# LECCIONES DE ESTADÍSTICA

Estos apuntes fueron realizados para impartir el curso de Métodos Estadísticos y numéricos en el I.E.S. A Xunqueira I de Pontevedra.

Es posible que tengan algún error de transcripción, por lo que no me responsabilizó de las consecuencias que dichos errores puedan inducir.

**J.M. Ramos**
**Pontevedra 2008**

# ÍNDICE

# TEMA I. ESTADÍSTICA DESCRIPTIVA

## Fenómenos determinísticos

Llamados también causales, son aquellos en los que se obtienen los mismos resultados, siempre que se realicen en las mismas condiciones. En ellos es posible predecir el resultado final conociendo el estado inicial y las condiciones de realización. Están sujetos a leyes naturales que pueden ser formuladas mediante ecuaciones matemáticas.

Se comprende fácilmente la imposibilidad de realizar un fenómeno en las mismas condiciones absolutas, pues la imperfección de la mente y de los sentidos del hombre hace que no podamos representarnos todas las causas que intervienen en un determinado fenómeno. Estas causas desconocidas las sustituimos por lo que se llama azar, pero en el caso de los fenómenos determinísticos el azar juega un papel tan ínfimo que suele despreciarse su efecto.

En muchos fenómenos (físicos especialmente) la presencia del azar es mínima, sin embargo en otros fenómenos, como los sociales, lo imprevisible es de tal magnitud que hace que no podamos predecir el estado final.

Obviamente los fenómenos determinísticos son objeto de estudio por parte de ciencias tales como La Física, la Química...etc.

Un ejemplo de fenómeno determinístico es la caida libre y en vacío de un objeto desde una altura h. Se sabe a priori que la velocidad final con la que va a llegar al suelo (estado final) viene determinada por la ley $v = \sqrt{2gh}$ . Otro sería que si tenemos un gas comprimido en un recipiente de volumen V y sometido a una presión P, su temperatura viene dada por la fórmula : $T = \frac{P.V}{nR}$, donde n es el número de moles de gas y R es una constante química.

Observemos que en ambos casos podemos saber con seguridad el estado final siempre y cuando realicemos el experimento con un estado inicial conocido y dándose la particularidad de que si lo volvemos a realizar tantas veces como deseemos, los resultados finales van a ser idénticos.

Todo ello es lo que caracteriza a los fenómenos determinísticos o causales.

**Fenómenos aleatorios**

A diferencia de los anteriores, son aquellos en los que es imposible predecir el resultado final, aun repitiéndolo en las mismas condiciones y además, donde una pequeña variación en las condiciones iniciales produce una gran variación en los resultados finales. (El aleteo de una mariposa en china puede provocar un huracán en Panamá). Esta característica podría hacernos pensar en la imposibilidad de un estudio formal y util de este tipo de fenómenos. La presunta esterilidad del estudio de los fenómenos aleatorios queda refutada de inmediato debido a que todos presentan una importantísima propiedad empírica denominada regularidad estadística que analizaremos con un ejemplo.

Obviamente son fenómenos aleatorios, el lanzar un dado o una moneda; extraer una carta de una baraja o una bola de una urna, rellenar un boleto de la lotería primitiva...etcétera. Como se podrá intuir hay toda una infinidad de fenómenos aleatorios.

Estos fenómenos son objeto de estudio de la Estadística.

**Estadística. Concepto:**

Como ya se concluyó en el apartado anterior, la Estadística es la ciencia que estudia los fenómenos aleatorios.

Históricamente, parece ser que los datos más antiguos que se poseen acerca del uso de las técnicas estadísticas, se remontan a los censos chinos ordenados por el emperador Tao, 2200 años a.C.

También tuvieron importancia los censos romanos hacia el año 555 a.C.

Es en el año 1660 cuando se publica la obra "*Aritmética política o descripción de las cosas notables del Estado*" de Hoernán Conring. Es donde esta ciencia comienza a denominarse Estadística.

Más adelante citar los notables trabajos del alemán Mendel (1822-1844), abad del monasterio de Brünn que fue quién puso los cimientos de la actual Genética con sus estudios sobre la herencia.

Citar a Pascal y a otros colegas franceses, que motivados por problemas surgidos en los juegos de azar, comienzan a estudiar en serio esta Ciencia.

En el siglo XX, la Estadística adquiere un impulso renovador con los estudios de los ingleses Pearson, Galton y Weldon, así como del ruso Kolmogorov.

Hoy en día la Estadística forma parte de nuestra realidad más cotidiana, pues basta abrir un periódico o escuchar un noticiario para darnos cuenta de la cantidad de datos y conceptos estadísticos que se manejan.

**Estadística descriptiva**

Es aquella rama de la estadística donde las conclusiones que se obtienen de las experiencias o datos en estudio no rebasan los límites de los mismos. Tiene como propósito su representación mediante tablas, gráficos y reducciones de datos. Puede también comprender el análisis de los mismos, siempre que sus conclusiones no trasciendan más alla de dichos datos.

**Conceptos básicos de Estadística descriptiva.**

***Población o Universo Estadístico***.- Conjunto formado por todos los elementos que posean una serie de caracteres previamente estipulados. Cada uno de los elementos de la población se denomina individuo y éstos pueden ser simples (hombres, piezas) o colectivos (familias)

Toda vez que el número de los elementos de una población objeto de estudio estadístico, en la mayoría de los casos es muy numeroso, resulta caro y engorroso el estudio de los mismos, por lo que se recurre a un subconjunto representativo de la población.

***Muestra***.- Para evitar un estudio, a veces imposible, de una población debido a su gran número de individuos, se suele tomar un subconjunto representativo llamado muestra, bajo criterios previamente estudiados, de tal modo que el estudio en dicho conjunto nos permita inducir resultados en toda la población con un grado de certeza ciertamente grande. De la buena elección de la muestra, dependerá la bondad de los datos extraídos para la población.

***Tamaño de la población*****:** Obviamente el tamaño es el número de individuos que componen la población o muestra. Lo representaremos por **N**

***Variable estadística*****.-** Denominada también *carácter***,** es una característica o propiedad común a todos los individuos de una población objeto de estudio estadístico. Por ejemplo en una población de personas, el peso podría ser una variable estadística, el color de ojos...etc.

A todos los posibles resultados de una variable estadística se le denominan *modalidades* o valores de la variable (datos)

Normalmente representaremos por **X** una v.e y por $\mathbf{x_i}$ sus modalidades.

Un ejemplo de variable estadística podría ser el sexo, cuyas modalidades serían varon, hembra. Se trataría de una v.e. con dos modalidades; por el contrario si consideramos la v.e. "estatura", las modalidades pueden llegar a ser infinitas, dentro de un intervalo.

Estos dos últimos ejemplos me van a permitir clasificar las variables estadísticas según el siguiente esquema:

$$Variable\ estadística \begin{cases} Cuantitativas \begin{cases} Discretas \\ Continuas \end{cases} \\ \quad Cualitativas \end{cases}$$

Vamos a analizar esta clasificación:

a) ***Variables estadísticas cuantitativas***: Son aquellas cuyas modalidades vienen representadas por un valor numérico; es decir son de algún modo medibles. Por ejemplo, la estatura, el peso, en valor de un dado...etc.

a.1) Son ***discretas*** cuando la cantidad de modalidades en un conjunto finito o infinito numerable (los valores de un dado, el número de bolas blancas que se extraen de una urna con devolución hasta que aparezca una negra)

a.2) Son ***continuas*** cuando la cantidad de modalidades es infinito y toma cualquier valor en un intervalo real. (el peso de una persona)

**b)** ***Variables estadísticas cualitativas***: Sus modalidades son atributos o cualidades no medibles y por tanto carecen de representatividad numérica (color de ojos, cara o cruz...etc)

***Parámetro*.-** Es toda función definida sobre los valores numéricos de una población (Media aritmética de las altura de todos los alumnos de Bachillerato de Galicia)

***Estadístico*.-** Es toda función definida sobre los valores numéricos de una muestra (Media aritmética de las altura de los alumnos de una muestra de 10 alumnos por centro de bachillerato de Galicia)

Sea una población o muestra, de tamaño N, sobre la que vamos a estudiar una variable estadística discreta con p modalidades: $x_1, x_2, x_3 \ldots x_p$

***Frecuencia absoluta de la modalidad $x_i$*.-** Es el número entero de veces que dicha modalidad aparece en la población o muestra. Lo representaremos por $\mathbf{n_i}$.

