuciones continuas el problema es más sencillo. Por ejemplo, si tenemos una distribución $N(0, 1)$ la región crítica quedará delimitada por el cuantil $z_{0,05/2}$ y su opuesto $-z_{0,05/2}$, pues la probabilidad de obtener un valor más extremo que éstos es exactamente 0.05, como se indica en el gráfico.



En los casos como éste en que la región crítica queda delimitada por un el cuantil, dado que el mismo se calcula a partir de un modelo de probabilidad teórico, éste se denomina **valor teórico**, en contraposición con el valor experimental. De hecho, el test de hipótesis consiste en comparar un valor experimental que proporciona la muestra con un valor teórico que proporciona la hipótesis inicial.

En definitiva, los elementos de un test de hipótesis son, en general, los siguientes:

1. **Hipótesis inicial:** modelo probabilístico teórico cuya validez se juzga mediante una muestra. Si la muestra resulta extrema y poco verosímil según este modelo supondrá una prueba significativa contra el mismo y tendremos que rechazarlo.
2. **Valor experimental:** un número que resumirá en lo posible la información relevante que aporta la muestra.

3. **Potencia:** capacidad del test de detectar la falsedad de la hipótesis inicial. Cuanto más información relativa al problema sea capaz de recoger el valor experimental mayor será la potencia del test.

4. **Región crítica:** es un conjunto tal que la pertenencia del valor experimental al mismo es poco verosímil (probabilidad inferior a 0.05) según la hipótesis inicial, aunque se procura que sea más verosímil para modelos alternativos (extremo), por lo que este hecho debe implicar el rechazo de la hipótesis inicial en favor de alguno de estos modelos. En el caso continuo la región crítica se construye a partir de un cuantil denominado valor teórico.

2.2.2. P-valor

Este concepto es fundamental porque viene a expresar el resultado final de un problema de contraste de hipótesis, lo cual puede convertirlo en el parámetro definitivo en un estudio más envergadura.

Tal y como hemos construido el test, aceptaremos o rechazaremos la hipótesis inicial dependiendo exclusivamente de si el valor experimental está o no dentro de una región extrema y poco probable, estableciendo un tope del 5% para lo que llegamos a entender como raro o poco probable. Debemos tener en cuenta primeramente que esta cota es puramente convencional. En segundo lugar, no es exactamente lo mismo que el valor experimental este muy cerca de la región crítica, como ha sido el caso de nuestro ejemplo, que lejos de ésta, aunque en ambos casos la decisión final sea la misma: aceptar H_0 .

Retomando el ejemplo anterior, no sería lo mismo, aunque ambos casos habría conducido a aceptar la hipótesis inicial (en el pueblo no ocurre nada especial), que nacieran 5 u 8 hembras. Tampoco sería lo mismo, aunque ambos casos habría conducido a rechazar la hipótesis inicial (en el pueblo pasa algo), que nacieran 9 o 10 hembras. Al margen del tope que delimita el 5% de casos extremos, constituido por el par 1-9, nos gustaría dar una medida del grado de verosimilitud de nuestra muestra según la hipótesis inicial. Esto nos lo proporciona el denominado P-valor o probabilidad de significación, que se define como la probabilidad de obtener una muestra al menos tan extrema como la nuestra según la hipótesis inicial.

Concretamente, con la secuencia de nacimientos obtenida tenemos 8 hembras y 2 varones. La probabilidad de obtener un caso igual de extremo o más que el nuestro, es decir, la probabilidad de obtener 8,2,9,1,10 o 0 hembras según el modelo $B(10, 0.5)$ es $P = 0,110$. Ese es el P-valor que corresponde a nuestra muestra. Al ser mayor que 0.05 se dice que es un resultado no significativo. Si hubiéramos obtenido, por ejemplo, 9 hembras, habría que sumar las probabilidades de obtener 9,1,10 y 0, obteniendo $P = 0,022$. Este sí sería un resultado significativo porque $P < 0,05$, lo cual equivale a que 9 se encuentre en la región crítica. Que este valor de P sea inferior al anterior nos habla de una muestra menos verosímil según la hipótesis inicial. De hecho, es tan poco verosímil que nos invita a rechazar la hipótesis. Sin embargo, si el número de hembras hubiera sido 10 (el caso más extremo) el P-valor sería 0.002, lo cual nos habla de una situación casi inverosímil según la hipótesis inicial, lo cual invitaría a rechazarla pero con mayor seguridad que en el caso $P = 0,022$.

