# DISTRIBUCIONES DE PROBABILIDAD

**ÍNDICE:** **Pág.**

# DISTRIBUCIONES DE PROBABILIDAD

## 1. INTRODUCCIÓN.

En el tema anterior hemos comprobado cómo muchos fenómenos de la vida cotidiana y de las ciencias sociales son susceptibles de ser tratados en términos probabilísticos, y esta parte de la matemática se convierte en una herramienta valiosa para estudiar cualquier fenómeno aleatorio. Ahora vamos a continuar profundizando en el tema, utilizando el cálculo de probabilidades sobre poblaciones, es decir, considerando una determinada propiedad aleatoria sobre un conjunto de individuos y estudiando cuántas veces puede darse en dicho conjunto. Así, veremos que si en una clase de 300 alumnos de Derecho la probabilidad de que un alumno haya repetido es 0'15, podemos calcular qué probabilidad hay de que haya 25, 50, 11 alumnos suspensos del total.

## 2. VARIABLE ALEATORIA.

Para estudiar un fenómeno en términos probabilísticos, lo primero que vamos a hacer es cuantificarlo. Para ello tendremos que asignarle una propiedad que agrupe a los sucesos elementales y les asocie un número. Veamos un ejemplo:

Sea el experimento aleatorio consistente en lanzar tres veces una moneda.

El espacio muestral es $E = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$

Sabemos además que estos ocho sucesos elementales son equiprobables, con lo que cada uno tendrá una probabilidad de $\frac{1}{8}$.

Supongamos que lo que queremos estudiar es cuántas caras aparecen.

El fenómeno "número de caras" puede tomar los valores 0 (no aparecen caras), 1 (aparece una cara en las tres tiradas), 2 (dos de los tres resultados son caras) y 3 (todos los resultados son caras).

Luego a cada resultado del experimento aleatorio, a cada suceso elemental, le estamos asignando un número:

Esto es lo que llamamos una variable aleatoria.

Por tanto, una variable aleatoria es una función que asocia a cada elemento del espacio muestral E un número real.

**Ejemplo 1:**

Dado el experimento aleatorio consistente en lanzar dos dados distintos, construimos la variable aleatoria "suma de los resultados de las caras superiores" La variable aleatoria es:

A cada elemento del espacio muestral le asociamos la suma de las puntuaciones de las dos caras. Así, a $\{3,4\}$ le asociamos 7, a $\{2,3\}$ le asociamos 5, etc.

**Ejemplo 2**

Consideramos el experimento consistente en elegir al azar cien judías verdes de una plantación y medir su longitud. La función que asocia a cada judía su longitud es una variable aleatoria.

**Ejercicio 1:**

Construye la variable aleatoria asociada al experimento lanzar simultáneamente un dado y una moneda si la ley que la rige es: si sale cara, le asociamos al suceso el número del dado, y si sale cruz, el doble de éste.

## 2.1. TIPOS DE VARIABLES ALEATORIAS.

Si nos fijamos en los ejemplos 1 y 2, hay una diferencia fundamental entre las variables que aparecen: mientras que en la primera los resultados son enteros y finitos (se pueden contar), en la segunda los resultados pueden ser cualquier valor (12'5 cm, 7'87777 cm, $2\pi$ cm,...) y dados dos valores, siempre podremos encontrar otro intermedio. La primera variable diremos que es discreta, y la segunda, continua.

Diremos que una variable aleatoria es discreta cuando sólo puede tomar unos ciertos valores enteros.

Diremos que una variable aleatoria es continua cuando puede tomar todos los valores posibles dentro de un cierto intervalo de la recta real.

## 2.2. FUNCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA.

Supongamos que hemos lanzado 240 veces un dado perfecto y hemos obtenido los siguientes resultados:

| Cara | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Nº de veces | 40 | 39 | 42 | 38 | 42 | 39 |

Las frecuencias, absolutas y relativas de los resultados obtenidos son:

| RESULTADOS OBTENIDOS | | |
|---|---|---|
| **CARA** | **FRECUENCIA ABSOLUTA** | **FRECUENCIA RELATIVA** |
| 1 | 40 | $\frac{40}{240}=0'1667$ |
| 2 | 39 | $\frac{39}{240}=0'1625$ |
| 3 | 42 | $\frac{42}{240}=0'175$ |
| 4 | 38 | $\frac{38}{240}=0'1583$ |
| 5 | 42 | $\frac{42}{240}=0'175$ |
| 6 | 39 | $\frac{39}{240}=0'1625$ |
| SUMA | 240 | 1 |
| DISTRIBUCIÓN DE LA FRECUENCIA | | |

Sabemos, por el tema anterior de probabilidad, que cuando el número de repeticiones del experimento aumenta, la frecuencia relativa tiende a estabilizarse en torno a un número que denominamos probabilidad del suceso (ley de los grandes números). Pues bien, esto podemos aplicarlo trasladando la probabilidad de los sucesos asociados a un valor de la variable aleatoria y definiendo lo que llamamos **función de probabilidad**.

Así, para el ejemplo anterior, dada la variable aleatoria X= "Resultado al lanzar un dado", que toma los valores $\{1,2,3,4,5,6\}$, la probabilidad de cada uno de ellos forma lo que denominamos su función de probabilidad, que en este caso sería:

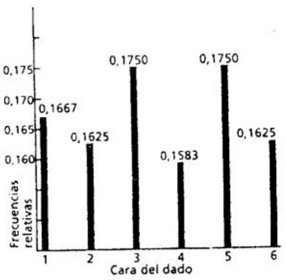
| RESULTADOS ESPERADOS | | |
|---|---|---|
| CARA | No DE VECES | PROBABILIDAD |
| 1 | 40 | $\frac{1}{6}$ |
| 2 | 40 | $\frac{1}{6}$ |
| 3 | 40 | $\frac{1}{6}$ |
| 4 | 40 | $\frac{1}{6}$ |
| 5 | 40 | $\frac{1}{6}$ |
| 6 | 40 | $\frac{1}{6}$ |
| SUMA | 240 | 1 |

## DISTRIBUCION DE PROBABILIDAD

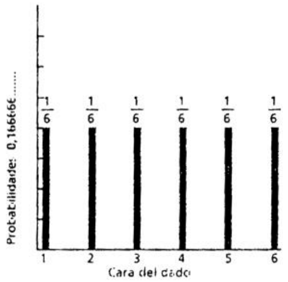
De manera que podemos definir: Dada una variable aleatoria $X$ que toma los valores $x_1, x_2, x_3, \cdots, x_n$, llamamos **función de probabilidad de *X*** a la aplicación que asocia a cada valor $x_i$ su probabilidad:

$$p_i = p\left(X = x_i\right) \quad \forall i = 1, 2, 3, \cdots, n$$

En las siguientes ilustraciones vemos la representación gráfica tanto de la distribución de frecuencia como de la distribución de probabilidad. Recordemos una vez más que la distribución de frecuencias se parecerá más a la distribución de probabilidad cuanto mayor sea el número de veces que se repite el experimento.




Distribución de frecuencias.




Distribución de probabilidad.

Pero es importante que sepamos distinguir lo que llamamos "resultados obtenidos" de los denominados "resultados esperados". Los primeros corresponden a la distribución de frecuencias, esto es, a los resultados concretos obtenidos a partir de una muestra concreta, variarán por tanto si realizamos un nuevo experimento; en cambio, los segundos, corresponden a la distribución de probabilidad, son los resultados teóricos que cabe esperar, de manera que los resultados obtenidos se parecerán más a éstos cuanto mayor sea el número de veces que se repite el experimento o el tamaño de la muestra.

**Propiedad:**

En toda función de probabilidad se verifica que $p_1 + p_2 + p_3 + \cdots + p_n = 1$ ya que se trata de la suma de probabilidades de un espacio completo de sucesos.

**Ejemplo 3:**

Consideremos el experimento aleatorio que consiste en lanzar simultáneamente tres monedas distintas. Sea la variable aleatoria X= "Número de caras que aparecen". Su distribución de probabilidad es la siguiente:

1º) El espacio muestral es $E = \left\{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\right\}$.

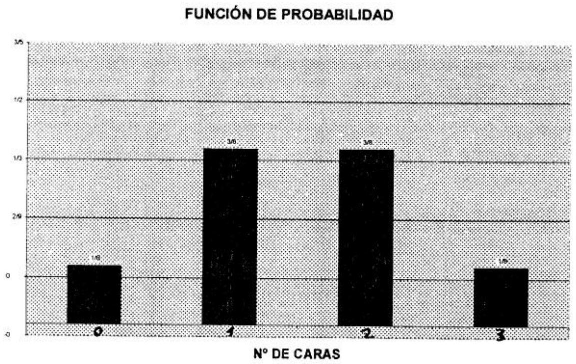
2º) Los valores que toma la variable son: 0, 1, 2, 3.

La función de probabilidad está formada por la probabilidad de estos valores:

$$p(0\text{ caras}) = p(\{XXX\}) = \frac{1}{8} \qquad p(1\text{ cara}) = p(\{CXX, XCX, XXC\}) = \frac{3}{8}$$

$$p(2\text{ caras}) = p(\{CCX, CXC, XCC\}) = \frac{3}{8} \qquad p(3\text{ caras}) = p(\{CCC\}) = \frac{1}{8}$$

Su gráfica es:




## 2.3. MEDIA Y VARIANZA DE UNA VARIABLE ALEATORIA DISCRETA.

Consideremos el ejemplo anterior del lanzamiento de un dado. Si nos fijamos en la distribución experimental, tenemos una variable que toma una serie, de valores cada uno con una determinada frecuencia absoluta y relativa. Estamos en condiciones, pues, de calcular los parámetros estadísticos correspondientes a esta distribución, en particular, la media aritmética, la varianza y la desviación típica.

Para ello no tenemos más que añadir las siguientes columnas a la tabla de la distribución experimental (recordar lo estudiado en cursos anteriores):

| **RESULTADOS OBTENIDOS (Distribución experimental)** | | | | | |
|---|---|---|---|---|---|
| **CARA** $(x_i)$ | **FREC. ABS.** | **FREC. REL.** $(h_i)$ | $x_i \cdot h_i$ | $x_i^2$ | $x_i^2 \cdot h_i$ |
| 1 | 40 | $\frac{40}{240} = 0'1667$ | 0’1667 | 1 | 0’1667 |
| 2 | 39 | $\frac{39}{240} = 0'1625$ | 0’325 | 4 | 0’65 |
| 3 | 42 | $\frac{42}{240} = 0'175$ | 0’525 | 9 | 1’575 |
| 4 | 38 | $\frac{38}{240} = 0'1583$ | 0’6332 | 16 | 2’5328 |
| 5 | 42 | $\frac{42}{240} = 0'175$ | 0’875 | 25 | 4’375 |

| | | | | | |
|---|---|---|---|---|---|
| 6 | 39 | $\frac{39}{240}=0'1625$ | 0'975 | 36 | 5'85 |
| SUMA | 240 | 1 | 3'4999 | | 15'1495 |
| DISTRIBUCIÓN DE LA FRECUENCIA | | | | | |

De manera que la media de la distribución experimental es $\overline{x}=\sum_{i=1}^{6} x_i \cdot h_i = 3'4999$.

Su varianza: $s^2=\sum_{i=1}^{6} x_i^2 \cdot h_i - \overline{x}^2 = 15'1495 - 3'4999^2 = 2'9002$, y su desviación típica: $s=\sqrt{2'9002}=1'7029$

Observación:

Para calcular los parámetros anteriores hemos utilizado la frecuencia relativa en vez de usar la frecuencia absoluta y luego dividir entre el número de elementos porque la operación es la misma pero así el proceso es más rápido. Es equivalente multiplicar los datos por la frecuencia absoluta y luego dividir por N que multiplicar directamente por la frecuencia relativa.

Pues bien, actuando de manera análoga sobre la tabla de la función de probabilidad tendremos la media aritmética, la varianza y la desviación típica de la variable aleatoria (distribución teórica).
En este caso, la probabilidad de cada resultado es el equivalente a la frecuencia relativa, de modo que si en la tabla de la distribución de probabilidad (resultados esperados) añadimos las siguientes columnas, tendremos calculados los parámetros estadísticos:

| **RESULTADOS ESPERADOS** | | | | | |
|---|---|---|---|---|---|
| **CARA** $(x_i)$ | **Nº DE VECES** | **PROB.** $(p_i)$ | $x_i \cdot p_i$ | $x_i^2$ | $x_i^2 \cdot p_i$ |
| 1 | 40 | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 | $\frac{1}{6}$ |
| 2 | 40 | $\frac{1}{6}$ | $\frac{2}{6}=\frac{1}{3}$ | 4 | $\frac{4}{6}=\frac{2}{3}$ |
| 3 | 40 | $\frac{1}{6}$ | $\frac{3}{6}=\frac{1}{2}$ | 9 | $\frac{9}{6}=\frac{3}{2}$ |
| 4 | 40 | $\frac{1}{6}$ | $\frac{4}{6}=\frac{2}{3}$ | 16 | $\frac{16}{6}=\frac{8}{3}$ |
| 5 | 40 | $\frac{1}{6}$ | $\frac{5}{6}$ | 25 | $\frac{25}{6}$ |
| 6 | 40 | $\frac{1}{6}$ | $\frac{6}{6}=1$ | 36 | $\frac{36}{6}=6$ |
| SUMA | 240 | 1 | $\frac{21}{6}$ | | $\frac{91}{6}$ |
| DISTRIBUCIÓN DE PROBABILIDAD | | | | | |

La media aritmética de la distribución teórica será $\mu = \sum_{i=1}^{6} x_i \cdot p_i = \frac{21}{6} = 3'5$.