Son de destacar las siguientes propiedades, que se deducen inmediatamente de la definición:

**a)** $0 \leq n_i \leq N \quad \forall i = 1 \ldots p$

**b)** $\sum_{i=1}^{p} n_i = N$

***Frecuencia relativa de la modalidad $x_i$*.-** Es la proporción con la que aparece la modalidad $x_i$ en la población o muestra. Se obtiene dividiendo la frecuencia absoluta entre N.

En muchas ocasiones esta proporción viene expresada en tanto por ciento. Basta multiplicar la frecuencia relativa por 100.

La frecuencia relativa la representaremos por $\mathbf{f_i}$.

En consecuencia: $f_i = \dfrac{n_i}{N}$

Son propiedades triviales de la frecuencia relativa, que se deducen de las indicadas para la frecuencia absoluta, las siguientes:

**a)** $0 \leq f_i \leq 1 \quad \forall i = 1 \ldots p$

**b)** $\sum_{i=1}^{p} f_i = 1$

***Frecuencias acumuladas*.-** Ordenados en sentido creciente o decreciente los valores o modalidades de una v.e., definiremos la frecuencia acumulada (absoluta o relativa) como la suma de frecuencias hasta un valor determinado de la variable. Cuando la modalidad es la última de la ordenación creciente, la frecuencia acumulada será igual a N si se trata de frecuencias absolutas, y 1 si lo que se acumulan son las frecuencias relativas.

La frecuencia acumulada absoluta para la modalidad $x_i$, se representará por $\mathbf{N_i}$, mientras que si se trata de la relativa, la representaremos por $\mathbf{F_i}$

***Tablas de recogida de datos.-*** Toda la información recogida anteriormente se dispone en una tabla de la siguiente manera:

| **Datos** | **Fr. Absoluta** | **Fr. relativa** | **Fr.Abs.Acu.** | **Fr.Rel.Acu.** | **%** |
|---|---|---|---|---|---|
| $x_1$ | $n_1$ | $f_1$ | $F_1=n_1$ | $F_1 = f_1$ | $100.f_1$ |
| $x_2$ | $n_2$ | $f_2$ | $F_2 = n_1+n_2$ | $F_2 = f_1+f_2$ | $100.f_2$ |
| ... | ... | ... | ... | ... | ... |
| $x_p$ | $n_p$ | $f_p$ | $F_p= N$ | $F_p = 1$ | $100.f_p$ |
| TOTALES | N | 1 | | | 100 |

Es de resaltar que estas definiciones anteriores están establecidas cuando el conjunto de datos en discreto.

En caso de que la variable sea cuantitativa continua, el número de modalidades es infinito y no toman valores aislados, por lo que no podemos representarlas mediante los $x_i$. ¿Cómo se procede en este caso?

***Intervalos de clase. Marcas de clase.-***

Cuando la variable estadística es cuantitativa continua o discreta pero con gran cantidad de modalidades, es necesario dividir el recorrido de la variable en intervalos, a ser posible de igual tamaño, denominados intervalos de clase. El número de intervalos influirá en la precisión de los estadísticos que se vayan a estudiar. Obviamente a mayor número de intervalos, mayor precisión.

Utilizaremos la siguiente notación. El intervalo de clase lo denotaremos por $(L_{i-1} - L_i)$, siendo $L_{i-1}$ y $L_i$ los límites inferior y superior, respectivamente, del intervalo.

La diferencia entre ambos límites nos determinarán el tamaño o longitud del intervalo, denominado amplitud y que representaremos por $c_i$. Así pues, $c_i=L_i-L_{i-1}$.

Como vamos a seguir teniendo necesidad de utilizar valores numéricos para obtener los distintos estadísticos de la muestra, consideraremos como valores representativos de los intervalos de clase, sus valores centrales, que denominaremos **marcas de clase**. Las marcas de clase son los equivalentes a los $x_i$ en el caso discreto. Su valor viene dado por: $x_i = \frac{L_{i-1} + L_i}{2}$.

Las frecuencias absolutas y relativas se referirán a los intervalos de clase, de tal modo que diremos que el intervalo $(L_{i-1} - L_i)$ tiene frecuencia absoluta $n_i$ cuando el número de individuos, cuya modalidad caiga dentro de dicho intervalo, sea precisamente $n_i$.

Análogamente haremos lo mismo para la frecuencia relativa.

Es importante que cada dato esté en un solo intervalo, por lo que los extremos han de ser indicados abiertos o cerrados, según el caso. En ocasiones para evitar esto, los valores extremos de los intervalos suelen ser tomados con una cifra decimal más que la que poseen los datos.

Este procedimiento de clasificar los datos en intervalos va a producir una inevitable pérdida de información, puesto que no se considerarán los resultados exactamente, sino por aproximación: no se dirá que un elemento tiene un carácter cuya medida es $x_i$, sino que dicho valor se encuentra en el intervalo $(L_{i-1} - L_i)$.

Así pues, lo que interesa es elegir una amplitud (a ser posible constante) de los intervalos lo suficientemente pequeña para que la pérdida de información sea la menor posible y, al mismo tiempo, lo suficientemente grande para que el agrupamiento presente una distribución de no demasiados intervalos, pues de lo contrario dicho agrupamiento perdería su finalidad, es decir, la comodidad del tratamiento.

**Tablas de recogida de datos**

| Intervalos | Marcas | Fr. Absoluta | Fr. relativa | Fr.Abs.Acu. | Fr.Rel.Acu. | % |
|---|---|---|---|---|---|---|
| $L_0$-$L_1$ | $x_1$ | $n_1$ | $f_1$ | $F_1=n_1$ | $F_1 = f_1$ | $100.f_1$ |
| $L_1$-$L_2$ | $x_2$ | $n_2$ | $f_2$ | $F_2 = n_1+n_2$ | $F_2 = f_1+f_2$ | $100.f_2$ |
| … | ... | ... | ... | ... | ... | ... |
| $L_{p-1}$-$L_p$ | $x_p$ | $n_p$ | $f_p$ | $F_p= N$ | $F_p = 1$ | $100.f_p$ |
| TOTALES | | N | 1 | | | 100 |

**Representaciones gráficas**

Partiendo de la máxima "vale más una imagen que mil palabras", el objeto de las representaciones gráficas es precisamente hacer valer esta frase hecha, de modo que el impacto visual de la representación responda a la realidad y, por consiguiente, el método seguido deberá basarse en principios geométricos ortodoxos.

Veamos los casos de representación para el caso de variables discretas no agrupadas en intervalos.

a) **Diagrama de barras**: Se elabora señalando en las abscisas de un sistema de ejes de coordenadas los valores de la variable construyendo sobre ellos unas columnas de altura igual a la frecuencia de cada uno de los valores, medida en el sentido del eje de ordenadas.

También puede realizarse un diagrama de barras para frecuencias acumuladas.

b) **Diagrama de sectores**: Consiste en un círculo con sectores de área proporcional a las frecuencias de cada uno de los valores. El ángulo correspondiente al sector de la modalidad $x_i$, viene dado por $360.f_i$.