En definitiva, no sólo distinguiremos entre resultados significativos ($P < 0,05$) y no significativos ($P > 0,05$) sino que hablaremos de distintos grados de significación. Concretamente, lo más habitual es clasificar el resultado según la tabla siguiente:

| | |
|-----------------------|-----------------------------------|
| $0,05 < P$ | Resultado no significativo |
| $0,01 < P \leq 0,05$ | Resultado significativo |
| $0,001 < P \leq 0,01$ | Resultado muy significativo |
| $P \leq 0,001$ | Resultado altamente significativo |

Ya hemos dicho que establecer el tope de lo que consideramos como raro en el 5% es puramente convencional. De hecho, en ocasiones estamos dispuestos a ser más críticos con la hipótesis inicial elevando ese tope al 10%, lo cual supondría ampliar la región crítica. Eso se hace en ocasiones para compensar la ventaja que para la hipótesis inicial supone una muestra pequeña. En otras circunstancias, queremos adoptar una actitud más conservadora disminuyendo el tope al 1% o incluso al 0.1%, lo cual implicaría reducir la región crítica. Esto ocurre cuando rechazar la hipótesis inicial puede implicar en la práctica serias modificaciones por lo que precisamos estar especialmente convencidos de ello. El tope que establezcamos finalmente, que suele ser del 5%, se denomina nivel de significación del test, y se definiría formalmente como la probabilidad de la región crítica según el modelo H_0 .

Pero en definitiva, debemos tener presente que una misma muestra puede conducir a distintas conclusiones según el grado de riesgo que estemos dispuestos a asumir respecto a la hipótesis inicial. Sin embargo, ello no debe suponer ninguna complicación si conocemos el P -valor correspondiente a nuestra muestra, porque sabemos en qué medida es verosímil para H_0 y ése es el resultado final del test de hipótesis. Que esa verosimilitud sea considerada grande o pequeña es completamente subjetivo, aunque se convenga en establecer 0.05 como cota habitual. El P -valor es el resultado final que aporta un programa estadístico cuando aplica un contraste de hipótesis.

Muestra \leftrightarrow Valor experimental \leftrightarrow P -valor

Si el estadístico de contraste sigue una distribución continua, obtener el P -valor es más sencillo aún. Se trata de calcular el área de las colas que delimitan el valor experimental por un lado y su opuesto por otro en la densidad de la distribución. Retomemos los datos del ejemplo 8. Supongamos que, por estudios previo, se considera que la estatura media de las mujeres de esa franja de edades es de $\mu_0 = 164,5\text{cm}$. Vamos a utilizar los $n = 40$ datos de nuestra muestra para contrastar la veracidad de dicha afirmación.

$$H_0 : \mu = 164,5$$

En primer lugar, debemos calcular a partir de la muestra un número (valor experimental) que resuma lo mejor posible la información relevante para este problema. Aunque no deje de ser intuitivo, es necesario un conocimiento más profundo de la Estadística para poder determinar el valor experimental. No obstante, no parece mala idea que la media aritmética intervenga en el cálculo, aunque también sabemos que por sí sola difícilmente podrá resumir suficientemente la información, por lo que la desviación típica se antoja también necesaria. Tal vez un único número calculado a partir de ambas podría recoger suficiente información, al menos en lo relativo a este problema, máxime teniendo en cuenta (1.1). En definitiva, el estadístico de contraste que proponemos en este caso es el siguiente