Su varianza $\sigma^2 = \sum_{i=1}^{6} x_i^2 \cdot p_i - \mu^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = 15'167 - 12'25 = 2'917$

Y su desviación típica $\sigma = \sqrt{2'917}$

**Observación:**

Como podemos ver, los parámetros de la distribución experimental no coinciden con los de la distribución teórica. Es muy importante saber distinguirlos, pues unas veces necesitaremos unos y otras veces los otros.

Cuanto mayor sea el número de elementos del espacio muestral o más veces se repita el experimento, más se aproximarán los parámetros experimentales a los teóricos pero no coincidirán. Los parámetros experimentales, $\overline{x}$ y $s$, se calculan a partir de los datos de una muestra, mientras que $\mu$ y $\sigma$ se hallan a partir de la probabilidad (teórica) de la distribución. A $\mu$ también la llamamos **esperanza matemática**, pues es el valor medio esperado a partir de los datos probabilísticos que disponemos.
Ejercicio:

Sea X una variable aleatoria cuya función de probabilidad es:

| | | | | | | |
|---|---|---|---|---|---|---|
| *x* | 0 | 1 | 2 | 3 | 4 | 5 |
| *p* | 0’1 | 0’2 | 0’1 | 0’4 | 0’1 | 0’1 |

a) Representa gráficamente la función de probabilidad.
b) Calcula los parámetros de la distribución.
c) Halla las siguientes probabilidades: $p(X \le 4'5)$, $p(X \ge 3)$ y $p(3 \le X < 4'5)$

## 2.4. FUNCIÓN DE DISTRIBUCIÓN DE UNA VARIABLE ALEATORIA DISCRETA.

El último apartado del ejercicio anterior nos sirve para darnos cuenta que muchas veces no necesitamos calcular probabilidades puntuales en un cierto valor, sino que lo que nos interesa es la probabilidad de un cierto intervalo Así, por ejemplo, si estamos jugando con un dado y hemos apostado a que sale un número menor o igual que 2, lo que nos hace falta conocer es la $p(X \le 2)$.

Esto da pie a que definamos lo que denominamos función de distribución de una variable aleatoria discreta:

Dada una variable aleatoria *X* que toma los valores $x_1, x_2, x_3, \cdots, x_n$, llamamos función de distribución y la representamos por la letra *F* a la función que viene definida por:

$$F(a) = p\left(X \leq a\right), \text{ para cualquier valor que demos a "}a\text{".}$$

**Ejemplo 4:**

Dada la variable aleatoria *X* = "Número de veces que sale cara en el lanzamiento de dos monedas", su función de distribución será:

$$F(0) = p\left(X \leq 0\right) = p\left(X = 0\right) = \frac{1}{4}$$

$$F(1) = p\left(X \leq 1\right) = p\left(X = 0\right) + p\left(X = 1\right) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

$$F(2) = p\left(X \leq 2\right) = p\left(X = 0\right) + p\left(X = 1\right) + p\left(X = 2\right) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$$

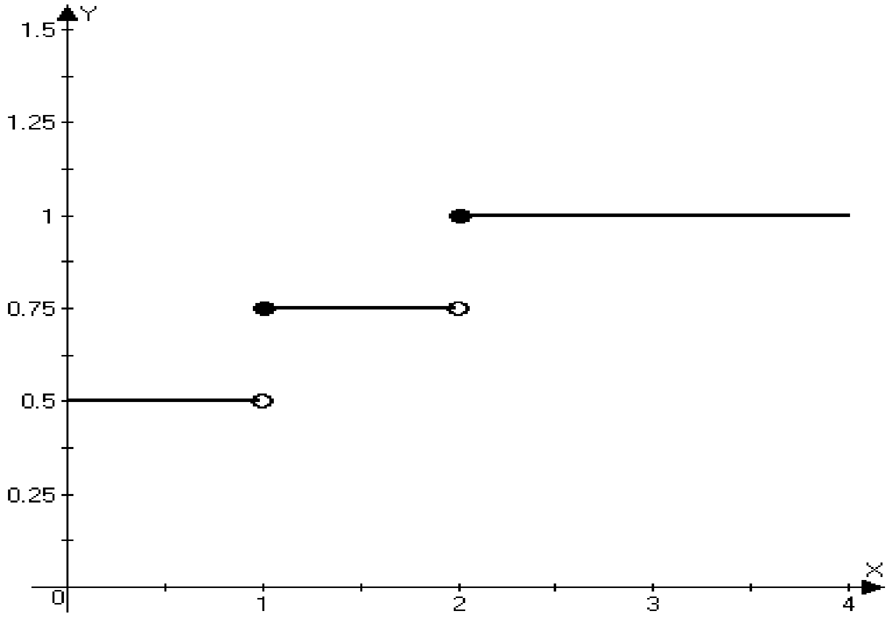
La función también puede tomar valores intermedios:

$$F(1'7) = p\left(X \leq 1'7\right) = p\left(X = 0\right) + p\left(X = 1\right) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

Podemos hacer una tabla para representarla:

| *X* | 0 | 0'5 | 0'7 | 0'9 | 1 | 1'5 | 1'7 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|
| *F(X)* | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | 1 | 1 |

Su gráfica es:




Como podemos observar es una función definida a trozos escalonada. Es creciente, puesto que vamos añadiendo, probabilidades, y el mayor máximo que toma es 1, puesto que este es el valor del suceso seguro.

**Ejercicio 2:**

Calcula y dibuja la función de probabilidad correspondiente a la variable aleatoria X=Suma de las caras en el lanzamiento de dos dado?.

## 3. LA DISTRIBUCIÓN BINOMIAL.

### 3.1. DEFINICIÓN.

Nos vamos a fijar en un experimento aleatorio en el que sólo haya dos sucesos elementales. A uno lo llamaremos “Éxito”, y al otro, su contrario, “Fracaso”.

$A = \text{"Éxito"}$, $\overline{A} = \text{"Fracaso"}$

Cada uno tendrá una probabilidad que llamaremos $p(A) = p$ y $p\left(\overline{A}\right) = q$ ($q = 1 - p$, pues ambos sucesos forman el espacio de sucesos y son incompatibles, luego la suma de sus probabilidades es 1)

Este tipo de experimentos es muy frecuente. Siempre que estemos estudiando la posibilidad de que ocurra un determinado suceso, este y su contrario se convertirían en los sucesos éxito y fracaso, respectivamente.

Por ejemplo, en el experimento “Número de asignaturas suspensas de un alumno de lº de bachillerato”, podemos considerar los sucesos:

Éxito: El alumno aprueba todas las asignaturas.
Fracaso: El alumno suspende al menos una asignatura.

0 también, en el lanzamiento de una moneda, si apostamos a que va a salir cara, tendríamos:

Éxito: Obtener cara.
Fracaso: Obtener cruz.

Si nos fijamos, estos experimentos podemos repetirlos el número de veces que nos interese:

Si una clase consta de 25 alumnos, podemos considerar cuantos éxitos ha habido, es decir, cuantos alumnos han aprobado todas las asignaturas.

Podemos lanzar cuatro veces un dado y estudiar cuántas veces ha salido cara.

Pues bien, cuando tenemos una variable de este tipo decimos que es una variable aleatoria o distribución binomial.

Es decir, una variable aleatoria es binomial cuando se repite un número, determinado de veces un experimento aleatorio con las siguientes características:

1) En cada prueba del experimento sólo hay dos posibles resultados: el suceso *A*, éxito, y su contrario, el suceso $\overline{A}$, fracaso.
2) El resultado obtenido en cada prueba es independiente de los anteriores.
3) La probabilidad del suceso *A* se mantiene constante, no varía de una prueba a otra. Representaremos por $p$ a esta probabilidad, y por $q$ a la de su contrario fracaso ($q = 1 - p$).

Toda variable aleatoria binomial es discreta, ya que mide el número de éxitos al realizar n veces un experimento.

Al repetir n veces un experimento, la variable podrá tomar los valores 0, 1, 2, 3,..., n (pues podrá haber desde 0 hasta n éxitos).

**Ejemplo 5:**

Tenemos una moneda perfecta y la lanzamos 4 veces. La variable aleatoria *X*="Número de caras que aparecen" es una variable binomial.

En efecto, al lanzar una moneda, tenemos dos posibilidades: *A*="Salir cara," con $p(A) = \frac{1}{2}$, y $\overline{A}$="Salir cruz", con $p(\overline{A}) = \frac{1}{2}$.

Y este experimento se realiza cuatro veces.

Luego la variable aleatoria *X*="Número de veces que aparece cara al lanzar cuatro veces una moneda", es una variable binomial que toma los valores 0, 1, 2, 3 y 4.

Como la probabilidad de éxito es $p = \frac{1}{2}$ y el experimento se realiza cuatro veces ($n = 4$), diremos que esta variable es una variable binomial de parámetros $n = 4$ y $p = \frac{1}{2}$, y lo notaremos escribiendo $X \sim B\left(n = 4, p = \frac{1}{2}\right)$

**<u>Ejemplo 6:</u>**

En una ciudad el porcentaje de fumadores es del 35%. Se escoge una muestra de 100 individuos esa población y se considera la variable *X*="Número de fumadores". Este es otro ejemplo de variable binomial.

En efecto:

1) Los sucesos éxito y fracaso son, respectivamente: *A*="Ser fumador", con $p(A) = 0'35 = p$ y $\overline{A}$="No ser fumador", con $p(\overline{A}) = 0'65 = q$
2) Cada prueba es cada una de las personas que forma la encuesta. El resultado de cada una de ellas es independiente, pues el hecho de que una persona fume o no, no influye en lo que haga la siguiente persona.
3) La probabilidad de ser fumador no varía de unas personas a otras, se mantiene constante cada vez que se elige una persona de la población.

Como $p = 0'35$ y la muestra es $n = 100$, la variable *X*="Número de fumadores en una muestra de cien personas", sigue una distribución binomial de parámetros 0'35 y 100, $X \sim B(100, 0'35)$

## 3.2. FUNCIÓN DE PROBABILIDAD DE UNA VARIABLE BINOMIAL.

Vamos a estudiar a partir de un ejemplo sencillo cuál es la ley o fórmula que sigue la función de probabilidad de una variable aleatoria binomial.

Supongamos para ello que tenemos un dado perfecto y apostamos a que sale un múltiplo de tres. Si lo lanzamos cuatro veces, tenemos:

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

*A* = Éxito = "Salir múltiplo de tres". $\rightarrow p = p(A) = \frac{2}{6} = \frac{1}{3}$

$B = \overline{A} =$ Fracaso = "No salir múltiplo de tres". $q = p(\overline{A}) = \frac{2}{3}$

Si se realiza el experimento cuatro veces, tenemos las siguientes opciones:

- Ningún múltiplo de tres: $X = 0$ B, B, B, B

  $$p(X = 0) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \left(\frac{2}{3}\right)^4$$

- Un múltiplo de tres: $X = 1$ A, B, B, B<br>B, A, B, B<br>B, B, A, B<br>B, B, B, A

  Estos sucesos tienen la misma probabilidad, $\left(\frac{1}{3}\right) \cdot \left(\frac{2}{3}\right)^3$, luego

  $$p(X = 1) = 4 \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^3$$

- Dos múltiplos de tres: $X = 2$ A, A, B,, B<br>A, B, A, B<br>A, B, B, A<br>B, A, A, B<br>B, A, B, A<br>B, B, A, A

  Estos sucesos tienen la misma probabilidad, $\left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2$, y como se repiten seis veces, resulta que $p(X = 2) = 6 \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2$

- Tres múltiplos de tres: $X = 3$ A, A, A, B<br>A, A, B, A<br>A, B, A, A<br>B, A, A, A

  Estos sucesos tienen la misma probabilidad, $\left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)$, y como se repiten cuatro veces, resulta que $p(X = 3) = 4 \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)$

- Cuatro múltiplos de tres: $X = 4$ A, A, A, A

  $$p(X = 4) = \left(\frac{1}{3}\right)^4$$

De manera que la función de distribución de esta variable binomial $X \sim B\left(4, \frac{1}{3}\right)$ es:

$$p(X=0)=\left(\frac{2}{3}\right)^4$$

$$p(X=1)=4\cdot\frac{1}{3}\cdot\left(\frac{2}{3}\right)^3$$

$$p(X=2)=6\cdot\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^2$$

$$p(X=3)=4\cdot\left(\frac{1}{3}\right)^3\cdot\left(\frac{2}{3}\right)$$

$$p(X=4)=\left(\frac{1}{3}\right)^4$$

Como podemos observar, es fácil deducir una ley o fórmula que nos permita generalizar los resultados anteriores, pues en las probabilidades aparecen *p* y *q* elevados, respectivamente, al número de éxitos o de fracasos correspondientes.