También puede representarse en un semicírculo, por lo que los ángulos vendrían dados por $180.f_i$

c) **Pictograma**: No es más que un diagrama de barras, pero en vez de simples columnas, se ilustra con figuras alusivas a los datos estudiados (animales, personas, máquinas...etc)

Cuando los datos vienen agrupados por intervalos, disponemos de otro tipo de representación gráfica: **Los histogramas.**

Los histogramas son rectángulos de base igual a los intervalos de clase y altura proporcional a la frecuencia establecida para dichos intervalos. (también se pueden hacer para frecuencias acumuladas)

**Polígono de frecuencias:**

Es el polígono limitado en un histograma por el eje de abscisas y la línea quebrada resultante de unir el punto medio del lado superior de cada rectángulo (con frecuencias absolutas) o la línea quebrada resultante de unir en cada rectángulo el vértice del polígono anterior con el suyo.

Se realiza tanto para histogramas de frecuencias simples como acumuladas.

### *REDUCCIÓN DE DATOS. MEDIDAS CARACTERÍSTICAS*

Esta denominación de reducción de datos se debe a Fisher, y consiste en sustituir la tabla estadística, la cual nos da una idea de difícil comparación con otras tablas, por unos números (estadísticos o medidas características) que midan las características más importantes de la distribución de los datos.

Estos valores o medidas características pueden ser de dos tipos: de centralización y de dispersión.

#### *Medidas características de centralización:*

Son valores de tendencia central en torno a los cuales se encuentran los valores de la variable estadística, con arreglo a un cierto critero de equiiibrio para las frecuencias. Es la característica más importante de la distribución y se mide a traves de los promedios, entre los cuales están los siguientes:

**a) Media aritmética** (Cuando se habla de media a secas, nos referiremos a la media aritmética): Se obtiene multiplicando cada valor de la v.e. por su frecuencia relativa y sumando todos los productos obtenidos. Se representa por $\bar{x}$ .

$$\bar{x} = \sum_{i=1}^{p} x_i . f_i \quad = \quad \sum_{i=1}^{p} x_i . \frac{n_i}{N} \quad = \quad \frac{1}{N} \sum_{i=1}^{p} n_i$$

Se llama también *centro de gravedad* de la distribución por admitir la siguiente interpretación mecánica: Consideremos que sobre cada punto $x_i$ actúa una fuerza de valor $n_i$. En este caso $\bar{x}$ corresponde al punto del eje en el cual, aplicando la fuerza de valor N, produce el equilibrio del sistema.

Para evitar largos y tediosos cálculos de aritmética elemental, la media aritmética la obtendremos mediante el uso de calculadora científica.

Si consideramos los valores de desviación, $\bar{x}$ - $x_i$ , se tiene como propiedad interesante que

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

c) **Media aritmética ponderada.-**

En determinadas ocasiones queremos que algún dato tenga más valor, o “pese” más a la hora de ser considerado dentro de la distribución. Asi pues, en este caso, asociamos a los datos $x_1$, $x_2$, ..., $x_p$, ciertos factores peso (o pesos) $w_1$, $w_2$, ..., $w_p$, dependientes de la relevancia asignada a cada número. En tal caso, la expresión:

$\bar{x} = \sum_{i=1}^{p} \frac{x_i . w_i}{w_i}$ se llama media aritmética ponderada de pesos $w_1, w_2, ..., w_p$

**d) Media geométrica**

$$G = \sqrt[N]{x_1^{n_1} . x_2^{n_2} ... x_p^{n_p}} \quad = \quad (\prod_{i=1}^{p} x_i^{n_i})^{\frac{1}{N}}$$

No tiene unas propiedades tan sencillas y claras como la media aritmética; no obstante se suele usar cuando la variable sigue progresión geométrica. También en la elaboración de los números índices (que se verán más adelante), muestra una propiedad interesante.

**e) Media armónica**

$$H = \frac{N}{\sum_{i=1}^{p} \frac{1}{x_i} . n_i}$$

**f) Media cuadrática**

$$\mathbf{C} = \frac{\sum_{i=1}^{p} x_i^2 . n_i}{N}$$

Esto puede generalizarse definiendo la **media general de orden m**

$$M(m) = \sqrt[m]{\frac{1}{N} \sum_{i=1}^{p} x_i^m . n_i}$$

resultando la media armónica para m=-1; la media geométrica para m=0, la media aritmética para m=1 y la media cuadrática para m=2

Se verifica: $M(-1) \leq M(0) \leq M(1) \leq M(2)$, es decir que:

$$H \leq G \leq \bar{x} \leq C$$

**Mediana**

Ordenados los datos de menor a mayor o viceversa, la mediana es el valor de la v.e, que deja el 50% de los datos a un lado y el 50% restante al otro. En otras palabras, la mediana divide por la mitad a los datos.

El cálculo de la mediana difiere en función del tipo de distribución, es decir si se trata de valores aislados o distribuidos por intervalos de clase.

En el primer caso, la mediana se obtiene del siguiente modo:

a) Si el número de datos es par, la mediana es la media aritmética de los dos valores centrales que se obtienen al ordenar dichos datos, es decir que si los N datos ordenados son $a_1$, $a_2$, ... , $a_N$, la mediana es:

$$\textbf{Me} = \frac{a_{\frac{N}{2}} + a_{\frac{N}{2}+1}}{2}$$

b) Si el número de datos es impar, solamente habrá un valor central. Este será precisamente la mediana, y dicho valor, siguiendo la notación anterior es.

$$a_{Ent(\frac{N}{2})+1}$$

En el caso de distribución por intervalos el valor de la mediana es:

$Me = L_{i-1} + \dfrac{(N/2) - N_{i-1}}{n_i}.c_i$. , siendo $L_{i-1}$ el límite inferior del intervalo donde se encuentra la mediana (aquel donde se alcance el 50% del total de la población N); $N_{i-1}$ es la frecuencia acumulada absoluta hasta dicho intervalo; $n_i$ es la frecuencia absoluta en el intervalo y $c_i$ es la amplitud del intervalo.

Esta fórmula se demostrará en clase.

**Ejemplo:** Sea la distribución siguiente:

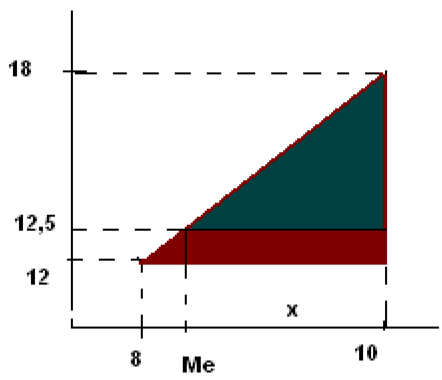
| Intervalos | Marcas | Frecu. Absoluta | Frec. Abs. acum.. |
|---|---|---|---|
| 2-4 | 3 | 5 | 5 |
| 4-6 | 5 | 4 | 9 |
| 6-8 | 7 | 3 | 12 |
| 8-10 | 9 | 6 | 18 |
| 10-12 | 11 | 7 | 25 |

Primero detectamos el intervalo donde se encuentra la mediana. Como N=25, calculamos N/2 = 12,5. Esta frecuencia, viendo la tabla de frecuencias acumuladas, se encuentra en el intervalo 8-10, por tanto $L_{i-1}=8$, $N_{i-1}=12$, $n_i=6$.

De ahí se obtiene que $Me = 8 + \dfrac{12{,}5 - 12}{6}.2 = 8{,}1\overline{6}$ .

**Método gráfico del cálculo de la mediana**

En el ejemplo anterior sería:




Como ambos triángulos son semejantes

6/2 = 5,5/x ; de donde x = 1,83

Me = 10 – 1,83 = 8,17

**Moda**

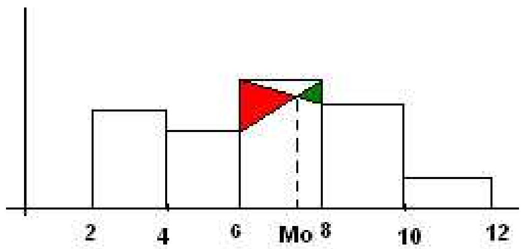
Es el valor de la v.e. que más se repite. Según esta definición puede haber más de una moda, llamándose la distribución unimodal, bimodal...etc, según tenga una, dos ...etc modas.