$$\mathcal{T} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

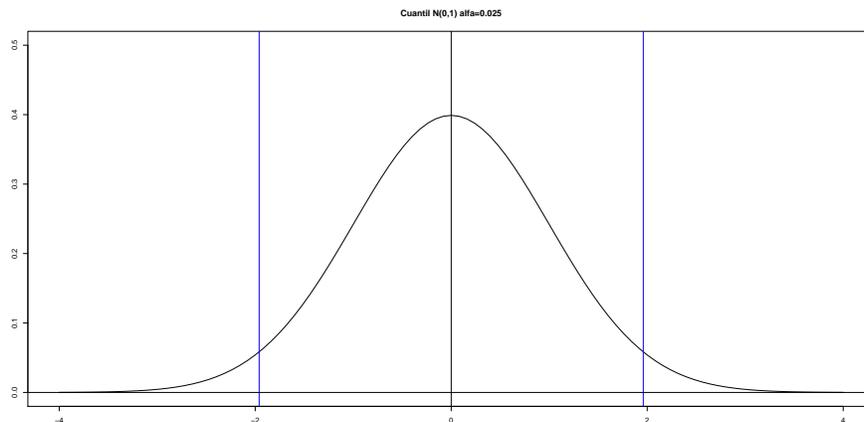
Según (1.1), si la media de la población μ coincidiera efectivamente con $\mu_0 = 164,5$, el estadístico de contraste seguiría (aproximadamente) un modelo de distribución $N(0, 1)$. Por lo tanto, la hipótesis inicial puede expresarse de la forma

$$H_0 : \mathcal{T} \sim N(0, 1)$$

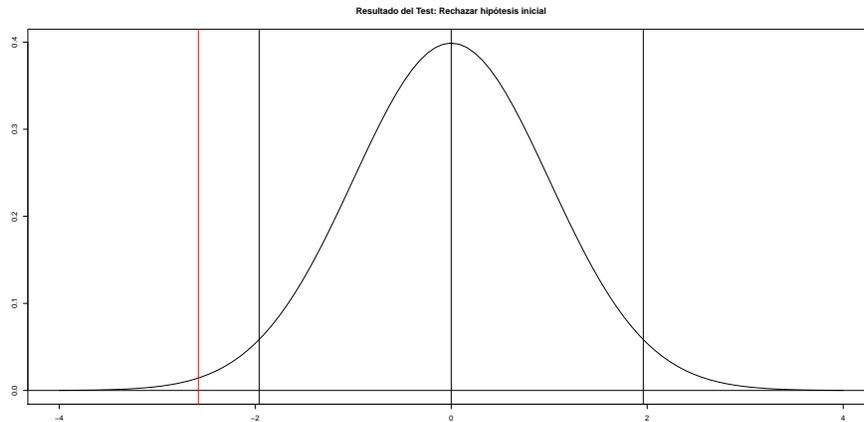
El valor experimental, que se obtiene sustituyendo en la expresión anterior los datos correspondientes a nuestra muestra es

$$\mathcal{T}_{exp} = \frac{162,3 - 164,5}{5,2/\sqrt{40}} = -2,67$$

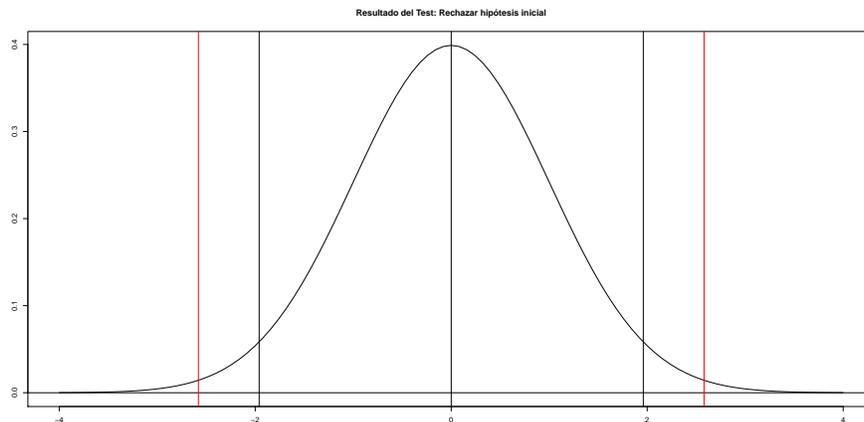
Si la hipótesis inicial fuera cierta, es decir, si $\mu = 164,5$, cabría esperar que la media aritmética de la muestra fuera próxima a 164.5 o, equivalentemente, un valor experimental próximo a 0. Si la verdadera media μ dista mucho de μ_0 , la distribución de \mathcal{T} será normal pero no de media 0 sino $\mu - \mu_0$. Si esta diferencia es positiva, los valores alejados de 0 por la derecha, que eran poco verosímiles según el modelo $N(0, 1)$, serán más verosímiles según el modelo alternativo. Si la diferencia $\mu - \mu_0$ es negativa, será más verosímiles los valores a la izquierda. Así pues, queda claro que un valor experimental lejano a 0, es decir, extremo, debe conducir a rechazar la hipótesis inicial. También parece razonable, por simetría, que la decisión que corresponde un determinado valor experimental debe ser la misma que corresponde a su opuesto. Así pues, la región crítica del test estará constituida por dos colas simétricas respecto a 0 de la curva $N(0, 1)$ que sumen un área o probabilidad de 0.05. La cola de la derecha debe ser pues la que queda delimitada por el cuantil $z_{0,05/2}$ y la de la izquierda, por el valor $-z_{0,05/2}$, como se indica en la figura