Lo único que "es extraño" es el número que aparece multiplicando al principio. Si nos fijamos en ellos, 4 son las formas de ordenar un éxito entre cuatro experimentos, 6 son las formas de ordenar dos éxitos entre cuatro experimentos, 4 son las formas de ordenar tres éxitos entre cuatro experimentos,...

Son lo que llamamos **números combinatorios,** que no hemos estudiado, y que tienen una fórmula para calcularse:

$$\binom{n}{r}=\frac{n!}{r!\cdot(n-r)!}$$

También, con el llamado **triángulo de Pascal,** podemos calcular estos números. Cada número es suma de los dos que se encuentran inmediatamente encima de él, a izquierda y derecha. La suma de los elementos de cada fila es la suma de casos posibles. Así, si una familia que tenga un hijo tiene dos casos posibles: niño y niña; si tiene dos hijos, hay cuatro casos posibles: dos niños, dos niñas, niño-niña y niña-niño, etc.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | | 1 | | | | | | |
| | | | | | 1 | | 2 | | 1 | | | | | |
| | | | | 1 | | 3 | | 3 | | 1 | | | | |
| | | | 1 | | 4 | | 6 | | 4 | | 1 | | | |
| | | 1 | | 5 | | 10 | | 10 | | 5 | | 1 | | |
| | 1 | | 6 | | 15 | | 20 | | 15 | | 6 | | 1 | |
| 1 | | 7 | | 21 | | 35 | | 35 | | 21 | | 7 | | 1 |

Triángulo de Tartaglia o de Pascal

De esta manera, la expresión de la función de probabilidad de una distribución binomial de parámetros *n* y *p* sería:

$$\boxed{p(\text{obtener } r \text{ éxitos}) = p(X = r) = \binom{n}{r} p^r q^{n-r}}$$

**<u>Ejemplo 7:</u>**

Si la probabilidad de que al nacer un niño sea varón es 0'48, calcula la función de probabilidad para el número de hijas en cuatro nacimientos.

Tenemos la variable aleatoria *X*= "Número de niñas de cuatro hijos", que es una binomial de parámetros $n = 4$ y $p = 0'52$

Considerando Éxito=*A*="Nacer niña" $(p = 0'52)$ y Fracaso=$\overline{A}$="Nacer niño", $(q = 0'48)$ y fijándonos en el triángulo de Pascal para calcular los números combinatorios, tenemos la siguiente función de probabilidad:

Probabilidad de que no nazcan niñas: $p(X = 0) = 1\cdot(0'52)^0 (0'48)^4 = 0'053$

Probabilidad de que nazca una niña: $p(X = 1) = 4\cdot(0'52)^1 \cdot (0'48)^3 = 0'23$

Probabilidad de que nazcan dos niñas $p(X = 2) = 6\cdot(0'52)^2 \cdot (0'48)^2 = 0.3738$

Probabilidad de que nazcan tres niñas $p(X = 3) = 4\cdot(0'52)^3 \cdot (0'48)^1 = 0'27$

Probabilidad de que nazcan tres niñas $p(X = 4) = 1\cdot(0'52)^4 \cdot (0'48)^0 = 0'073$

**<u>Ejercicio 3:</u>**

La probabilidad de que un alumno llegue tarde al instituto es 0'08. Si hacemos el cómputo de una semana:

a) ¿Cuál es la probabilidad de que llegue dos días tarde?

b) ¿Y de que sea todos los días puntual?

**Ejercicio 4:**

Tenemos un dado trucado en el que la probabilidad de cara es doble que la de cruz. Si lo lanzamos cinco veces, ¿a qué apostarías?

a) Salir tres caras.
b) Salir menos de tres cruces.

## 3.3. USO DE TABLAS DE LA DISTRIBUCIÓN BINOMIAL.

Aunque tengamos la fórmula anterior, realizar los cálculos probabilísticos puede resultar un poco largo. Para simplificarlos, tenemos la tabla de la distribución binomial que nos permite calcular las probabilidades para determinados valores de los parámetros.

En efecto, si nos fijamos en la tabla, podemos observar que aparece una columna a la izquierda con los datos correspondientes a *n* (número de veces que se realiza el experimento) y a *r* (número de casos favorables a éxito, o valor que toma *X*), y una fila en la parte superior con distintos valores de probabilidad (*p*). Cuando tengamos que buscar las probabilidades correspondientes a una distribución binomial, tendremos que ver cuales corresponden a los parámetros de la misma en esa tabla. Veámoslo con un ejemplo:

La probabilidad de que un alumno llegue tarde a clase es 0’15. Si nos fijamos en una semana (cinco días lectivos), calculemos las siguientes probabilidades:

a) Llegar dos días tarde.
b) Llegar siempre puntual.
c) Llegar siempre tarde.

La distribución es binomial de parámetros $n=5$ y $p=0'15$, siendo *X*=“Nº de días que llega tarde”. $X \sim B(5,0'15)$

Para calcular las probabilidades con la tabla, nos fijamos en la columna correspondiente a $p=0'15$ y en las filas correspondientes a $n=5$. Tenemos:

| n | r | p | 0.15 |
|---|---|---|---|
| 5 | 0 | | 0.4437 |
| | 1 | | 0.3915 |
| | 2 | | 0.1382 |
| | 3 | | 0.0244 |
| | 4 | | 0.0022 |
| | 5 | | 0.0001 |

De manera que las probabilidades pedidas son:

a) El suceso dos días tarde es $X=2$, luego su probabilidad es $p\left(X=2\right)=0'1382$
b) Llegar siempre puntual es $X=0$, por tanto tenemos $p\left(X=0\right)=0'4437$ e) Llegar siempre tarde es $X=5$, que ocurre con probabilidad $p\left(X=5\right)=0'0001$

**Ejercicio 5:**

Lanzamos un dado perfecto diez veces. Si apostamos a que sale múltiplo de tres:

a) ¿Cuál es la probabilidad de que acertemos cuatro veces?
b) ¿Y de que acertemos al menos dos veces?

**Observación:**

Como hemos visto en el último apartado del ejercicio anterior, no vamos a calcular únicamente probabilidades de un solo valor de *X*, sino que muchas veces nos preguntarán por probabilidades de intervalos (utilizando desigualdades). Veamos cómo resolverlo a partir del siguiente ejemplo:

Una prueba de inteligencia está compuesta por diez preguntas, cada una de las cuales tiene cuatro respuestas, siendo sólo una de ellas correcta. Un alumno tiene prisa por acabar la prueba y decide contestar a lo loco, es decir, aleatoriamente. Calculemos:

a) Probabilidad de contestar exactamente cuatro preguntas correctamente:

   De lo primero que hemos de percatamos es que tenemos una distribución binomial en la que Éxito = $A$ = "Contestar la pregunta correctamente" $\left(p\left(A\right)=p=0'25\right)$ y Fracaso= $\overline{A}$ ="Contestar equivocadamente la pregunta" $\left(p\left(\overline{A}\right)=q=0'75\right)$ y repetimos el experimento diez veces, luego *X*="Nº de respuestas correctas al contestar diez preguntas" es una distribución binomial $B\left(10,0'25\right)$.
   Contestar 4 preguntas correctas es que $X=4$, y mirando en la tabla, $p\left(X=4\right)=0'1460$

b) Probabilidad de no acertar ninguna:

   Es $p\left(X=0\right)=0'0563$

c) Probabilidad de aprobar:

Aprobará cuando conteste correctamente cinco o más preguntas, luego hemos de hallar $p(X \geq 5)$

$$p(X \geq 5) = p(X = 5) + p(X = 6) + p(X = 7) + p(X = 8) + p(X = 9) +$$
$$+ p(X = 10)$$

Consultando la tabla, tenemos:

$$p(X \geq 5) = 0'0584 + 0'0162 + 0'0031 + 0'0004 + 0.0001 + 0 = 0.0782$$

0 sea, contestando aleatoriamente no llega al 8% la posibilidad de aprobar.

**Importante:**

La última probabilidad calculada, $p(X = 10)$, aparece como 0 en la tabla. Esto no es totalmente cierto. Si realizamos los cálculos con la fórmula, obtenemos:

$$p(X = 10) = 1 \cdot (0'25)^{10} \cdot (0'75)^{0} = 9'8 \cdot 10^{-14} = 0'000000000000098$$

Es decir, teóricamente, tiene una cierta probabilidad de salir, pero que es prácticamente cero.

d) Probabilidad de obtener menos de tres preguntas correctas:

Es
$$p(X < 3) = p(X = 0) + p(X = 1) + p(X = 2) = 0'0563 + 0'1877 + 0'2816 =$$
$$= 0'5256$$

e) Probabilidad de suspender:

Es $p(X < 5)$. Como este es el suceso contrario de aprobar, su probabilidad será:

$$p(X < 5) = 1 - p(X \geq 5) = 1 - 0'0782 = 0'9218$$

Cada cual podrá sacar las consecuencias oportunas.

## 3.4. MEDIA Y VARIANZA DE LA DISTRIBUCIÓN BINOMIAL.

Como cualquier otra variable aleatoria, la binomial tiene su esperanza y varianza, que calcularemos a partir de la tabla de la distribución de probabilidad. Para simplificar los cálculos, vamos a calcular la media de una binomial de parámetros $n=3$ y $p$, y después generalizaremos para cualquier otro tamaño.

Sea, pues, $X \sim B(3, p)$. Fijándonos en el triángulo de Pascal, la función de probabilidad es:

$$p(X=0)=1\cdot q^3=q^3$$
$$p(X=1)=3\cdot p\cdot q^2$$
$$p(X=2)=3\cdot p^2\cdot q$$
$$p(X=3)=1\cdot p^3=p^3$$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 1 | | 1 | | | |
| | | 1 | | 2 | | 1 | | |
| | 1 | | 3 | | 3 | | 1 | |
| 1 | | 4 | | 6 | | 4 | | 1 |

Como la esperanza matemática es $\mu=\sum_{i=1}^{n} x_i\cdot p_i$, en este caso nos queda que $\mu=0\cdot q^3+1\cdot 3\cdot p\cdot q^2+2\cdot 3\cdot p^2\cdot q+3\cdot p^3=3p\left(q^2+2pq+p^2\right)=3p\left(p+q\right)^2=3p\cdot 1=3p$

(El resultado del paréntesis es 1, pues si nos fijamos en el triángulo de Pascal, corresponde a la suma de las probabilidades de una binomial para $n=2$).

Luego vemos que para $X \sim B(3, p)$, su esperanza matemática es $\mu=3p$, es decir, hemos multiplicado el número de veces que se repite el experimento por la probabilidad de éxito. En general se verifica:

Si $X \sim B(n, p)$ su esperanza matemática es $\boxed{\mu=n\cdot p}$

Razonando análogamente, añadiendo los cálculos necesarios, llegamos a que su varianza es:

$$\boxed{\sigma^2=n\cdot p\cdot q}$$

Ejemplo:

Lanzamos cinco veces una moneda perfecta. Consideramos la variable aleatoria $X$="Número de veces que aparece cara". Esta variable es una binomial $X \sim B\left(5,\frac{1}{2}\right)$; su media es $\mu=n\cdot p=5\cdot\frac{1}{2}=\frac{5}{2}=2'5$ y su varianza es $\sigma^2=n\cdot p\cdot q=5\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{5}{4}=1'25$.

## 3.5. AJUSTE DE UNA DISTRIBUCIÓN EMPÍRICA MEDIANTE UNA DISTRIBUCIÓN BINOMIAL.

Habrá ocasiones en las que tengamos un conjunto de datos (por ejemplo los resultados de una encuesta) y no sepamos si corresponden o no a una variable binomial. Para saberlo, en estos casos lo que vamos a hacer es lo que denominamos ajustar la distribución empírica (datos reales) mediante una distribución binomial (datos teóricos).

Para ello seguimos los siguientes pasos:

1) Comprobamos si los datos cumplen las características de una variable binomial.

2) Calculamos la media de la distribución empírica (media muestral).

3) Ajustamos como media de la distribución binomial la media muestral, $\mu = \overline{x}$, y como $\mu = n \cdot p$, de aquí inferimos el valor de $p$ despejando en la anterior expresión.

4) De los cálculos anteriores tenemos ya definida la variable $B(n, p)$, así que calculamos su función de probabilidad.