El cálculo para distribuciones con valores aislados es trivial; sin embargo para el caso de distribuciones agrupadas por intervalos no es tan evidente, y la fórmula que determina la moda es:

$Mo = L_{i-1} + \dfrac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}.c_i$, siendo $L_{i-1}$ – $L_i$ el intervalo donde se encuentra la moda.

**Método gráfico para el cálculo de la moda** (en el caso de distrib. agrupadas)
Ejemplo**:** Sea la distribución siguiente:

| Intervalos | Marcas | Frecu. Absoluta | Frec. Abs. acum.. |
|---|---|---|---|
| 2-4 | 3 | 5 | 5 |
| 4-6 | 5 | 4 | 9 |
| 6-8 | 7 | 7 | 16 |
| 8-10 | 9 | 6 | 22 |
| 10-12 | 11 | 2 | 24 |




Estudiamos primero cual es el intervalo modal (de mayor frecuencia) y trazamos los segmentos que unen los vértices superiores del rectángulo del histograma correspondiente en dicho intervalo con los extremos de los rectángulos contiguos, obteniéndose dos triángulos (el verde y el rojo) semejantes pues tienen ángulos iguales. Si llamamos x a la altura (respecto del lado vertical) del triángulo rojo, resulta que 2-x es la altura del verde. Por otra parte la base (lado vertical) del rojo vale 7-4=3, y la del verde vale 7-6=1. Estableciendo la razón de semejanza, se obtiene que x/3 = (2-x)/1; de donde x= 1,5; de donde la Moda será 6+1,5 = 7,5.

**Relación empírica entre media, mediana y moda.**

Para distribuciones de frecuencia unimodal que sean poco asimétricas se tiene que:
Media – Moda = 3(media – mediana)

***MEDIDAS POSICIONALES***

**Los cuantiles**

Se llama cuantil de orden m a un valor $x_m$ que deja por debajo de él al m por 100 de los elementos de la población.

**Cuartiles:**

Son los valores que dividen a la población en 4 partes iguales. Existen por tanto tres cuartiles: $Q_1$ (primer cuartil), $Q_2$ (segundo cuartil), $Q_3$ (tercer cuartil). Resulta obvio que $Q_2$ = Me.

**Cálculo de los cuartiles.**

a) Caso discreto o valores aislados: Ordenados de menor a mayor los datos y para i=1,2,3. Si N.i/4 no corresponde a ningún valor de la frecuencia acumulada (está entre $N_k$ y $N_{k+1}$) se le da a $Q_i$ el valor de la variable que corresponde a la frecuencia acumulada $N_{k+1}$.

En caso de que N.i/4 corresponde a un valor $N_k$ de la frecuencia acumulada, el cuartil es la media aritmética entre $x_k$ y $x_{k+1}$

b) Caso continuo o por intervalos:

Se utiliza la fórmula:

$$Q_r = L_{i-1} + \frac{r.(N/4) - N_{i-1}}{n_i}.c_i \text{, para r=1,2,3}$$

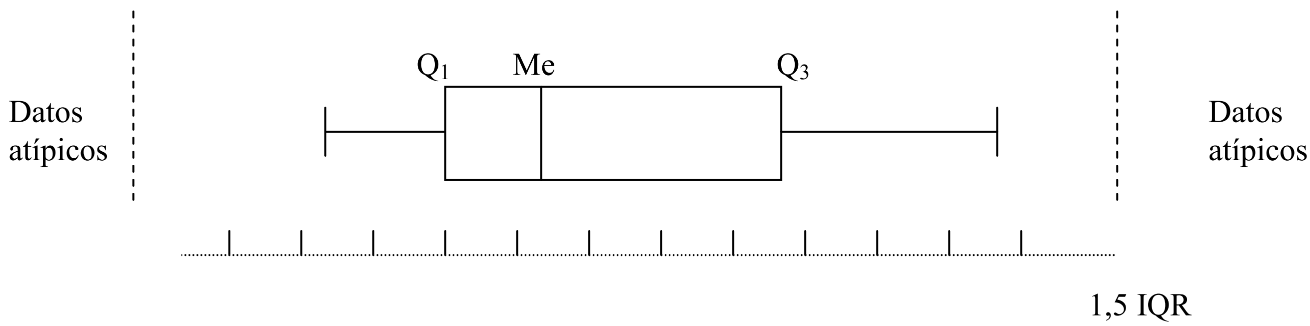
**Recorrido intercuartílico o I.Q.R.**: Es la diferencia entre $Q_3$ y $Q_1$. La idea es dividir los datos en cuatro grupos iguales y ver lo distantes que son los extremos de esos grupos

**Box plots o diagramas de Tukey de caja y bigotes**

Para realzar el IQR, John Tuker inventó un tipo de representación llamado Diagrama de caja y bigotes. Los extremos de la cja son los cuartiles $Q_1$ y $Q_3$. La Mediana se dibuja dentro de la caja. Si un punto está a más de 1,5 veces el IQR alejado de un extremo de la caja, se denomina atípico y se dibuja de forma aislada. Finalmente se extienden los "bigotes" hasta los puntos más lejanos que no sean atípicos (es decir dentro de 1,5 veces el IQR de los cuartiles).

Estos diagramas son especialmente buenos para realzar las diferencias entre grupos.

Veamos como son:



**Quintiles:**

Son los valores que dividen a la población en 5 partes iguales. Existen 4 quintiles: $K_1$, $K_2$, $K_3$, $K_4$.
Su cálculo es exactamente igual que en el caso de los cuartiles y su fórmula es:

$$K_r = L_{i-1} + \frac{r.(N/5) - N_{i-1}}{n_i}.c_i, \quad \text{para } r=1,2,3,4$$

**Deciles:**

Son los valores que dividen a la población en 10 partes iguales. Existen 9 deciles, siendo estos: $D_r$, con r =1...9

Su cálculo es exactamente igual que en el caso de los cuartiles y además se verifica:
$D_2=K_1$ ; $D_4=K_2$ ; $D_5$ = Me ; $D_6=K_3$ ; $D_8=K_4$
La formula para los deciles es:

$$D_r = L_{i-1} + \frac{r.(N/10) - N_{i-1}}{n_i}.c_i \quad \text{para } r=1, 2, 3... 9$$

**Centiles o percentiles:**

Son los que dividen a la población en 100 partes iguales. Hay 99 centiles que se representan por $C_r$ o $P_r$ indistintamente, donde r= 1...99.

Para su cálculo nos remitimos al cálculo de los cuartiles.
Se producen ciertas identidades tales como: $C_{25}=Q_1$; $C_{50}$ = Me...etc.

La fórmula para centiles es:

$$C_r = L_{i-1} + \frac{r.(N/100) - N_{i-1}}{n_i}.c_i \quad \text{para } r=1,2,3...99$$

## ANEXO

### *RECOGIDA DE DATOS CON LA CALCULADORA*

Dependiendo de las marcas comerciales de las calculadores científicas, el modo de trabajar con datos estadísticos normalmente varía de una a otra.

*INTRODUCCIÓN DE DATOS CON LA CALCULADORA CASIO* **fx-100D**

Paso 1) Activar el modo de trabajo en Estadística:

MODE + 3 (En la cabecera de la pantalla tiene que aparecer la leyenda “SD”)

Paso 2) Borrado de posibles datos en memoria de trabajos anteriores:

KAC = SHIFT + AC

Paso 3) Comprobar que, en efecto, no tenemos datos

n = SHIFT + 3 (Este n es el equivalente a nuestro N). En la pantalla tiene que aparecer 0.