Si el valor experimental queda en las colas, debemos rechazar la hipótesis inicial. Si queda entre ellas, la aceptaremos. En nuestro caso ocurre lo primero pues $\mathcal{T}_{exp} = -2,67$



Es decir, rechazamos la hipótesis inicial cuando $|\mathcal{T}_{exp}| > z_{0,05/2}$. En este caso, ocurre efectivamente que $2,67 > 1,96$. No obstante, conviene en todo caso calcular el P -valor, que hemos definido como la probabilidad, según H_0 (es decir, según el modelo $N(0,1)$), de obtener un resultado al menos tan extremo como \mathcal{T}_{exp} . Viendo el anterior gráfico, se trataría del área que queda a la izquierda de la línea roja multiplicada por 2, puesto que habría que sumarle la cola simétrica.



El valor exacto de P se obtendría pues resolviendo la ecuación

$$z_{\frac{P}{2}} = |\mathcal{T}_{exp}|$$

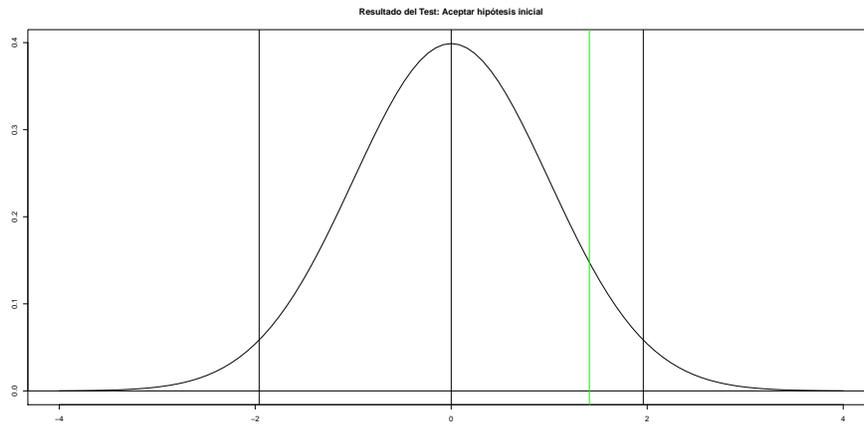
aunque cualquier programa estadístico lo calcula automáticamente. En este caso se obtiene $P = 0,0076$. Como cabía espera, el P -valor es inferior a 0.05. Es más, incluso queda por debajo de 0.01, lo que supone un resultado muy significativo. Para valorarlo en su justa medida debemos tener muy presente el significado del P -valor:

El P -valor viene a ser una medida del grado de verosimilitud de la muestra desde el punto de vista del modelo inicial, es decir, informa de lo rara y extrema que es nuestra muestra según dicha hipótesis.

Así pues, cuanto más pequeño sea el P -valor, mayor será la contradicción entre nuestra muestra y la hipótesis inicial y, en consecuencia, más significativas serán las pruebas en su contra. En

este caso, tenemos pruebas muy significativas contra la hipótesis inicial de que la media sea 164.5, lo que nos induce a pensar que esta hipótesis es falsa.