5) Comparamos los datos reales de frecuencia relativa con los anteriores de probabilidades teóricas. Si son similares significa que la distribución empírica se ajusta como binomial. Si son distintos el ajuste daría lugar a error.

6) Podemos calcular qué frecuencia absoluta correspondería a los datos teóricos y compararlas con las de los datos de la muestra.

Ejemplo:

De quinientas unidades de un producto farmacéutico, tal que las unidades llevan dos precintos, se han observado las frecuencias de precintos rotos que vienen dadas por la siguiente tabla:

| Nº de precintos rotos | Nº de unidades |
|---|---|
| 0 | 190 |
| 1 | 220 |
| 2 | 90 |

Ajustar, si es posible, mediante una distribución binomial, y calcular las frecuencias absolutas teóricas.

1) Veamos si se cumplen los requisitos de una distribución binomial:

i) En cada prueba sólo hay dos posibles resultados: precinto roto, precinto correcto.
ii) El resultado obtenido en cada prueba es independiente de los resultados anteriores (el que el precinto de un producto esté roto no influye en lo que le ocurra a los restantes).
iii) La probabilidad del suceso precinto roto es constante y, por lo tanto, no varía de una prueba a otra. Por tanto, se verifican todas las condiciones.

2) Calculamos la media muestral:

| Nº de precintos rotos $(x_i)$ | Nº de unidades $(f_i)$ | frec. rel. $(h_i)$ | $x_i \cdot h_i$ |
|---|---|---|---|
| 0 | 190 | 0.38 | 0 |
| 1 | 220 | 0.44 | 0.44 |
| 2 | 90 | 0.18 | 0.36 |
| suma | 500 | 1 | 0.8 |

$$\overline{x} = \sum_{i=1}^{3} x_i \cdot h_i = 0'8$$

3) Calculamos los parámetros de la distribución binomial:

Como la variable toma los valores 0, 1, 2, tenemos que $n = 2$.
Estimamos $\mu$ como $\overline{x}$: $\mu = 0'8$

Teniendo en cuenta que $\mu = 2p$, nos queda que $0'8 = 2p \Rightarrow p = \frac{0'8}{2} = 0'4$

Por tanto la distribución teórica es $X \sim B(2, 0'4)$

4) Su función de probabilidad es:

$p(X = 0) = 1 \cdot 0'4^0 \cdot 0'6^2 = 0'36$ $\quad$ $p(X = 1) = 2 \cdot 0'4 \cdot 0'6 = 0'48$

$p(X = 2) = 1 \cdot 0'4^2 \cdot 0'6^0 = 0'16$

5) Comparemos la distribución de frecuencias relativas (datos empíricos) con la distribución de probabilidad de la binomial (datos teóricos):

| Distribución empírica | | Distribución teórica | |
|---|---|---|---|
| Nº precintos rotos $(x_i)$ | frec. rel. $(h_i)$ | Nº precintos rotos $(x_i)$ | frec. rel. $(p_i)$ |
| 0 | 0.38 | 0 | 0.36 |
| 1 | 0.44 | 1 | 0.48 |
| 2 | 0.18 | 2 | 0.16 |

| | | | |
|---|---|---|---|
| | 1 | | 1 |

Como vemos, las frecuencias relativas están muy próximas a las probabilidades correspondientes, de manera que es correcto ajustar la distribución empírica mediante una binomial.

6) Las frecuencias absolutas teóricas serán:

Como el número de datos es 500, para calcular el número de elementos correspondientes a cada suceso multiplicamos 500 por su probabilidad:

| Distribución teórica | | |
|---|---|---|
| Nº de precintos rotos $(x_i)$ | Probabilidad $(p_i)$ | Nº de unidades (frec. abs.) $(f_i)$ |
| 0 | 0.36 | 500-0.36=180 |
| 1 | 0.48 | 500-0.48=240 |
| 2 | 0.16 | 500-0.16=80 |

Como vemos, las frecuencias absolutas teóricas varían poco de las muestrales, lo que da idea de la bondad del ajuste.

**Ejercicio:**

El departamento de control de calidad de una fábrica de aparatos de televisión realiza cuatro controles. De seiscientos televisores se han obtenido los siguientes datos:

| Nº de fallos | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Nº de televisores | 136 | 219 | 58 | 6 | 1 |

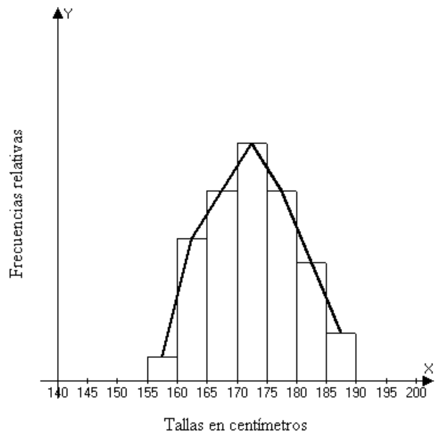
Comprueba si es prudente ajustar esta distribución empírica a una distribución binomial y halla las frecuencias absolutas esperadas.

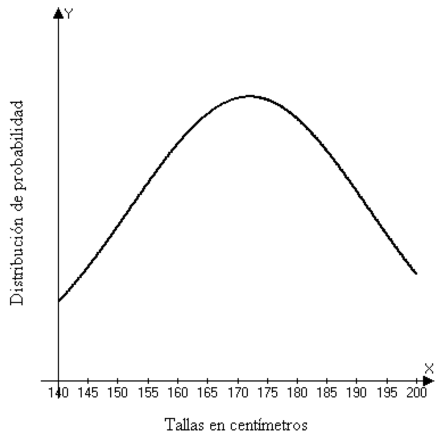
## 4. VARIABLES ALEATORIAS CONTINUAS.

### 4.1. FUNCIÓN DE DENSIDAD DE UNA VARIABLE ALEATORIA CONTINUA.

Hemos definido las variables aleatorias continuas como aquellas que pueden tomar cualquier valor en un intervalo. Ya en cursos pasados hemos estudiado variables continuas. Un caso típico es el estudio de las alturas de una población. En este caso, dada una muestra (datos experimentales) la distribución de frecuencias la representábamos a partir de un histograma, en el que la variable aleatoria venía dada en intervalos.

En el ejemplo que representamos a continuación, hemos medido las tallas de cuarenta alumnos varones de un centro escolar de 1º de bachillerato y hemos obtenido que se distribuyen según el histograma de frecuencias relativas de la izquierda. Pues bien, si eligiésemos una muestra lo suficientemente grande, representativa de toda la población de varones en ese curso en España, y fuésemos tomando intervalos de amplitud más pequeños, entonces la distribución de frecuencias se iría acercando cada vez más a la gráfica de la derecha.




El diagrama representado a la derecha nos da una idea de lo que denominamos **distribución de probabilidad continua**.

Es decir, la función de densidad de una variable aleatoria continua es la gráfica a la que tiende el histograma de frecuencias relativas cuando la amplitud de los intervalos va haciéndose cada vez más pequeña. A esa función la vamos a llamar **función de densidad.**

Observación 1:

De igual modo que la suma de frecuencias relativas de una distribución de frecuencias es 1, también la suma de todas las probabilidades de una distribución de probabilidad es 1. Esto se traduce en que el área encerrada bajo la curva de la función de densidad es 1.

Observación 2:

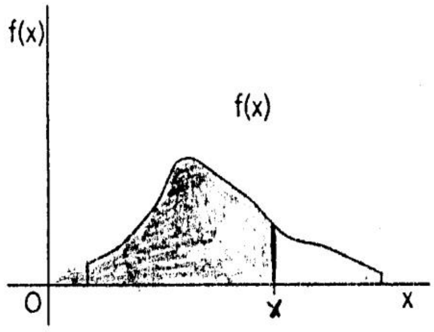
En el caso de variables aleatorias continuas, no tiene sentido hablar de la probabilidad de un punto, pues va a ser siempre cero. Lo que tendrá interés será calcular la probabilidad de un intervalo.

## 4.2. FUNCIÓN DE DISTRIBUCIÓN DE UNA VARIABLE ALEATORIA CONTINUA.

La idea de función de distribución en variables aleatorias continuas es la misma que la de función de distribución de una variable discreta. Medirá la probabilidad de que la variable tome un valor menor o igual que uno dado.

Es decir, la función de distribución de una V. A. *X* continua es $F(x) = p\left(X \leq x\right)$

Hemos dicho anteriormente que el área bajo la curva de la función de densidad es 1. Esto significa que el área de la figura representa la probabilidad. Gráficamente, esto nos lleva a que, si nos fijamos en un intervalo, su probabilidad será el área comprendida en ese rectángulo. Por tanto, la función de distribución de *x*, *F(x)*, indica el área bajo la función de densidad a la izquierda de *x*.



v. aleatoria continua

En la figura de la izquierda, observamos que la función de distribución de *x* es el área bajo la curva en el intervalo $\left(-\infty, x\right)$.

En cursos posteriores estudiaremos cómo hallar la expresión analítica de esta función Lo que sí es importante es que no olvidemos que toda función de densidad *f(x)* verifica:

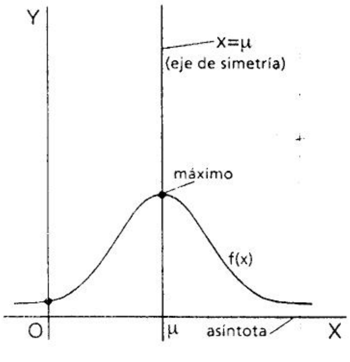
1) $f(x) \geq 0 \quad \forall x$
2) El área encerrada bajo su gráfica es 1.
3) El área bajo cualquier rectángulo es la probabilidad de dicho intervalo.

## 5. LA DISTRIBUCIÓN NORMAL.

### 5.1. INTRODUCCIÓN.

Denominamos como normal a la distribución continua de probabilidad más común.

Recibe este nombre porque en un principio se pensó que todos los fenómenos de la naturaleza venían regidos por esta ley de probabilidad. Su función de densidad recibe el nombre de **Campana de Gauss,** siendo simétrica respecto a la media.



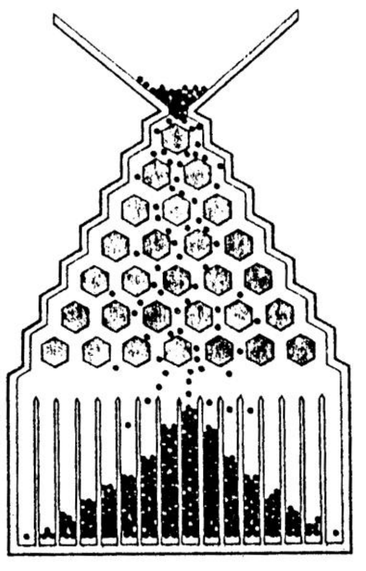

Como podemos observar en la figura, las normales son distribuciones que se reparten uniforme y simétricamente alrededor de la media, es decir, los valores con mayor probabilidad son los centrales, cuanto más lejos estén de la media menor será su probabilidad, siendo además simétrica respecto a dicha media.

La primera aproximación a la distribución normal la realizó **Francis Galton** (1822-1911), quien construyó un ingenioso dispositivo en el que experimentalmente obtenía la curva de la distribución normal.

El aparato, que aparece a la derecha, constaba de un tablero inclinado sobre el que estaban distribuidos regularmente un sistema de clavos. Éstos, permiten deslizar un gran número de bolas que proceden de un depósito situado en la parte superior. Las bolas, al chocar con los clavos, se alejan en mayor o menor medida de la línea central de caída según la ley del azar. Recogiendo estas bolas en compartimentos estrechos, distribuidos a lo largo del borde inferior del tablero, las alturas que alcanzan las bolas en las distintas columnas dan una idea bastante clara de la distribución normal.

**Este es el aparato original hecho por Galton en 1873 con la Empresa Tisley&Spiller**

Actualmente, existe un aparato análogo al que ideó Galton, denominado *demostrador mecánico de probabilidad,* construido por el Science Materials Center, bajo el nombre patentado de Hexstat. Las bolas de este depósito caen del depósito superior y pasan por una serie de obstáculos, siendo recogidos finalmente en unos compartimentos situados en la parte inferior.