Paso 4) Introducir los datos:
Para introducir, por ejemplo, el dato 8 con frecuencia 4:

8 X 4 M+, o bien 8 y se pulsa la tecla M+ cuatro veces.

Paso 5) Comprobar que el número de datos introducidos coincide con N

n = SHIFT + 3. En la pantalla tiene que aparecer N.

Paso 6) Calcular la media aritmética, desviaciones típicas...etc.

SHIFT + 1

*INTRODUCCIÓN DE DATOS CON LA CALCULADORA* **CASIO fx-570 ES**

1) SHIFT Mode ▼ 4(STAT) 1(FRECUENCIA ON)
2) SHIFT 1(STAT) 2 (DATA)
3) Introducir x1 = x2 = x3 = ... con el cursor ► ▲ pasar a la columna frec e introducir los valores de las frecuencias absolutas correspondientes (n1 = n2 =...)
4) AC
5) SHIFT 1(STAT) 5(VAR):

1 (n) = tamaño de la población
2 (x) = media aritmética
3 (xón) = desviación típica.

6) Borrar datos: SHIFT 1(STAT) 3(Edit) 2: (Del-A)

*INTRODUCCIÓN DE DATOS CON LA CALCULADORA* **CASIO fx-350 ES**

1) SHIFT Mode ▼ 3 (STAT) 1(FRECUENCIA ON)

2) SHIFT 1(STAT) 2 (DATA)
3) Introducir x1 = x2 = x3 = ... con el cursor ► ▲ pasar a la columna frec e introducir los valores de las frecuencias absolutas correspondientes (n1 = n2 =...)
4) AC
5) SHIFT 1(STAT) 5(VAR):

1 (n) = tamaño de la población
2 (x) = media aritmética
3 (xón) = desviación típica.
6) Borrar datos: SHIFT 1(STAT) 3(Edit) 2: (Del-A)

INTRODUCCIÓN DE DATOS CON LA CALCULADORA **casio fx-570 MS**

1) SHIFT MODE 1 (SD)
2) Comprobación de datos SHIFT 1 3(n)
3) Introducción de datos: xi **;** ni M+
4) Cálculo de parámetros SHIFT 2 1(x media) 2 (xon) desv. tipica.
5) Borrado de datos: SHIFT 1(Clr)

***CREAR TABLAS DE FRECUENCIAS EN EXCEL***

**Caso discreto:**

En un rango se escriben todos los datos. A continuación en una matriz columna se escriben las modalidades. Seleccionamos la columna donde queremos que aparezcan las frecuencias absolutas Se pulsa $f_x$ y vamos a las funciones estadísticas, dentro de las que escogeremos FRECUENCIA. Hay dos parámetros que hay que introducir: Datos y grupos. En datos seleccionamos la matriz de datos y en grupos la matriz de modalidades. y damos salida a los resultados, al tratarse de una matriz, con CTRL.+Mayúsculas.

**Caso continuo:**

Al igual que en el caso anterior pero en el parámetro grupos hemos de escribir los limites superiores de los intervalos de acogida de datos.

**EJERCICIOS PROPUESTOS**

1. Las edades de seis dependientes de un comercio son: 18,19,25,29,34,35 años. Calcular la media de dichas edades

2. La medida de la longitud de 50 varillas ha dado los siguientes resultados: de 5 cms, 8 varillas; de 7 cm., 6 varillas; de 8 cm., 6 varillas; de 9 cm., 9 varillas; de 10 cm., 11 varillas; de 12 cm., 7 varillas y de 13 cm., 3 varillas. Calcular la media de estas longitudes.

3. Calcular la media de la distribución correspondiente a la estatura de 40 chicos de primero de Bachillerato, siendo esta:

| Intervalos | 148,5-153,5 | 153,5-158,5 | 158,5-163,5 | 163,5-168,5 | 168,5-173,5 | 173,5-178,5 |
|---|---|---|---|---|---|---|
| Frec.abs. | 2 | 4 | 11 | 14 | 5 | 4 |

4. Calcular la media de los siguientes valores agrupándolos primero por intervalos de amplitud igual a 5 y después por intervalos de amplitud igual a 10. Estos valores son:

49,48,43,42,49,41,42,43,43,44,44,51,53,54,51,59,58,57,56,54,51,54,53,64,62,64, 63,62,61,62,68,68,67,66,69

5. Dada la distribución de la tabla, calcular la media aritmética, la media geométrica, la media armónica, la media cuadrática. Comprobar que relación existe entre ellas:

| $x_i$ | 2 | 3 | 8 | 12 | 17 |
|---|---|---|---|---|---|
| $n_i$ | 2 | 2 | 3 | 3 | 1 |

6. Calcular la frecuencia correspondiente al tercer intervalo de la siguiente distribución, sabiendo que la media aritmética es igual a 11,5

| Int. | 4-6 | 6-10 | 10-16 | 16-20 | 20-30 |
|---|---|---|---|---|---|
| $n_i$ | 4 | 5 | x | 3 | 1 |

7. Calcular la mediana de las siguientes distribuciones de frecuencias:

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $n_i$ | 10 | 12 | 7 | 7 | 3 |

| $x_i$ | 2 | 4 | 6 | 8 | 10 | 12 | 14 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 12 | 10 | 8 | 7 | 5 | 8 | 10 |

8. Dada la siguiente serie estadística de la distribución de salarios a los obreros de una empresa, calcular la mediana:

| SALARIO | NUM. OBREROS | NUM. OBREROS. ACUM |
|---|---|---|
| 20000-25000 | 100 | 100 |
| 25000-30000 | 150 | 250 |
| 30000-35000 | 200 | 450 |
| 35000-40000 | 180 | 630 |
| 40000-50000 | 41 | 671 |

9. Las calificaciones de la asignatura de Matemáticas de los 40 alumnos de una clase vienen expresadas por la siguiente tabla:

| Nota | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Alum. | 2 | 2 | 4 | 5 | 8 | 9 | 3 | 4 | 3 |

Calcular los cuartiles 1 y 3 así como los percentiles de orden 30 y 70

10. Se tiene la siguiente distribución continua, expresada por la tabla siguiente:

| Interv. | 38-44 | 44-50 | 50-56 | 56-62 | 62-68 | 68-74 | 74-80 |
|---|---|---|---|---|---|---|---|
| $N_i$ | 7 | 8 | 15 | 25 | 18 | 9 | 6 |

Calcular los cuartiles 1 y 3, así como los percentiles 40 y 90

11. Una zapatería de caballeros vende en un día 45 pares de zapatos de las siguientes tallas.

| Talla | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 |
|---|---|---|---|---|---|---|---|---|
| Zapatos | 1 | 3 | 5 | 8 | 12 | 9 | 5 | 2 |

Calcular a) la mediana, b) Cuartiles, c) ¿Qué percentiles corresponden a la talla 39?

12. El número de hijos de 20 familias seleccionadas al azar, es el siguiente: 3, 1, 2, 2, 1, 5, 2, 2, 0, 6, 3, 2, 4, 3, 4, 2, 3, 1, 7, 6

a) formar la tabla de frecuencias
b) Construir el correspondiente diagrama de barras
c) Construir el polígono de frecuencias.
d) Construir un diagrama de sectores o en su defecto indicar el ángulo que correspondería a cada modalidad en dicho diagrama.