Imaginemos que los datos de la muestra hubieran aportado una media de 165.7cm con una desviación típica de 3.8cm. En ese caso, habríamos obtenido un valor experimental $\mathcal{T}_{exp} = 1,41$, que representamos a continuación con una línea verde



Como podemos apreciar, queda dentro de la región de aceptación de la hipótesis inicial pues $|\mathcal{T}_{exp}| \leq 1,96$. Por lo tanto, el correspondiente P -valor debe ser superior a 0.05. Concretamente se tiene que $P = 0,1586$. En definitiva, esta muestra no aportaría una prueba significativa contra la hipótesis inicial. Es decir, no estamos en condiciones de rechazarla. La diferencia existente entre la media supuesta, 164.5cm, y la que presenta la muestra, 165.7cm, puede explicarse por el azar inherente al muestreo.

2.2.3. Relación entre test de hipótesis e intervalo de confianza

Los propios datos originales del ejemplo 8 pueden servir para evidenciar una clara vinculación entre el intervalo al 95 % de confianza y el test de hipótesis considerado, dado que ambos han sido construidos partiendo del resultado (1.1) del capítulo anterior

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{\text{aprox}}{\sim} N(0, 1)$$

Recordemos que, con estos datos, el intervalo de confianza al 95 % para la media de nuestra población es (160.7, 163.9). Es decir, con una confianza del 95 % y, por lo tanto, asumiendo una probabilidad de error de 5 %, afirmamos que la media μ se encuentra entre esos límites. De no ser así, significaría que nuestra muestra estaría entre el 5 % de muestras más extremas, cosa que por principio nos negamos a pensar. Dado que $\mu_0=164.5$ queda fuera del intervalo, debemos entender entonces que nuestra media no puede ser μ_0 , que es precisamente lo que se concluye tras aplicar el test de hipótesis. Si queremos analizar el porqué de esta coincidencia basta tener en cuenta que μ_0 pertenece al intervalo de confianza al 95 % cuando

$$|\bar{X} - \mu_0| \leq E_{\text{máx}}$$

es decir, cuando

$$\frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} \leq z_{\frac{0,05}{2}}$$

o, equivalentemente, cuando

$$|\mathcal{T}_{exp}| \leq z_{\frac{0,05}{2}},$$

que es lo que tiene que ocurrir para que el resultado del test de hipótesis sea no significativo. En nuestro caso el resultado es $P = 0,0078$, que es muy significativo. Puede probarse igualmente que eso equivale a que μ_0 quede fuera del intervalo al 99 % de confianza para la media, mayor que el anterior. Como no se trata de un resultado altamente significativo, podemos comprobar que μ_0 sí queda dentro del intervalo al 99.9 % de confianza. La regla es sencilla: el nivel de riesgo α que se asume al construir el intervalo debe concordar con la probabilidad asociada a la región crítica del test, es decir, con su nivel de significación.

Este vínculo entre intervalos de confianza y tests de hipótesis puede extrapolarse a muchos de los métodos que estudiaremos en el siguiente capítulo, y será de especial interés a la hora de contrastar la hipótesis inicial igualdad entre las medias μ_1 y μ_2 de sendas poblaciones. Efectivamente, en ese caso, la hipótesis inicial equivale a que la diferencia entre ambas medias sea nula. El programa estadístico proporcionará, además del resultado del test correspondiente en forma de P -valor, un intervalo al 95 % de confianza para la diferencia de medias $\mu_1 - \mu_2$. De esta forma, el resultado será significativo cuando el valor 0 quede fuera de dicho intervalo. Pero el intervalo tiene la virtud añadida de expresar el tamaño de la diferencia entre ambas medias.

2.2.4. Hipótesis alternativa: contrastes bilaterales y unilaterales

Hasta ahora nos hemos centrado en la hipótesis inicial H_0 y hemos hablado en términos muy vagos de su alternativa. Hemos entendido como hipótesis alternativa cualquier modelo teórico diferente del inicial H_0 . En el caso del ejemplo 9, si imponemos una serie de supuestos formales, esa familia de modelos se expresaría mediante $p \neq 0,5$. Esa hipótesis (o familia de hipótesis) se denota por H_1 . Así pues el contraste se plantea de la forma

$$\begin{cases} H_0 : p = 0,50 \\ H_1 : p \neq 0,50 \end{cases}$$

En el caso de los datos del ejemplo 8, las hipótesis a contrastar son

$$\begin{cases} H_0 : \mu = 164,5 \\ H_1 : \mu \neq 164,5 \end{cases}$$

También podemos contrastar si la media de dos poblaciones, μ_1 y μ_2 , son o no diferentes. En tal caso, la hipótesis inicial es $H_0 : \mu_1 = \mu_2$, mientras que la alternativa es la negación de H_0 , es decir $H_1 : \mu_1 \neq \mu_2$.