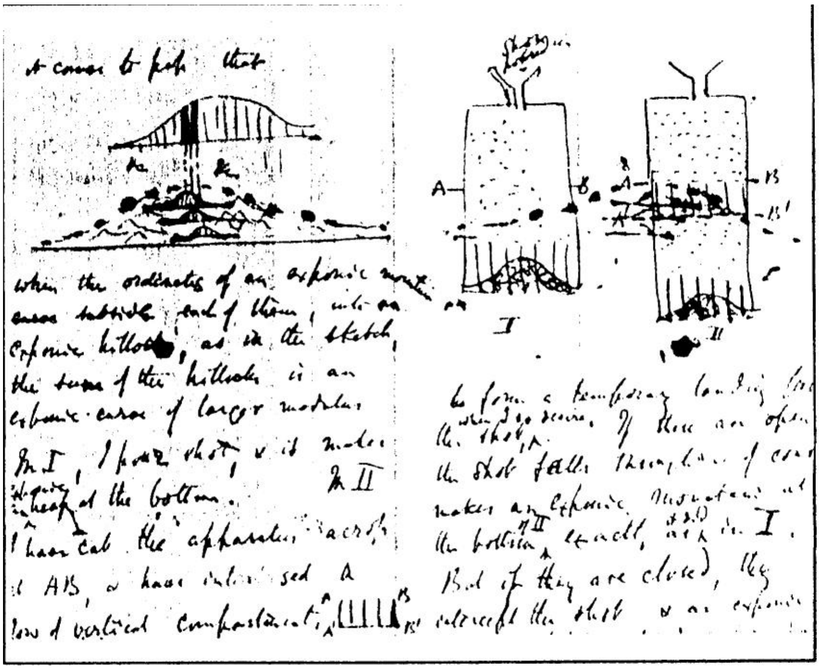
Hexstat construido por el Science Materials Center

En cada obstáculo, la probabilidad, de que la bola vaya a la izquierda es igual a la probabilidad de que vaya a la derecha, es decir, 1/2.

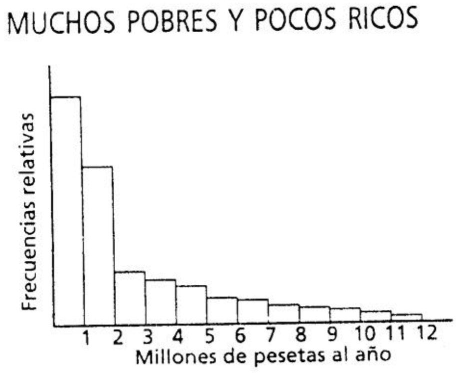
Este aparato proporciona, por vía empírica, una buena aproximación a la distribución normal.

Podemos ver en la página siguiente una carta enviada por Galton a su primo George Darwin el 12 de enero de 1877 en la que le explica los resultados obtenidos sobre la distribución normal.

En efecto, hay muchos fenómenos de la vida real y de la naturaleza cuyo histograma de frecuencias se corresponde a la ley normal. Por ejemplo, la distribución del peso de los individuos de cualquier especie, y existen multitud de fenómenos estudiados por la psicología, pedagogía, biología, etc., relativos a medidas psicométricas. También hay otros muchos fenómenos que responden a una distribución simétrica y que son suma de una serie de efectos parciales independientes que, independientemente, no se ajusten a una distribución normal, pero el fenómeno resultante sí puede tender a una distribución normal.



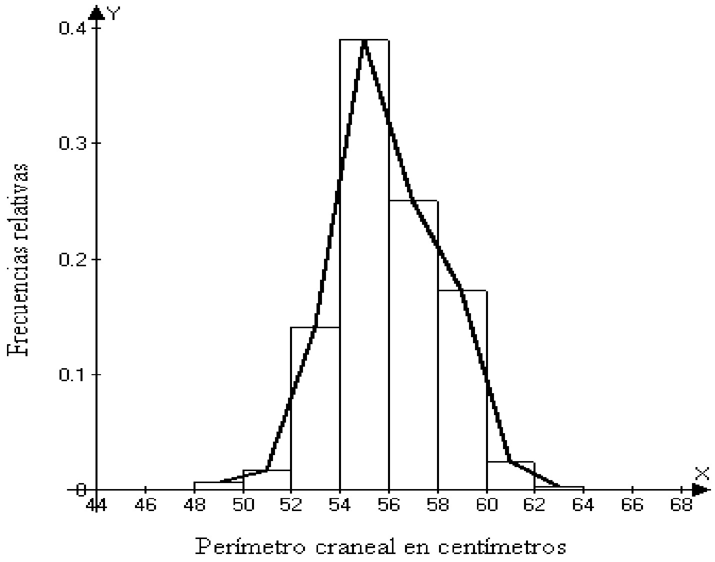
Sin embargo, hay otros casos de distribuciones que no se ajustan al modelo normal. En efecto, si observemos el ejemplo de la figura siguiente. Hemos clasificado a los ciudadanos españoles según el nivel de renta. Evidentemente, son muy pocos los españoles que poseen niveles de rentas altas, y en cambio son muchos los que poseen niveles de rentas bajas. Por tanto, la distribución no es simétrica y, en consecuencia, no se adapta a un modelo normal.




*Esta distribución no es normal.*

## 5.2. FUNCIÓN DE DENSIDAD DE LAS DISTRIBUCIONES NORMALES.

El histograma de la figura muestra la distribución de quinientos estudiantes de un instituto respecto a su perímetro craneal, que viene reflejado en la tabla de la derecha.




| Perímetro craneal (cm) | Número de alumnos |
|---|---|
| $[48, 50)$ | 3 |
| $[50, 52)$ | 8 |
| $[52, 54)$ | 70 |
| $[54, 56)$ | 195 |
| $[56, 58)$ | 125 |
| $[58, 60)$ | 86 |
| $[60, 62)$ | 12 |
| $[62, 64)$ | 1 |

Si tomamos una muestra cada vez mayor, y vamos haciendo la amplitud de los intervalos cada vez más pequeña, el histograma irá acercándose a la gráfica de la función de densidad de la distribución normal.

Esta función de densidad tiene la siguiente expresión analítica:

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

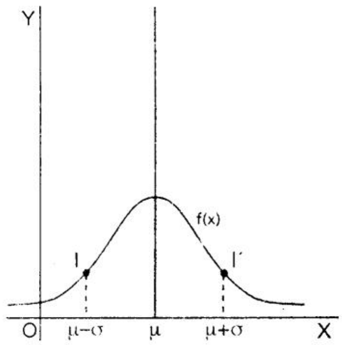
Donde:

- $\mu$ es su media y $\sigma$ su desviación típica.
- $x$, es la variable aleatoria, que puede tomar cualquier valor de $(-\infty, +\infty)$. Es decir, el dominio es $\mathbb{R}$
- $\pi = 3'141592\cdots$ y $e = 2'718281\cdots$

La expresión de la función de densidad es algo complicada y no vamos a necesitar utilizarla. Para dibujarla, podemos utilizar la máquina de Galton. Lo que nos importan son sus características:

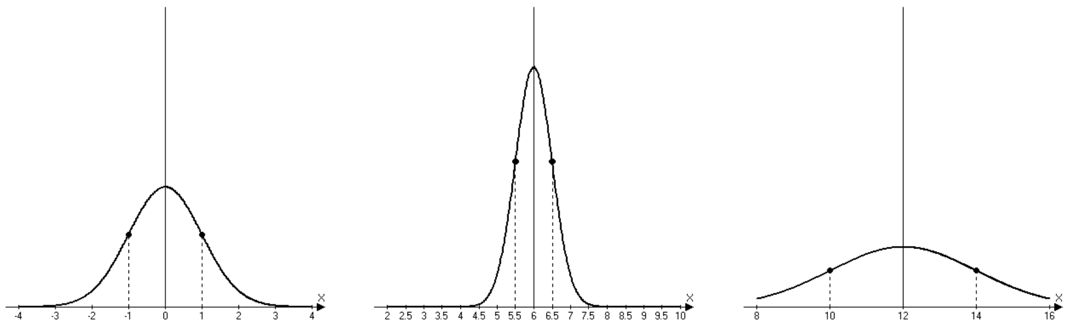
a. Dominio: Todos los números reales.
b. Es simétrica respecto a la recta $x = \mu$
c. Asíntota horizontal: El eje OX.
   Nótese que la ordenada vale prácticamente 0 para valores alejados de la media, tanto a su derecha como a su izquierda.
d. Crece en $(-\infty, \mu)$ y decrece en $(\mu, +\infty)$

e. Por tanto, el máximo se da para $x = \mu$.
f. El área encerrada bajo la curva es 1.




Dependiendo de los valores de la media y la desviación típica, la gráfica de la función de densidad de una distribución normal varía. Cuanto mayor sea a, habrá más dispersión, y los datos estarán más separados, luego la gráfica será más estilizada y más abierta. Podemos verlo en las siguientes gráficas:

$\mu = 0, \sigma = 1$ $\mu = 6, \sigma = 0'5$ $\mu = 12, \sigma = 2$



Cuando una variable aleatoria $X$ sea normal, estará definida por los valores de su media $\mu$ y su desviación típica $\sigma$, de manera que escribiremos que $X \sim N(\mu, \sigma)$.

Es importante que nos percatemos que existen infinitas distribuciones normales, dependiendo del valor de la media y la varianza. Evidentemente, según el fenómeno que estemos estudiando, los parámetros muestrales tomarán unos valores y unas unidades distintos. La forma de la gráfica será similar, pero no la misma (centrada en distinto lugar, más o menos dispersa,...).

De todas estas distribuciones, la más sencilla es la que tiene por media 0 y desviación típica 1. Es la distribución $N(0,1)$. Su función de densidad queda bastante simplificada:

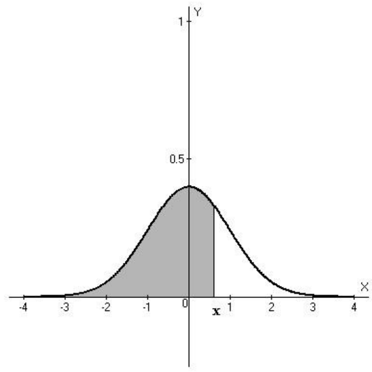
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Cuya representación es que sigue:

La función de distribución *F(x)*, sería el área del recinto sombreado bajo la curva.

Es decir, que la probabilidad $p(X \le x)$ vendría dada por dicha área.

Para facilitar su cálculo, existen tablas que lo facilitan, pues si no, tendríamos que hacer uso del cálculo integral. Veremos más adelante cómo se utilizan.




## 5.3. MANEJO DE LAS TABLAS DE UNA VARIABLE $X \sim N(0,1)$

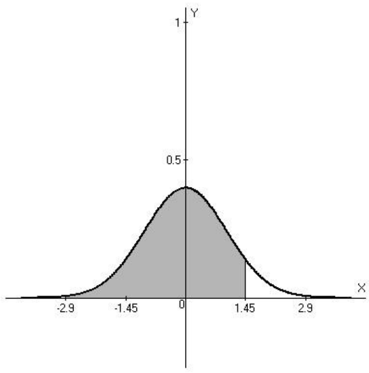
Sea $Z \sim N(0,1)$. Veamos con varios ejemplos, como utilizar la tabla de la normal para calcular algunas probabilidades:

a. $p(Z \le 1'45)$

La probabilidad pedida es igual al área sombreada de la figura de la derecha, y se encuentra en la tabla con sólo buscar 1'4 en la columna y 0'05 en la fila; su intersección nos da la probabilidad 0'9265.

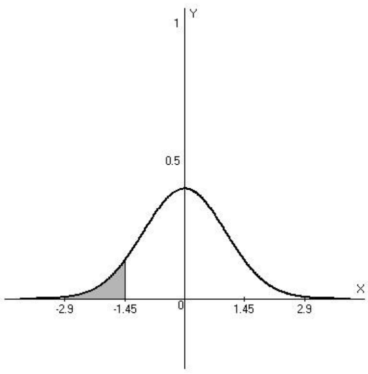
Es decir, $p(Z \le 1'45) = 0'9265$

Esto quiere decir que el 92'65 % de los datos se encuentran en el intervalo $(-\infty, 1'45)$.




b. $p(Z \le -1'45)$

La probabilidad pedida es igual al área sombreada de la figura de la derecha. La tabla sólo ofrece probabilidades para valores positivos de Z, pero teniendo en cuenta la simetría de la función de densidad, y que el área bajo la curva es igual a la unidad, resulta que:

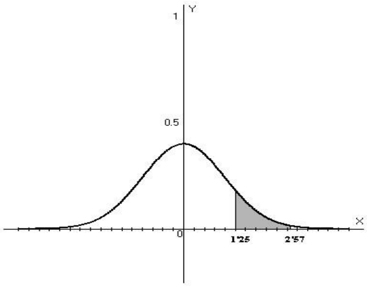



$$p(Z \le -1'45) = p(Z > 1'45) = 1 - p(Z \le 1'45) = 1 - 0'9265 = 0'0735$$

c. $p(1'25 < Z \leq 2'57)$

La probabilidad pedida es igual al área sombreada de la figura de la derecha.