13. Los valores el ph sanguíneo en 80 individuos son los siguientes:

7,33 7,32 7,34 7,40 7,28 7,29 7,35 7,33 7,34 7,28 7,31 7,35 7,32 7,33 7,33 7,36 7,32 7,31 7,35 7,36 7,26 7,39 7,29 7,32 7,34 7,30 7,34 7,32 7,39 7,30 7,33 7,33 7,35 7,34 7,33 7,36 7,33 7,35 7,31 7,33 7,37 7,38 7,38 7,33 7,35 7,30 7,31 7,33 7,35 7,33 7,27 7,33 7,32 7,31 7,34 7,32 7,32 7,32 7,31 7,36 7,30 7,37 7,33 7, 32 7,31 7,33 7,32 7,30 7,29 7,38 7,33 7,35 7,32 7,33 7,32 7,34 7,32 7,34 7,32 7,33

a) Formar la tabla de frecuencias utilizando 15 intervalos de clase.
b) Construir el histograma de frecuencias
c) Polígono de frecuencias
d) Construir el histograma de frecuencias acumuladas
e) Construir el polígono de frecuencias acumuladas

14. Se ha medido la longitud de 50 individuos adultos de una determinada especie de rana, obteniéndose los siguientes resultados:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 32,1 | 31,0 | 32,6 | 30,0 | 32,8 | 31,4 | 32,0 | 30,0 | 30,1 | 31,8 |
| 34,0 | 31,7 | 33,0 | 31,0 | 32,3 | 32,6 | 32,0 | 31,4 | 30,2 | 32,0 |
| 33,0 | 31,4 | 32,4 | 31,6 | 32,7 | 34,0 | 33,2 | 33,1 | 33,7 | 31,0 |
| 31,8 | 33,0 | 32,3 | 31,4 | 32,4 | 31,4 | 34,0 | 33,4 | 32,7 | 32,3 |
| 32,2 | 33,1 | 34,2 | 31,3 | 29,6 | 32,7 | 33,0 | 31,4 | 32,6 | 33,0 |

a) Formar la tabla de frecuencias en 5 intervalos de clase
b) Construir el histograma de frecuencias relativas
c) Polígono de frecuencias relativas
d) Construir el histograma de frecuencias relativas acumuladas
e) Construir el polígono de frecuencias relativas acumuladas

15. El número de accidentes mortales diarios en una gran ciudad en nueve días han sido: 6, 4, 8, 1, 5, 3, 3, 7, 2.

a) Hallar la media aritmética
b) Hallar la media geométrica
c) Hallar la media armónica
d) Hallar la media cuadrática
e) Relación entre estas medias.

16. El número de individuos muertos por cólera en un determinado páis por año, en el transcurso de 11 años han sido: 2, 17, 5, 8, 12, 3, 2, 8, 12, 12, 3.

f) Hallar la media aritmética
g) Hallar la media geométrica
h) Hallar la media armónica
i) Hallar la media cuadrática
j) Relación entre estas medias.

17. El número de pétalos de 13 flores de una determinada especie es el siguiente: 8, 10, 6, 5, 8, 11, 8, 10, 7, 10, 7, 10, 9

a) Calcular la mediana
b) Calcular la moda
c) Calcular los cuartiles de primer y tercer orden
d) Calcular el recorrido intercuartilico
e) Representar el Box Plot.

18. Considerando el valor teórico del metabolismo basal igual a 100, los valores observados en 50 individuos han dado los siguientes resultados.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 102 | 98 | 93 | 100 | 98 | 105 | 115 | 110 | 99 | 120 |
| 115 | 130 | 100 | 86 | 95 | 103 | 105 | 92 | 99 | 134 |
| 116 | 118 | 88 | 102 | 128 | 99 | 119 | 128 | 110 | 130 |
| 112 | 114 | 106 | 114 | 100 | 116 | 108 | 113 | 106 | 105 |
| 120 | 106 | 110 | 100 | 106 | 117 | 109 | 108 | 105 | 106 |

Calcular agrupando los datos en 10 clases, los siguientes valores:

a) Media aritmética
b) Mediana
c) Moda
d) Cuartiles de primer y tercer orden
e) Recorrido intercuartílico
f) Diagrama de Tukey de caja y bigotes

19. La distribución por pesos de 70 empleados de un hospital se expresa en la tabla siguiente:

| Kg. | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 | 100-110 |
|---|---|---|---|---|---|---|
| Nº empl. | 8 | 15 | 21 | 18 | 7 | 1 |

Calcular la media aritmética, la mediana y la moda.

20. Dada la siguiente distribución, ¿qué centil corresponde a 222?. ¿Qué centil corresponde a 230?

| Intervalo | 210-215 | 215-220 | 220-225 | 225-230 | 230-235 |
|---|---|---|---|---|---|
| $n_i$ | 2 | 10 | 11 | 5 | 2 |

***Medidas características de dispersión:***

El simple conocimiento de las medidas de centralización no sólo es insuficiente para darnos una idea de cómo están los datos distribuidos, sino que incluso puede llegar a ser engañoso. Pensemos en una población donde la media aritmético del sueldo de sus individuos es 500.000 ptas; enseguida nos viene a la mente que la gran mayoría de la población gana una cantidad entorno a esta cifra; sin embargo pudiera ocurrir que la mitad de la población gana 1.000.000 y la otra mitad nada. De este modo vemos que en ambos casos: Riqueza bien repartida y mal distribuida son dos casos en los que la media coincide, sin embargo las realidades son opuestas.

El ejemplo anterior demuestra que se hace necesario disponer de información acerca de cómo están distribuidos los datos alrededor de las medidas de centralización; dicho de otro modo..¿qué alejados o dispersos están los datos?.

Esto conduce a las medidas de dispersión, entre las que se encuentran las siguientes:

**Recorrido o rango**:

Ya utilizado con anterioridad, el recorrido de una variable estadística es la diferencia entre el valor más alto y el más bajo de dicha variable.

**Desviación a la media de la modalidad $x_i$.**

La desviación a la media de la modalidad a $x_i$, mide la distancia entre dicho valor $x_i$ y la media, por tanto su valor viene determinado por:

$$d_i = |x_i - \bar{x}|$$

Es importante destacar que dicha medida viene dada por el valor absoluto de la diferencia, ya que si omitiésemos dicha función, podría ocurrir que alguna medida fuese negativa y estos sería contradictorio con el concepto de distancia (alejamiento o dispersión) como un valor positivo.

Obsérvese que esta definición es válida solamente para los valores de la modalidad aislados, por tanto $d_i$ es independiente de la frecuencia $n_i$, asi como del resto de valores, por lo que es una medida de dispersión que no nos da una información de conjunto. Esto nos los cubre la siguiente medida.

**Desviación media**

Es la media aritmética de las desviaciones a la media $d_i$, esto es:

$$D_m = \sum_{i=1}^{p} \frac{d_i . n_i}{N} = \sum_{i=1}^{p} \frac{|x_i - \bar{x}| . n_i}{N}$$

**Varianza**

En ocasiones se suelen primar los alejamientos grandes y minimizar las pequeños dispersiones, esto produce la medida característica llamada varianza que no es más que la media cuadrática de las desviaciones a la media, esto es:

$$\sigma^2 = \sum_{i=1}^{p} \frac{d_i^{\,2}.n_i}{N} = \sum_{i=1}^{p} \frac{(x_i - \bar{x})^2 n_i}{N}$$

Esta medida tiene el inconveniente que viene expresada en unidades de la variable al cuadrado. Para evitar este problema se extrae la raiz cuadrada, y el valor obtenido se denomina:

La varianza también puede expresarse del siguiente modo:

$\sigma^2 = \sum_{i=1}^{p} \frac{(x_i - \bar{x})^2 n_i}{N} = \sum_{i=1}^{p} (x_i - \bar{x})^2 f_i = \sum_{i=1}^{n} x_i^{\,2}.f_i - \bar{x}^2$. (La media aritmética de los cuadrados de los $x_i$ menos el cuadrado de la media aritmética de los $x_i$

**Desviación típica:**

Llamada también desviación standard o desviación cuadrática media, es la <u>raiz cuadrada (positiva) de la varianza</u>

$$\sigma = \sqrt{\sum_{i=1}^{p} \frac{(x_i - \bar{x})^2 n_i}{N}}$$

En Excel es la función **DESVESTPA**

**Coeficiente de variación de Pearson:**

En ocasiones se hace necesario comparar dos o más distribuciones de datos de distinta naturaleza. Supongamos que queremos saber que distribución está más dispersa en los siguientes casos:

Una variable estadística X de pesos de individuos de media 170 cm. y desviación típica 20 cm., o una variable Y de producción de una ganadería vacuna de media 30 litros y desviación típica 6,5 litros.