Sin embargo, en ocasiones tenemos una idea más clara y por lo tanto restrictiva de la hipótesis alternativa. Volvamos al ejemplo 9: existe la teoría de que ciertos contaminantes ambientales no sólo están afectando a la capacidad de reproducción masculina sino que incluso está impidiendo que los embriones y fetos masculinos prosperen. En definitiva, de ser eso cierto, existiría una mayor tendencia a los nacimientos de niñas en las zonas con mayor exposición ($p > 0,50$). Supongamos que nuestro pueblo es una de esas zonas y que lo hemos seleccionado como muestra para contrastar dicha teoría. En ese caso, la hipótesis inicial es, como siempre²,

²Porque esta hipótesis debe identificarse con un fenómeno aleatorio concreto a partir del cual podamos calcular probabilidades.

$H_0 : p = 0,50$. Sin embargo, la hipótesis alternativa no es la negación de la anterior sino $H_1 : p > 0,50$. Así pues, nos planteamos el contraste

$$\begin{cases} H_0 : p = 0,50 \\ H_1 : p > 0,50 \end{cases}$$

Contrastes de este tipo se denominan **unilaterales** en contraposición de con los anteriores, denominados **bilaterales**. ¿En qué afecta este matiz al diseño del test de hipótesis? Pues viene a romper la simetría en la región crítica. En este caso, un número elevado de hembras en la secuencia de nacimientos puede resultar raro según la hipótesis inicial pero verosímil según la alternativa considerada, lo que debe conducirnos a optar por esta última. Sin embargo, un escaso número de hembras (muchos varones) puede resultar raro para la hipótesis inicial pero lo será mucho más para la alternativa, por lo que el Principio de Máxima Verosimilitud nos conduce a aceptar H_0 . Por lo tanto, la región crítica para este contraste debe estar constituida exclusivamente por los valores extremos y raros a la derecha de 5. Como no hay que sumar la probabilidad de las dos colas estamos en principio condiciones de ampliar la región crítica por este lado, es decir, vamos a ser más críticos que en el planteamiento bilateral si la muestra presenta más hembras que varones. Por contra, si presenta más varones que hembras la decisión será automáticamente H_0 .

¿Cómo afecta este nuevo diseño al P -valor? Pues en el caso de que haya más varones que hembras no se define siquiera. Si el número de hembras es mayor, el P -valor será la probabilidad de obtener una valor tan grande al menos como ése. Como no hay que considerar la región simétrica a la izquierda de 5, esta probabilidad será exactamente la mitad del P -valor correspondiente al contraste bilateral. Por lo tanto, con los datos de nuestro ejemplo, tendríamos

$$P = \frac{0,110}{2} = 0,055$$

Vemos que el P -valor ha disminuido, lo que supone un resultado más crítico hacia H_0 , aunque sigue sin ser significativo.

Ni que decir tiene que pueden considerarse hipótesis alternativas del tipo $H_1 : p < 0,50$. En ese caso, la región crítica quedaría a la izquierda y la regla para obtener el P -valor sería la misma pero al contrario. También podemos considerar hipótesis del tipo $H_1 : \mu_1 < \mu_2$, $H_1 : p_1 > p_2$, etcétera.

2.3. Cuestiones propuestas

1. Se estudia cierta variable X . Una muestra de tamaño n aportó un determinado intervalo de confianza para la media μ de la variable al 95 % de confianza. Razona si el intervalo de confianza al 99 % ha de ser más o menos amplio que el anterior.
2. En una muestra de 100 pacientes con infarto se ha medido el GOT a las 12 horas, obteniéndose una media de 80 y una desviación típica de 120. Construir un intervalo de confianza al 95 % para la media de todos los infartados. Según estudios anteriores el valor medio del GOT es de 85. Contrasta esta hipótesis calculando el correspondiente P -valor. Relacionar el resultado obtenido con el intervalo de confianza anterior.
3. Supongamos que el tiempo utilizado en la atención a un paciente es una variable aleatoria. Se pretende determinar de la manera más precisa posible el tiempo medio esperado de

atención a partir de una muestra supuestamente aleatoria de tamaño 50 que aportó una media aritmética de 34 minutos con una desviación típica de 2.3 minutos. ¿Qué podemos hacer? Según los organismos públicos el tiempo medio de atención no excede de los 30 minutos. Contrastar dicha hipótesis inicial calculando el P -valor.