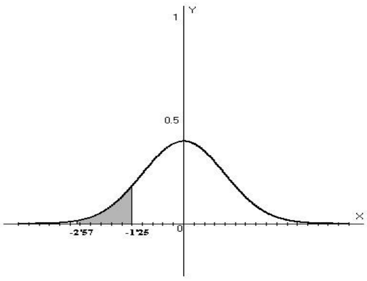
Su cálculo lo realizaremos restando al área mayor la menor:




$p(1'25 < Z \leq 2'57) = p(Z \leq 2'57) - p(Z < 1'25) =$
$= 0'9949 - 0'8944 = 0'1005$

d. $p(-2'57 < Z \leq -1'25)$

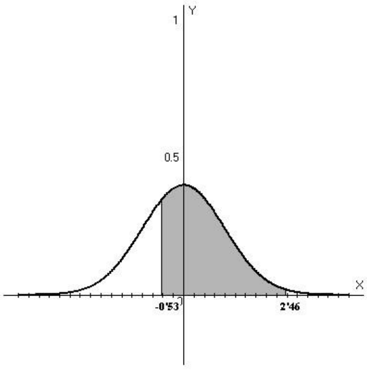
La probabilidad pedida es igual al área sombreada de la figura de la derecha, y como consecuencia de la simetría de la función tenemos:




$p(-2'57 < Z \leq -1'25) = p(1'25 < Z \leq 2'57) =$
$= 0'1005$

e. $p(-0'53 < Z \leq 2'46)$

La probabilidad pedida es igual al área sombreada de la figura de la derecha.




$p(-0'53 < Z \leq 2'46) =$
$= p(Z \leq 2'46) - p(Z \leq -0'53) =$
$= p(Z \leq 2'46) - p(Z > 0'53) =$
$= p(Z \leq 2'46) - (1 - p(Z \leq 0'53)) =$
$= 0'9931 - (1 - 0'7019) = 0'695$

Lo que significa que el 69'5 % de las observaciones se encuentran entre -0'53 y 2.46.

Cualquier otro caso que se pueda presentar, se podrá reducir a alguno de los anteriores.

Ejercicio:

Si $X \sim N(0,1)$, calcula las siguientes probabilidades:

a. $p(X \leq 1'32)$
b. $p(0 \leq X \leq 0'25)$
c. $p(-2'23 \leq X \leq 1'15)$

## 5.4. TIPIFICACIÓN DE UNA VARIABLE NORMAL.

Hemos dicho que existen infinitas distribuciones normales, dependiendo del valor de la media y de la desviación típica. También hemos comentado que la más fácil de manejar es la normal 0-1, para la cual tenemos tablas que nos ayudarán a calcular probabilidades. Evidentemente, lo que no podemos tener es una tabla para cada una de las infinitas distribuciones normales que existen.

Para solucionar este problema, lo que vamos a hacer, cuando el fenómeno que estemos estudiando sea normal pero no sea 0-1 es hacer una transformación que dé lugar a una $N(0,1)$.

Este proceso es lo que llamamos **tipificar la variable**.

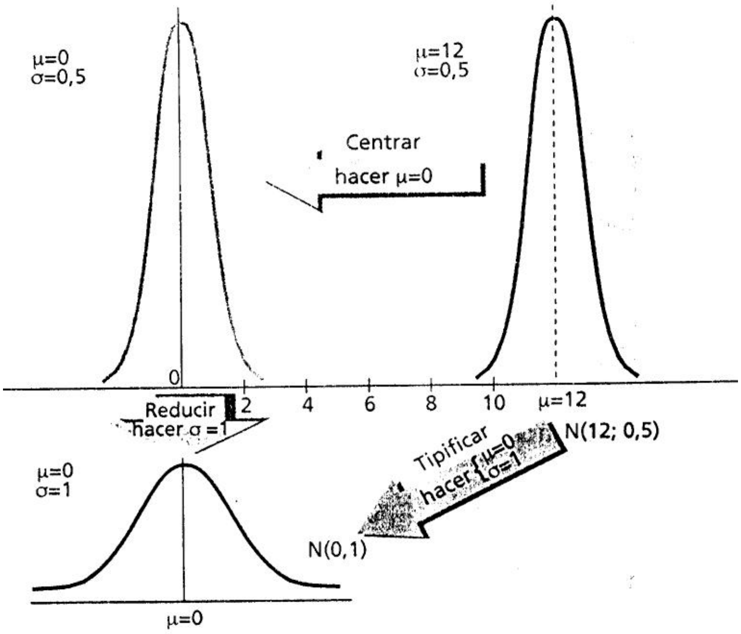
El proceso es el siguiente:

a) Tenemos una variable $X \sim N(\mu,\sigma)$.
b) Construimos, a partir de la anterior, otra variable *Z* que sea $N(0,1)$.
c) Calculamos las probabilidades de la variable *X* a partir de las probabilidades de la variable *Z*.

La variable *Z* va a ser $Z = \dfrac{X-\mu}{\sigma}$, que es $N(0,1)$

Las probabilidades para *Z* las sabemos calcular a partir de la tabla de la normal tipificada, y de aquí deduciremos las probabilidades para *X*, ya que ambas probabilidades coinciden.

En el esquema que puedes ver a continuación se observan los dos pasos en el proceso de tipificar una variable:



a)Centrarla: Trasladamos la media de la distribución al origen de coordenadas.
b)Reducir la desviación típica a 1: Esto equivale a dilatar o contraer la gráfica de la distribución.

Caso práctico 1:

El peso de los alumnos de secundaria de una localidad de la Sierra de Córdoba se distribuye normalmente con media 70 Kg. y desviación típica 6 Kg. Sabiendo que hay 2.000 estudiantes:

a) ¿Qué probabilidad hay de que pesen menos de 76 Kg?
b) ¿Qué número de alumnos se espera que pese menos de 76 Kg?
c) ¿Qué probabilidad tenemos de que un alumno pese entre 64 y 76 Kg?
d) ¿Cuántos alumnos esperamos que haya comprendidos entre estos pesos?

Lo primero es proceder a tipificar la variable. Si llamamos *X*="Peso de los alumnos de secundaria", $X \sim N(70, 6)$

La variable tipificada es $Z = \dfrac{X-\mu}{\sigma}$ y $Z \sim N(0,1)$

a) Si $X = 76$, entonces $Z = \dfrac{76-70}{6} = \dfrac{6}{6} = 1$, de manera que $p(X < 76) = p(Z < 1)$

Esta última probabilidad, como es de una distribución normal 0-1, podemos calcularla utilizando la tabla correspondiente.

$p(X < 76) = p(Z < 1) = 0'8413$

En términos porcentuales, el 84'13 % de los alumnos pesa menos de 76 kg.

b) Si $p(X < 76) = 0'8413$ y hay 2.000 alumnos, entonces se espera que pesen menos de 76 kg un total de $2000 \cdot 0'8413 = 1682'6 \simeq 1683$ alumnos.

c) La probabilidad pedida es $p(64 < X < 76) = p(X < 76) - p(X < 64)$
El primer término lo conocemos; calculemos el segundo. Volvemos a tipificar:
Si $X = 64$, $Z = \dfrac{64-70}{6} = \dfrac{-6}{6} = -1$, luego $p(X < 64) = p(Z < -1)$
$p(Z < -1)$ no aparece en la tabla, pero por simetría, $p(Z < -1) = p(Z > 1)$
$p(Z > 1) = 1 - p(Z \le 1) = 1 - 0'8413 = 0'1587$, de donde:

$p(X<64)=p(Z<-1)=p(Z>1)=1-p(Z\leq 1)=1-0'8413=0'1587$, con lo que:
$p(64<X<76)=p(X<76)-p(X<64)=0'8413-0'1587=0'6826$

d) Y el número de alumnos que se espera estén comprendidos entre 76 y 64 Kg. es de $2000\cdot 0'6826=1365'2\simeq 1365$ alumnos.

Caso práctico 2:

Otras veces, lo que nos pedirá el ejercicio será que calculemos el intervalo de valores de la variable que tiene una determinada probabilidad, es decir, el proceso inverso al planteado en el ejercicio anterior:

Hemos aplicado un test a cuatrocientos alumnos y ha resultado que las puntuaciones obtenidas siguen una distribución normal de media 60 puntos y desviación típica 5 puntos. ¿A partir de qué valor se encuentra el 67 % de las notas más bajas?

Si llamamos *X* = "Puntuación obtenida", $X \sim N(60,5)$

El enunciado nos dice que $p(X<x)=0'67$, y hemos de calcular el valor de *x*.

Tipificando $p(X<x)=p\left(Z<\frac{x-60}{5}\right)=0'67$, donde $Z \sim N(0,1)$

Buscando en las tablas, tenemos que $\frac{x-60}{5}=0'44$ y, despejando, resulta que $x=60+5\cdot 0'44=62'2$ puntos.

Por consiguiente, el 67 % de los alumnos ha conseguido menos de 62.2 puntos en el examen.

Ejercicio:

La duración media de un lavavajillas es de 15 años con una desviación típica de 0'5 años. Si la vida útil del electrodoméstico se distribuye normalmente, calcula:

a) La probabilidad de que al comprar un lavavajillas dure menos de 17 años.
b) La probabilidad de que dure más de 16 años.
c) La probabilidad de que dure entre 13 y 18 años.
d) ¿Cuál es la vida útil máxima del 70% de los lavavajillas que menos duran?

**Observación:**

Al operar con desigualdades, nos es indiferente utilizar el signo $<$ o $\leq$, puesto que la probabilidad de un punto en distribuciones continuas es 0.

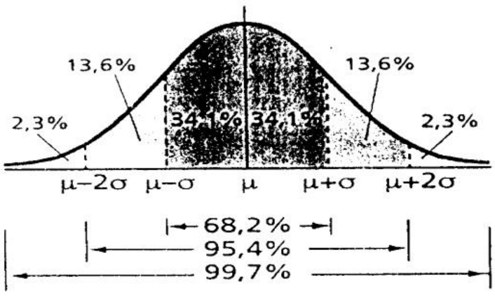
$p(X<5)=p(X\le 5)$ $\quad\quad$ $p(X>1)=p(X\ge 1)$

## 5.5. ALGUNOS CASOS PARTICULARES DE INTERVALOS CON UNA PROBABILIDAD DETERMINADA.

Hemos visto en la pregunta anterior que a veces lo que nos interesa no es calcular la probabilidad de un determinado intervalo, sino a qué intervalo le corresponde una determinada probabilidad.

Teniendo en cuenta que la distribución normal es simétrica respecto a la media, utilizando las tablas llegamos a las siguientes conclusiones:

1) El 68’26% de las observaciones está en el intervalo $(\mu-\sigma,\mu+\sigma)$
2) El 95’4% de las observaciones está en el intervalo $(\mu-2\sigma,\mu+2\sigma)$
3) El 99’7 % de las observaciones está en el intervalo $(\mu-3\sigma,\mu+3\sigma)$




*Proporción de individuos que se distribuyen en intervalos de la forma $(\mu - k\sigma,\ \mu + k\sigma)$.*

En la figura anterior podemos observar los resultados anteriores de proporción de individuos en intervalos simétricos de la forma $(\mu-k\sigma,\mu+k\sigma)$

Obsérvese que fuera del intervalo $(\mu-3\sigma,\mu+3\sigma)$ no quedan prácticamente individuos: el 0’3%

Calculemos analíticamente los resultados del 1er apartado:

Teniendo en cuenta que si $X \sim N(\mu,\sigma)$, entonces $Z=\dfrac{X-\mu}{\sigma}$ sigue una $N(0,1)$, entonces:

$$p(\mu-\sigma<X<\mu+\sigma)=p\left(\frac{\mu-\sigma-\mu}{\sigma}<Z<\frac{\mu+\sigma-\mu}{\sigma}\right)=p(-1<Z<1)=$$

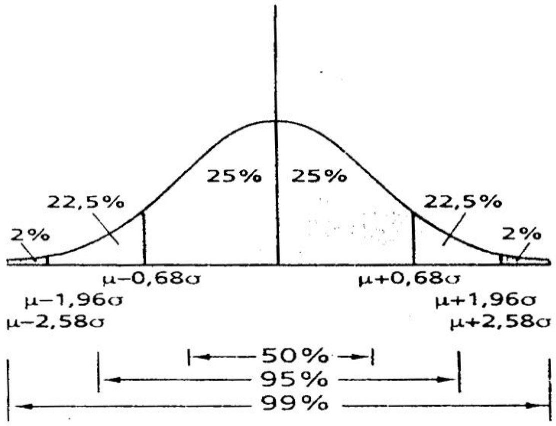
$$= p(Z < 1) - p(Z < -1) = 0'8413 - 0'1587 = 0'6826$$

Las otras probabilidades se calcularían análogamente.

También podemos realizar el proceso inverso, buscar en qué intervalos simétricos respecto a la media, se da una probabilidad determinada. Haciendo los cálculos necesarios, resulta que:

a) El 50% de los individuos está en el intervalo $(\mu - 0'68\sigma, \mu + 0'68\sigma)$.
b) El 95 % de los individuos está en el intervalo $(\mu - 1'96\sigma, \mu + 1'96\sigma)$.
c) El 99 % de los individuos está en el intervalo $(\mu - 2'58\sigma, \mu + 2'58\sigma)$

La figura de siguiente ilustra lo que estamos diciendo.




Ejemplo:

En la población del caso práctico 1, alumnos de un instituto que se distribuyen normalmente con media 70 y desviación típica 6, ¿entro qué pesos se encuentra, alrededor del peso medio, el 50 % de los alumnos?