Obviamente no se pueden comparar litros con centímetros, pero para saber comparativamente cuál está más concentrada o tiene los datos menos dispersos con respecto a la media, se utiliza el llamado coeficiente de variación de Pearson

$$v = \frac{\sigma}{\bar{x}}$$

Es por tanto una medida adimensional y establece un valor de relación independiente de las unidades en las que se esté trabajando. A mayor coeficiente, mayor dispersión de datos.

***MOMENTOS***

Generalizando lo anterior, tanto en las medidas de centralización como en las de dispersión, podemos definir los llamados momentos potenciales respecto al origen y respecto a la media. Son interesantes porque nos van a proporcionar unos valores para obtener más información acerca de la distribución.

**Momento de orden r respecto al origen:**

$a_r = \sum_{i=1}^{p} \frac{x_i^{\,r}.n_i}{N}$. Obsérvese que esto es una generalización de algunas medidas de concentración, ya que si r=1 se obtiene la media aritmética, si r=2 se obtiene la media cuadrática.

**Momento de orden r respecto a la media:**

$$m_r = \sum_{i=1}^{p} \frac{(x_i - \overline{x})^r.n_i}{N}$$

Si r=1 su valor es 0. Si r=2 obtenemos la varianza

**Relaciones entre momentos:**

Todos los momentos respecto a la media pueden expresarse en función de los momentos respecto al origen mediante las siguientes igualdades:

$m_2 = a_2 - a_1^2$ ; $m_3 = a_3 - 3a_2.a_1 + 2a_1^3$ ; $m_4 = a_4 - 4a_3.a_1 + 6a_2.a_1^2 - 3a_1^4$

Demostrar como ejercicio la primera igualdad.

***MEDIDAS DE FORMA***

Nos dan información acerca de cómo es la gráfica de la distribución o curva envolvente del histograma, entre las más importantes tenemos las siguientes:

***Medidas de asimetría o sesgo:***

Elaboran un indicador que permite establecer el grado de simetría (o asimetría) que presenta la distribución sin realizar su representación gráfica.

Pueden ser campaniformas y en forma de U y se llaman sesgadas a la derecha cuando la "cola" de la distribución se prolonga hacia la derecha. Análogamente para las sesgadas a la izquierda.

Para medir el sesgo tenemos:

***Coeficiente de asimetría de R.A. Fisher***

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

Si $g_1$= 0, la distribución es simétrica

Si $g_1$>0, la distribución es sesgada a la derecha.

Si $g_1<0$, la distribución es sesgada a la izquierda

***Coeficiente de K. Pearson***

$$A_p = \frac{\bar{x} - Mo}{\sigma}$$

Su sencillez tiene el inconveniente de que solamente es segura para distribuciones campaniformes unimodales y moderadamente asimétricas.

Decide según los mismos criterios que el coeficiente de Fisher.

***Coeficiente de asimetría de Bowley-Yule***

$$A_B = \frac{Q_3 + Q_1 - 2.Me}{Q_3 - Q_1}$$

Sigue los mismos criterios de signos que el de Fisher

***Coeficiente absoluto de asimetría:***

$$A_B = \frac{Q_3 + Q_1 - 2.Me}{\sigma}$$

Sigue los mismos criterios de signos que los anteriores

***Medidas de apuntamiento o curtosis***

Cuando la distribución es campaniforme, unimodal y ligeramente asimétrica puede ser —según la moda esté en baja, media o alta frecuencia— platicúrtica (achatada), mesocúrtica (sin exceso) y leptocúrtica (con exceso)

Esta característica nos la mide la curtosis o coeficiente de apuntamiento, cuyo valor viene dado por:

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

Si $g_2 = 0$. la distribución es mesocúrtica (o sin exceso)

Si $g_2 > 0$. la distribución es leptocúrtica (o con exceso)

Si $g_2 < 0$. la distribución es platicúrtica (o achatada)

## EJERCICIOS PROPUESTOS

1.) Calcular la desviación típica y varianza de las distribuciones de los ejercicios que figuran en las páginas 13, 14 y 15 de estos apuntes.

2) Dada la siguiente distribución, calcular los cuatro primeros momentos respecto al origen

| | | | |
|---|---|---|---|
| $x_i$ | -2 | 1 | 3 |
| $n_i$ | 2 | 1 | 1 |

3) Dada la siguiente distribución, calcular los cuatro primeros momentos respecto a la media

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | 4 | 5 |
| $n_i$ | 3 | 10 | 4 | 2 | 1 |

4) Dada la siguiente distribución, calcular el tercer y cuarto memento respecto a la media a partir de los momentos respecto al origen.

| | | | | | | |
|---|---|---|---|---|---|---|
| $x_i$ | -2 | -1 | 0 | 1 | 2 | 3 |
| $n_i$ | 2 | 4 | 6 | 5 | 2 | 1 |

5) Dada la siguiente distribución de frecuencias, calcular el coeficiente de asimetría de Fisher y su curtosis.

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 0 | 10 | 20 | 30 | 40 |
| $n_i$ | 2 | 4 | 7 | 5 | 2 |

6) Calcular el coeficiente de asimetría de Pearson de la siguiente distribución de frecuencias.

| | | | | | | |
|---|---|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
| $n_i$ | 2 | 8 | 3 | 5 | 7 | 5 |

7) La distribución por intervalos de un test de Economía realizado a 1230 opositores puntuando de 0 a 800, da los siguientes resultados:

| PUNTUACIÓN TEST | % OPOSITORES |
|---|---|
| Hasta 60 | 6,30 |
| 60-84 | 9,83 |
| 84-120 | 20,87 |
| 120-180 | 27,63 |
| 180-240 | 15,00 |
| 240-480 | 18,27 |
| 480-700 | 1,65 |
| Mas de 700 | 0,45 |

Calcular los cuartiles y el coeficiente de asimetría de Bowley-Yule

# TEMA 2. DISTRIBUCIONES BIDIMENSIONALES. CORRELACIÓN Y REGRESIÓN.

**Variables estadísticas bidimensionales:**

Son los resultados de la observación de un fenómeno respecto de dos características.

A estas variables estadísticas se le denominan bidimensionales.

Se representan por el par de valores (X,Y), siendo X una variable unidimensional que toma los valores $x_1$, $x_2$, ... ,$x_p$ e Y una variable unidimensional que toma los valores $y_1$, $y_2$ ... $y_k$

Las variables estadísticas bidimensionales toman los valores

$(x_1,y_1)$, $(x_2,y_2)$....$(x_p,y_k)$

**Representación gráfica.**

Cuando los valores x e y no están agrupados en intervalos, a cada valor x (eje de abscisas) y a cada valor y (eje de ordenadas), tomados conjuntamente les corresponde un punto en el plano. Este conjunto de puntos se denomina "nube de puntos" o "diagrama de dispersión".

Si cada pareja tuviera una frecuencia distinta de la unidad, se puede expresar con un número, la frecuencia correspondiente al lado del punto o contruir una gráfica tridimensional donde la tercera dimensión z, representaría la frecuencia.