4. En el contraste de hipótesis del ejemplo 9 se ha obtenido un P -valor de 0.110, lo cual supone un resultado no significativo. ¿Significa eso que se ha demostrado que no existen en el pueblo factores ambientales que alteren la simetría en el sexo de los bebés? Si no se está de acuerdo con dicha afirmación, ¿qué deberíamos hacer?
5. Diseñar un test de hipótesis para contrastar la hipótesis inicial anterior pero partiendo en esta ocasión de una muestra de 100 nacimientos. Indica que P -valor se obtendrá si la proporción de hembras en la muestra es del 80 %.
6. Cuando hemos construido el test para el contraste bilateral de una media hemos afirmado que el estadístico de contraste

$$\mathcal{T} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

calculado a partir de la media aritmética y la desviación típica, podría recoger suficiente información de la muestra, al menos en lo relativo a este problema. Decimos “podría” porque ello ocurre bajo ciertas condiciones. ¿Puedes intuir en que condiciones el estadístico \mathcal{T} es idóneo?

7. ¿En qué sentido puede influir el tamaño de la muestra en un test de hipótesis?
8. Al contrastar la hipótesis inicial $\mu = 164,5$ con los datos del ejemplo 8 se ha obtenido un resultado muy significativo. ¿Estamos entonces seguros de que la media de la población difiere de 164.5?
9. En un problema de contraste de hipótesis se obtiene como resultado final $P > 0,05$. ¿Significa eso que se ha demostrado la autenticidad de H_0 ? ¿Cómo debe interpretarse un valor $P < 0,05$? ¿Cuál ha de ser el valor de P para tener certeza absoluta de la falsedad de H_0 ?
10. Partiendo de una muestra aleatoria de tamaño $n = 250$ de una variable, se obtuvo como resultado $\bar{x} = 13,1$ y $s = 2,2$. El intervalo al 95 % de confianza para la media es el siguiente:
 - A: (11.73 , 14.44)
 - B: (12.09 , 14.10)
 - C: (12.82 , 13.37)
 - D: (2.74 , 3.75)
 - E: (2.72 , 3.77)
11. En relación con la cuestión anterior, se plantea el problema de decidir si la media de la variable estudiada es igual a 13.3. Indica si son verdaderas o falsas cada una de las siguientes afirmaciones:

- A:** El resultado del test a partir de los datos de la muestra no es significativo.
- B:** Tenemos una confianza del 95 % de que la hipótesis inicial se da con absoluta seguridad.
- C:** El resultado del test a partir de los datos de la muestra es muy significativo.
12. ¿Por qué podemos afirmar que el resultado del test para contrastar la hipótesis inicial $H_0 : \mu = \mu_0$ es muy significativo cuando μ_0 queda fuera del intervalo al 99 % de confianza para la media μ ?
13. Con los datos del problema 8, describe la región crítica del test para contrastar la hipótesis inicial $H_0 : \mu = 164,5$ contra la alternativa unilateral $H_1 : \mu < 164,5$. Calcula de manera directa el P -valor correspondiente a este contraste.
14. Aunque en el ejemplo 8 hemos supuesto que, si no concurren agentes ambientales externos, la proporción de nacimientos de varones ha de ser idéntica a la de nacimientos de hembras, lo cierto es que se sabe desde hace tiempo que esto no es así y que, de hecho, depende de la composición étnica de la población. Tradicionalmente, se ha venido registrando en Europa Occidental año tras año proporciones de nacimientos varones en torno al 51 %. Por lo tanto, si se aplicara un test para contrastar la hipótesis inicial $H_0 : p = 0,50$, el resultado del mismo sería significativo, ¿o no? Comenta de manera crítica esta última afirmación.