Sin más que sustituir, el intervalo pedido es:

$$(\mu - 0'68\sigma, \mu + 0'68\sigma) = (70 - 0'68 \cdot 6, 70 + 0'68 \cdot 6) = (65'92, 74'8)$$

El 50% del alumnado pesa entre 65’92 kg y 74.8 kg

## 5.6. AJUSTE DE UNA DISTRIBUCIÓN EMPÍRICA MEDIANTE UNA DISTRIBUCIÓN NORMAL.

Como hemos dicho en un apartado anterior, frecuentemente, los datos recogidos en una muestra para realizar diversos estudios de tipo antropométrico (estatura, peso, diámetro craneal), morfológico (longitud de las hojas de un árbol, de las vainas de una plantación de guisantes), sociales (aceptación de una norma por parte de una comunidad), económico (impacto de un nuevo producto en el mercado), psicológicos (coeficiente intelectual de un conjunto de individuos) o físicos (distribución de errores en una serie de medidas) suelen distribuirse normalmente.

Sin embargo, esto no lo sabemos de antemano. Para estudiarlo, existe una primera aproximación que consiste en comprobar si la distribución experimental se ajusta a las propiedades de la distribución normal (distribución teórica). Para ello, lo más sencillo es dibujar el histograma de frecuencias relativas de la distribución experimental y comprobar si se ajusta a la forma de la campana de Gauss (simetría, distribución de valores alrededor de la media con valores que cumplan las condiciones de la normal [el 50% de las observaciones en el intervalo $\left(\mu-0'68\sigma,\mu+0'68\sigma\right)$,...).

También existen otros métodos matemáticos que predicen con mayor rigor si un conjunto de datos se ajusta a una distribución normal, como el llamado **test de normalidad de Kolmogorov,** que se estudiará en cursos posteriores. Nosotros, como primera aproximación, simplemente dibujaremos el histograma de frecuencias relativas y comprobaremos si se asemeja a la campana de Gauss.

Los pasos a seguir serán los siguientes:

1) Dibujamos el histograma de frecuencias relativas y comprobamos que se asemeja a la función de densidad de la distribución normal. En caso contrario, no tiene sentido hacer el ajuste.
2) Calculamos la media y desviación típica muestrales.
3) Ajustamos los datos como una $N\left(\overline{x},s\right)$ (distribución teórica)
4) Con la distribución anterior calculamos la probabilidad que nos pidan.

Caso práctico:

Se ha medido la distribución de una variable antropométrica, que representamos por X, de 50 hombres comprendidos entre 20 y 30 años, obteniéndose los resultados que muestra la tabla siguiente:

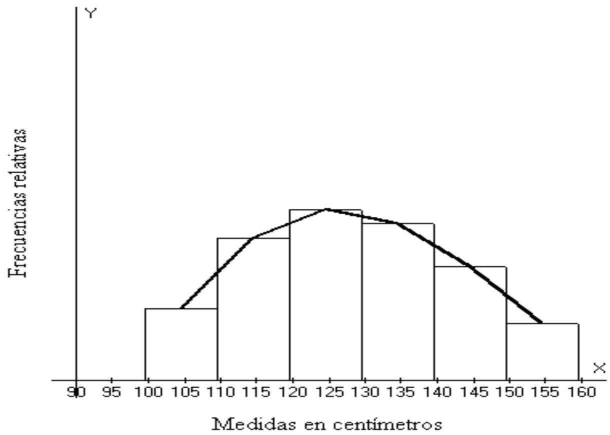
| Variable $X$ (en cm.) | Nº de hombres $\left(f_i\right)$ |
|---|---|
| [99’5, 109’5) | 5 |
| [109’5, 119’5) | 10 |
| [119’5, 129’5) | 12 |
| [129’5, 139’5) | 11 |
| [139’5, 149’5) | 8 |
| [149’5, 159’5) | 4 |
| SUMA | |

a) Comprobar si la anterior muestra se ajusta a una distribución normal.
b) En caso afirmativo, calcula la probabilidad de que un hombre tenga una medida antropométrica menor o igual que 105 cm.
c) Calcula también la probabilidad de que sea mayor que 125 cm.

a) Según lo que hemos dicho, lo primero es dibujar el histograma de frecuencias relativas. Para ello, añadimos la columna correspondiente a la tabla anterior.

| Variable $X$ (en cm) | Nº pers $(f_i)$ | Marca clase $(x_i)$ | frec. relat. $(h_i)$ | $x_i \cdot h_i$ | $x_i^2$ | $x_i^2 \cdot h_i$ |
|---|---|---|---|---|---|---|
| [99’5, 109’5) | 5 | 104’5 | 0’1 | 10’45 | 10920’25 | 1092’025 |
| [109’5, 112’5 | 10 | 114’5 | 0’2 | 22’9 | 13110’25 | 2622’05 |
| [119’5, 129’5) | 12 | 124’5 | 0’24 | 29’88 | 15500’25 | 3720’06 |
| [129’5, 139’5) | 11 | 134’5 | 0’22 | 29’59 | 18090’25 | 3979’855 |
| [139’5, 149’5) | 8 | 144’5 | 0’16 | 23’12 | 20880’25 | 3340’84 |
| [149’5, 159’5) | 4 | 154’5 | 0’08 | 12’36 | 23870’25 | 1909’62 |
| SUMA | N=50 | | 1 | 128’3 | | 16664’45 |

Representando obtenemos el siguiente histograma:




Que, como podemos observar nos recuerda la forma de la campana de Gauss, luego podemos ajustar la distribución experimental como normal.

b) El siguiente paso consiste en calcular la medía y varianza muestrales. Para ello hemos realizado los cálculos que aparecen en la tabla anterior. Como sabemos, tenemos que calcular la marca de clase de cada intervalo y con ella realizar cálculos análogos a los que hemos visto para variables discretas, de manera que:

$\overline{x} = \sum_{i=1}^{6} x_i h_i = 128'3$ cm. y $s^2 = \sum_{i=1}^{6} x_i^2 \cdot h_i - \overline{x}^2 = 16664'45 - 128'3^2 = 203'56$, de donde $s = \sqrt{203'56} = 14'26$

c) Con estos datos, la distribución teórica la ajustamos como $N(\bar{x}, s) = N(128'3, 14'26)$

d) La probabilidades pedidas las calculamos a partir de la anterior distribución tipificando tal como sabemos:

$Z = \frac{X - 128'3}{14'26}$, $Z \sim N(0,1)$, de manera que:

$$p(X \le 105) = p\left(Z \le \frac{105 - 128'3}{14'26}\right) = p(Z \le -1'63) = 1 - p(Z \le 1'63) =$$
$$= 1 - 0'9484 = 0'0516$$
$$p(X > 125) = 1 - p(X \le 125) = 1 - p\left(Z \le \frac{125 - 128'3}{14'26}\right) = 1 - p(Z \le -0'23) =$$
$$= 1 - p(Z > 0'23) = 1 - (1 - p(Z \le 0'23)) = p(Z \le 0'23) = 0'591$$

Ejercicio:

Las puntuaciones obtenidas por cien individuos sometidos a un test de habilidad manual vienen dadas en la siguiente tabla:

| $X$ | $f_i$ |
|---|---|
| 32 | 4 |
| 35 | 20 |
| 38 | 41 |
| 41 | 23 |
| 44 | 12 |

a. Comprueba que estos datos se ajustan a una distribución normal (se dan ya las marcas de clase, el primer intervalo sería $[30'5, 33'5)$.
b. Halla la probabilidad de que un individuo obtenga más de 40 puntos.
c. Halla la probabilidad de que un individuo obtenga menos de 39’5 puntos.
d. ¿Qué número de personas se espera que obtengan menos de 39.5 puntos?
e. Compara los datos experimentales y teóricos y deduce alguna conclusión.

## 6. APROXIMACIÓN DE UNA DISTRIBUCIÓN BINOMIAL MEDIANTE UNA NORMAL:

Como hemos podido comprobar experimentalmente, el uso práctico de la distribución binomial suele necesitar la realización de cálculos engorrosos, pues las tablas de las que disponemos sólo nos sirven para unos determinados valores de los parámetros, pero no para cualquiera de ellos.

Por ejemplo, si queremos calcular la probabilidad de que haya más de quince piezas defectuosas en una muestra de 110 piezas, con la condición de que la probabilidad de pieza defectuosa es 0'9, tendríamos una binomial $B(110, 0'09)$, y el cálculo pedido sería $p(X > 15) = \sum_{x=16}^{110} \binom{110}{x} \cdot (0'09)^x \cdot (0'91)^{110-x}$, y tendríamos que hacer 95 cálculos de paciente obtención.

Estos inconvenientes podemos superarlos si la distribución binomial $B(n, p)$ cumple las siguientes condiciones:

$$n \cdot p \geq 5 \qquad \text{y} \qquad n \cdot q \geq 5$$

Esto es así porque, como podemos comprobar en la siguiente gráfica, a medida que crece el número de pruebas realizadas, **la función de probabilidad de la binomial tiende a la curva de la función de densidad de la distribución normal**. La siguiente gráfica muestra la distribución del número de caras al ir aumentando el número de veces que lanzamos la moneda:
Como podemos observar, al lanzar treinta veces la moneda, la función de probabilidad de la binomial es muy parecida a la campana de Gauss.

Consecuencia:

Dada una variable aleatoria $X \sim B(n, p)$, si $n \cdot p \geq 5$ y $n \cdot q \geq 5$, entonces $X$ podemos aproximarla como normal con los siguientes parámetros:

$$X' \sim N\left(n \cdot p, \sqrt{n \cdot p \cdot q}\right)$$

(recordemos que la media de una distribución binomial es $n \cdot p$, y su varianza $n \cdot p \cdot q$, luego lo que estamos diciendo es que aproximamos la binomial por una normal con sus mismos parámetros)

Cuanto mayor sea *n* y más próximo a 0'5 sea *p*, mejor será la aproximación realizada.

Observación importante:

Una diferencia fundamental entre las distribuciones binomial y normal es que la primera es discreta y la segunda continua, de manera que la primera tiene probabilidad no nula para valores puntuales y la segunda no. Esto nos va a llevar a que, cuando aproximemos una binomial por una normal, para calcular probabilidades puntuales tomemos "intervalos pequeños" alrededor de ese punto. Es lo que llamamos corrección de continuidad, quedando del siguiente modo:

$$\boxed{p(X = k) = p(k - 0'5 \leq X' \leq k + 0'5)}$$

Y para otras situaciones:

$$\boxed{\begin{array}{l} p(X \le k) = p(X' \le k + 0'5) \\ p(X < k) = p(X' \le k - 0'5) \\ p(X > k) = p(X' \ge k + 0'5) \\ p(X \ge k) = p(X' \ge k - 0'5) \end{array}}$$

Ejemplo:

Se ha encuestado a la población masculina de cierto municipio, encontrándose que un 34% son hinchas del equipo de fútbol local. Elegidos cincuenta hombres al azar:

a) ¿Cuál es la probabilidad de que haya exactamente dieciocho aficionados?
b) ¿Y más de veinte?

La variable del problema *X*="Nº de personas que son hinchas del equipo de fútbol en una muestra de cincuenta" es Binomial. $X \sim B(50, 0'34)$

Comprobamos que podemos hacer el ajuste:

$$n \cdot p = 50 \cdot 0'34 = 17 \text{ y } n \cdot q = 50 \cdot 0'66 = 33$$

Como ambos valores son superiores a cinco, el ajuste es bueno.

$X' \sim N\left(n \cdot p, \sqrt{n \cdot p \cdot q}\right) = N\left(50 \cdot 0'34, \sqrt{50 \cdot 0'34 \cdot 0'66}\right) = N(17, 3'34)$.

a) $p(X = 18) = p(18 - 0'5 \le X' \le 18 + 0'5) = p(17'5 \le X' \le 18'5)$.

Tipificando:

$$p(17'5 \le X' \le 18'5) = p\left(\frac{17'5 - 17}{3'34} \le Z \le \frac{18'5 - 17}{3'34}\right) = p(0'15 \le Z \le 0'45) =$$
$$= p(Z \le 0'45) - p(Z \le 0'15) = 0'6736 - 0'5596 = 0'114$$

b) $p(X > 20) = p(X' \ge 20'5) = p\left(Z \ge \frac{20'5 - 17}{3'34}\right) = p(Z \ge 1'04) =$

$$= 1 - p(Z \le 1'04) = 1 - 0'8508 = 0'1492$$

Ejercicio:

Un examen contiene cuarenta preguntas del tipo verdadero-falso. El examen se aprueba si se contestan correctamente al menos 22 preguntas. Si un alumno responde a las preguntas lanzando una moneda perfecta, contesta:

a) Probabilidad de aprobar el examen.
b) Probabilidad de que el número de respuestas acertadas esté entre veinticinco y treinta.

# RELACIÓN DE EJERCICIOS DE DISTRIBUCIONES DE PROBABILIDAD

## VARIABLES ALEATORIAS DISCRETAS

1) Al lanzar dos dados trescientas cincuenta veces y observar la suma de las puntuaciones de sus caras superiores, se han obtenido los resultados siguientes:

| Suma total | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| frec. abs. | 13 | 16 | 31 | 37 | 46 | 61 | 49 | 35 | 29 | 21 | 12 | 350 |

   a) Calcula la distribución de probabilidad asociada a este experimento.
   b) Calcula la media y varianza muestrales.
   c) Calcula la media y varianza teóricas.
   d) Compara ambos resultados y obtén consecuencias.

2) Sea $X$ una variable aleatoria cuya función de probabilidad viene dada por $p(X=n)=\frac{1}{8}$, $n=2,3,\cdots,9$. Calcula:

   a) La función de probabilidad y su gráfica.
   b) La función de distribución y su gráfica.
   c) La media y la desviación típica.
   d) Las probabilidades $p(X\geq 6)$, $p(4<X\leq 7)$ y $p(X<-4)$
   e) ¿Por qué sabernos que la función definida en el apartado a es de probabilidad?

3) Se lanzan dos dados y se anota el número de caras que presentan números primos. Calcula:

   a) La distribución de probabilidad.
   b) La media y la desviación típica.

4) Una variable aleatoria discreta tiene la siguiente distribución de probabilidad:

| $x_i$ | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| $p(x_i)$ | $\frac{1}{9}$ | $\frac{1}{18}$ | $\frac{1}{3}$ | $\frac{5}{18}$ | $m$ | $\frac{1}{6}$ |

   a) Calcula *m*.
   b) Halla la media y la desviación típica.
   c) ¿Qué puedes decir de la dispersión?

5) Un dado ha sido manipulado con el fin de alterar las probabilidades de obtener las diferentes caras. Así, si $x$ representa la puntuación obtenida en una tirada, tenemos que:

$p(X=1)=\frac{1}{6}-2k \qquad p(X=2)=\frac{1}{6}-k \qquad p(X=3)=p(X=4)=\frac{1}{6}$

$p(X=5)=\frac{1}{6}-k \qquad p(X=6)=\frac{1}{6}-2k$

Determina $k$ para que la esperanza matemática de la variable sea 4.

6) Juzga la equidad de un juego que consiste en apostar 1 € y lanzar dos dados, ganándose 5 € al sacar parejas y nada en caso contrario. (Calcula su esperanza matemática).

7) Se ha pasado una prueba de fluidez verbal a un grupo de niños de una comarca socialmente deprimida, y se ha detectado que el 35% tiene una fluidez verbal prácticamente nula; el resto se puede considerar aceptable. De una muestra aleatoria formada por siete niños, calcula:

   a) La media y la varianza.
   b) La función de probabilidad.

8) Una determinada raza de perros tiene cuatro cachorros en una camada. Si la probabilidad de que un cachorro sea macho es 0'55, halla:

   a) La probabilidad de que en una camada haya dos hembras.
   b) La probabilidad de que en una camada al menos dos sean hembras.

9) Un jugador de tenis tiene una probabilidad de $\frac{1}{4}$ de ganar una partida. Si juega cuatro partidos, calcula la probabilidad de que gane más de la mitad.

10) Se lanza una moneda al aire cuatro veces. Halla la probabilidad de que salgan más caras que cruces.

11) Quinientos opositores han participado en una prueba escrita que consta de tres ejercicios. Los resultados obtenidos son los que constan en la siguiente tabla:

| Nº de ejercicios aprobados | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Nº de opositores | 136 | 223 | 120 | 121 |

Ajusta esta distribución empírica a una distribución binomial, halla las frecuencias teóricas esperadas e indica la bondad del ajuste.

12) Sobre una variable aleatoria $X$, se efectúan 1.000 observaciones, con los siguientes resultados:

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $f_i$ | 610 | 198 | 100 | 57 | 35 |

   a) Ajusta los datos a una distribución binomial.
   b) Comprueba la bondad del ajuste.

c) Compara la media y varianza teóricas con las experimentales. Explica a qué son debidas las diferencias.

**VARIABLES ALEATORIAS CONTINUAS**

13) Dada una variable aleatoria $X$ con distribución $N(0,1)$,calcula las probabilidades de los siguientes valores:

a) $X \geq 2$
b) $X < 0'8$
c) $X \leq 0'8$
d) $\frac{1}{3} \leq X < \frac{1}{2}$
e) $X \leq -0'7$
f) $-1'4 \leq X \leq -0'5$

14) Sea $X$ una variable aleatoria $N(0,1)$. Halla el valor de $a$ tal que:

a) $p(X \leq a) = 0'7967$
b) $p(0 \leq X \leq a) = 0'2291$
c) $p(a \leq X \leq 0'5) = 0'6107$

15) Dadas las siguientes distribuciones normales, calcula las probabilidades que se piden:

a) $N(3,1)$; $p(X \leq 5'5)$
b) $N(-4'3,1'5)$; $p(X > 0)$
c) $N(235,24)$; $p(220 \leq X \leq 225)$

16) Las notas de clase de una asignatura tienen una distribución $N(5,0)$. ¿Qué se puede decir de las notas de esa clase?

17) Las estaturas de los quinientos estudiantes de un instituto de secundaria tienen una distribución normal $N(1'58,0'12)$. Halla el número de estudiantes con estaturas entre:

a) 1’60 y 1’70
b) 1’45 y 1’60

18) Una máquina hace arandelas con un diámetro medio de 8 mm y una desviación típica de 0’5 mm. Suponiendo que la distribución es normal, ¿cuál es la probabilidad de que una pieza tomada al azar tenga un diámetro:

a) mayor que 8’5 mm?
b) menor que 7’5 mm?
c) mayor que 8’8 mm?

19) Un fabricante de ropa para jóvenes produce pantalones de cinco tallas destinados a individuos con estaturas entre 1’50 y 1’58 m; 1’58 y 1’66 m; 1’66 y 1’74 m; 1’74 y

1’82; 1’82 y 1’90 m. Suponiendo que la estatura de la población se distribuye normalmente $N(1'70, 0'08)$, ¿qué porcentaje deberá fabricar de cada talla?

20) En 200 muestras de un fármaco se han encontrado las siguientes muestras de arsénico (en mg por Kg.):

| Intervalo | Frecuencia |
|---|---|
| $[2'3, 2'35)$ | 4 |
| $[2'35, 2'4)$ | 16 |
| $[2'4, 2'45)$ | 34 |
| $[2'45, 2'5)$ | 45 |
| $[2'5, 2'55)$ | 49 |
| $[2'55, 2'6)$ | 38 |
| $[2'6, 2'65]$ | 14 |

Ajusta una distribución normal a esta muestra y compara los resultados dados con los teóricos.

21) Si las notas de matemáticas tienen una distribución normal $N(6'3, 2'4)$, halla las notas límites que clasifican a los alumnos en cuatro categorías con igual número de individuos.

22) Si la clase anterior se compone de treinta alumnos, ¿cuántos de ellos han aprobado?

23) El peso de la producción de los naranjos de un huerto se distribuye normalmente. El 25% tiene más de 51 Kg. de fruta, y el 60% más de 40. ¿Cuál es la media y la desviación típica de esta distribución?

24) Se realiza una encuesta de opinión sobre un tema determinado. Se envían mil cuestionarios por correo. Estos cuestionarios suelen ser contestados en el 32% de los casos. ¿Qué probabilidad hay de recoger, al menos, trescientos cuestionarios? ¿Y de recoger justamente la mitad?

25) Las notas de inglés han tenido una distribución normal $N(6'5, 1)$ y las de literatura $N(7'5, 2)$. Juan ha obtenido, respectivamente, un 8 y un 9 en cada asignatura. ¿En cuál ha destacada más sobre el resto de la clase? Si Pedro tiene un 7 en inglés, ¿cuál es la nota equivalente en literatura?

26) En una determinada población, una de cada 10.000 personas está esquizofrénica. Una población sólo puede ofrecer asistencia adecuada a 25 enfermos. Si la población es de 200.000 personas, ¿cuál es la probabilidad de que se hagan insuficientes las posibilidades de asistencia?

27) Lanzamos cien veces una moneda perfecta. ¿Cuál es la probabilidad de que obtengamos entre 45 y 55 caras?

28) Se ha realizado un estudio de los resultados obtenidos en matemáticas por los alumnos que han ido pasando un por un instituto de secundaria, obteniendo que la probabilidad de que un alumno apruebe es 0'8. Si el estudio se refiere a una población de 5.000 alumnos y llamamos *X* a la variable número de alumnos aprobados:

a) Determina la función de probabilidad.
b) Calcula su media y varianza.
c) Halla la probabilidad de que el número de alumnos aprobados esté entre 4.000 y 4.500.
d) Halla la probabilidad de que haya más de 4.000 alumnos aprobados.

29) El 11% de los billetes de lotería reciben algún tipo de premio, aunque sea el reintegro. Si en una familia juegan a 46 números, ¿cuál es la probabilidad de que resulten premiados, al menos, diez de ellos?

30) Las notas de un examen de estadística siguen una distribución normal. El 15'87% de los alumnos tiene una nota superior a 7, y el 15'87% inferior a 5.

a) ¿Cuál es la nota media del examen?
b) ¿Qué porcentaje de alumnos tiene una nota entre 6 y 7?

31) Cierta enfermedad tiene una probabilidad muy baja de ocurrir, $p = \dfrac{1}{100.000}$. Calcula la probabilidad de que en Málaga (540.000 habitantes) haya más de tres personas que la padezcan.

## EJERCICIOS DE RECOPILACIÓN

32) ¿Se puede decir que la siguiente tabla se ajusta a una distribución binomial?

| Nº mascotas familia | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Frecuencia | 28 | 44 | 24 | 4 |

Halla los datos teóricos y compáralos con los experimentales.

33) El número de calzado de un grupo de 74 alumnos de bachillerato es:

| $x_i$ | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $f_i$ | 1 | 2 | 13 | 15 | 8 | 10 | 6 | 8 | 5 | 2 | 1 | 3 |

Calcula sus parámetros.

34) La probabilidad de que un piloto de carreras sufra un reventón en un circuito es 0’04. Si en una carrera participan 200 conductores, ¿cuál es la probabilidad de que se produzcan entre 12 y 18 reventones?

35) La probabilidad de cara en una moneda trucada es 0’45. Se lanza la moneda 7 veces. Calcula la probabilidad de que:

a) Salgan exactamente tres caras.
b) Salgan al menos tres caras.
c) Salgan a lo sumo tres caras.

36) En un determinado juego se gana cuando al lanzar dos dados se obtiene una suma de 10 o más puntos. Un jugador tira en 12 ocasiones los dados. Halla las siguientes probabilidades:

a) Ganar justamente entres ocasiones.
b) Perder las doce veces que juega.

37) En un examen de matemáticas se ha evaluado de 0 a 20 puntos, obteniéndose el 67% de los alumnos puntuaciones iguales o inferiores a 12’2, y el 9% puntuaciones superiores a 16’7. La distribución de las puntuaciones es normal. Calcula la media y la desviación típica.

38) Determina los extremos de un intervalo simétrico respecto a la media en una distribución normal que contiene el 80% de las observaciones. Aplícalo para $\mu = 5$ y $\sigma = 2$.

39) Las notas de un examen hecho a una clase de 36 alumnos sigue una distribución $N(4'2, 1'3)$. Calcula:

a) Número de alumnos que han obtenido una nota de aprobado.
b) Número de alumnos que ha obtenido una nota entre 4 y 6.

40) Aplicando un test a un grupo de 300 personas, se ha obtenido una distribución normal de media 50 y desviación típica 5. Se pide:

a) El percentil 33.
b) Las puntuaciones que delimitan el 30% de la distribución.
c) El número de personas que obtiene en el test más de 56 puntos o menos de 47.

41) Una fábrica de relojes produce un modelo para el que los controles de calidad detectan la aparición de un defecto con probabilidad 0’1. pero que un reloj sea defectuoso es independiente de que los otros lo sean o no. En el curso de la fabricación retiramos 5 relojes al azar. Calcula las siguientes probabilidades:

a) Al menos uno de los relojes extraídos es defectuoso.

b) Exactamente dos relojes son defectuosos.

42) Las puntuaciones obtenidas en una asignatura siguen una distribución normal. Sabiendo que $p(X \leq 3) = 0'1587$ y $p(X > 9) = 0'0228$, determina la media y la desviación típica.

43) Una compañía de autobuses conoce que el retraso en la llegada sigue una ley normal de media 5 minutos y que el 68'26% de los autobuses llega con un retraso comprendido entre los 2 y los 8 minutos.

a) ¿Cuál es la desviación típica?
b) ¿Cuál es la probabilidad de que un autobús llegue puntual o antes de la hora?
c) ¿Y de que un autobús llegue con un retraso de más de 10 minutos?