Si los datos están agrupados en intervalos, la representación gráfica tendríamos el plano dividido en p.k rectángulos, siendo h el numero de intervalos de x y k el numero de intervalos de y y sobre cada uno de estos rectángulos se levanta un prisma cuya altura es proporcional a la frecuencia de la pareja $(x_i,y_i)$

**Tablas de frecuencias.**

**Tabla de columnas:**

Si la frecuencia de cada par $(x_i,y_i)$ es 1

| $x_i$ | $y_i$ | $x_i.y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|
| ------------ | --------------- | ---------------- | ---------------- | --------------- |
| --------------- | --------------- | -------------- | ---------------- | -------------- |
| totales | totales | totales | totales | totales |

Si las frecuencias no son 1, habría que añadir una columna con los valores de dicha frecuencia $n_i$.

**Tabla de doble entrada:**

| | $y_1$ | $y_2$ | | | $y_k$ | |
|---|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ | | | $n_{1k}$ | |
| $x_2$ | $n_{21}$ | $n_{22}$ | ........ | | $n_{2k}$ | |
| | | | | | | |
| $x_p$ | $n_{p1}$ | $n_{p2}$ | | | $n_{pk}$ | |
| | | | | | | N |

En las zones sombreadas se consignarán las sumas de los $n_{ij}$ por filas y columnas, teniendo que sumar por ambos lados N

**Distribuciones marginales:**

Se llaman así a las distribuciones de las dos variables que intervienen X e Y, consideradas de forma aislada.

Según esto las distribuciones marginales son las que figuran en la zona sombreada de la gráfica de doble entrada anterior.

**Momentos en distribuciones bidimensionales:**

**Respecto al origen:**

Momento de orden r, s respecto al origen:

$$a_{rs} = \sum_{i=1}^{p}\sum_{j=1}^{k} x_i^r \cdot y_j^s \cdot \frac{n_{ij}}{N}$$

$a_{01}$ es la media de y, mientras que $a_{10}$ es la media de x.

**Momentos de orden r, s respecto a las medias**

$$m_{rs} = \sum_{i=1}^{p}\sum_{j=1}^{k} (x_i - \overline{x})^r \cdot (y_j - \overline{y})^s \cdot \frac{n_{ij}}{N}$$

Según esta definición $m_{10} = m_{01} = 0$

$m_{02}$ es la varianza de y

$m_{20}$ es la varianza de x

$m_{11}$ se denomina covarianza.

Por tanto la covarianza es:

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

Suele representarse por $\sigma_{xy}$ o $S_{xy}$

Se puede demostrar que $\sigma_{xy} = \overline{x.y} - \overline{x}.\overline{y}$, es decir la media del producto menos el producto de las medias. (Hágase como ejercicio)

Cuando dos variables X e Y son independientes, la covarianza es 0

**Correlación**

Es la teoría que estudia la relación de dependencia entre las dos variables (x, y) de una distribución bidimensional.

Hay varios tipos de correlación, pero nos vamos a centrar en la lineal, esto es cuando la nube de puntos se condensa más o menos en torna a una linea recta.

Podemos distinguir:

**Correlación positiva o directa:** Cuando una variable crece también lo hace la otra. En la correlación lineal esto se traduciría en que la recta sería creciente, o lo que es lo mismo, de pendiente positiva.

**Correlación negativa o inversa:** Cuando una variable crece la otra decrece. En la correlación lineal esto se traduciría en que la recta sería decreciente, o lo que es lo mismo, de pendiente negativa.

**Correlación nula:** Cuando no existe ninguna relación entre las variables. Se dice que las variables están incorreladas.

**Regresión matemática:**

Es el resultado de sustituir la nube de puntos o diagrama de dispersión de una distribución bidimensional por la función matemática que mejor se aproxima a ella. Nosotros vamos a centrarnos en la regresión lineal solamente, que se da cuando la función que se ajusta a la nube de puntos es una recta.

**Recta de regresión de y sobre x:**

Es la recta que hace mínimos la suma de los cuadrados de las diferencias entre los valores observados experimentalmente $y_i$ y los teóricos y que se obtienen mediante la recta, medidos paralelamente al eje Y (mínimas sumas $(y_i\text{-}y)^2$. Su ecuación es :

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

**Recta de regresión de x sobre y:**

Es la recta que hace mínimos la suma de los cuadrados de las diferencias entre los valores observados experimentalmente $x_i$ y los teóricos x que se obtienen mediante la recta, medidos paralelamente al eje X (mínimas sumas $(x_i\text{-}x)^2$. Su ecuación es :

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

Es importante hacer notar que ambas rectas se cortan en el punto $(\bar{x}, \bar{y})$, llamado centro de gravedad de la distribución conjunta.

**Coeficientes de regresión:**

Se llaman así a las pendientes de las rectas anteriores, que son:

$b = \dfrac{\sigma_{xy}}{\sigma_x^2}$ que es el coeficiente de regresión de y sobre x y representaremos por $\beta_{21}$

$b' = \dfrac{\sigma_{xy}}{\sigma_x^2}$ que es el coeficiente de regresión de sobre y, y representaremos por $\beta_{12}$

Una propiedad importante de estos coeficientes es que el producto de ambos coeficientes es menor que 1.

Coeficiente de correlación lineal:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Puesto que $-1 \le r \le 1$, se obtiene lo siguiente:

-Si $r= 1$: Correlación perfecta positiva (funcional)
-Si $r=-1$: Correlación perfecta negativa (funcional)
-Si $r=0$ : Correlación nula, las rectas son $y = \bar{y} \quad x = \bar{x}$
-Si $-1<r<0$ :Correlación negativa, más fuerte cuanto más se aproxime el valor a –1
-Si $0<r<1$ :Correlación positiva, más fuerte cuanto más se aproxime el valor a 1

**Problemas propuestos**

1º) Dada la siguiente distribución que representa a 60 hombres de la misma edad con respecto a los caracteres: altura (x) y peso (y)

| | 1,55-1,65 | 1,65-1,75 | 1,75-1,85 |
|---|---|---|---|
| 55-65 | 3 | 2 | 0 |
| 65-75 | 6 | 10 | 4 |
| 75-85 | 4 | 11 | 5 |
| 85-95 | 1 | 6 | 8 |

Calcular: La distribución marginal de la variable y. Media y varianza marginales de la variable y.

2º) Dada la siguiente distribución bidimensional:

| | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 0 | 2 | 5 | 10 | 12 |
| 1 | 6 | 10 | 28 | 8 |
| 2 | 15 | 12 | 6 | 6 |

Se piden:
a) Distribuciones marginales
b) Media y moda de las distribuciones marginales
c) Desviación típica y asimetría de las distribuciones marginales.

3º) Dadas las siguientes series de valores de las variables x e y, calcular la recta de regresió de y xobre x y de x sobre y

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 2 | 3 | 6 | 10 | 12 |
| $y_i$ | 150 | 130 | 125 | 120 | 100 |

4º) Dada la siguiente distribución bidimensional, obtener las rectas de regresión y el coeficiente de correlación:

| | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| 10 | 2 | | | |
| 20 | 5 | 4 | 1 | |
| 30 | 3 | 8 | 6 | 3 |
| 40 | | 3 | 6 | 6 |
| 50 | | | 2 | 1 |

5º) Las notas obtenidas por 10 universitarios en Anatomía y Fisiología son:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Anatomía | 6 | 4 | 4 | 8 | 5 | 3,5 | 7 | 5 | 10 | 5 |
| Fisiología | 6,5 | 4,5 | 5 | 7 | 5 | 4 | 8 | 7 | 10 | 6 |

Se pide: Calcular las rectas de regresión. Dibujarlas y dibujar la nube de puntos o diagrama de dispersión.

Calcular el coeficiente de correlación

¿Cuál sería la nota esperada en Fisiología de un universitario que haya obtenido 8,3 en Anatomía.

6º) Dada la siguiente distribución

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $x_i$ | 2 | 2 | 2 | 4 | 7 | 7 | 10 | 10 |
| $y_i$ | 3 | 4 | 5 | 5 | 4 | 5 | 3 | 5 |
| $n_i$ | 5 | 10 | 17 | 19 | 20 | 16 | 9 | 4 |

Calcular la covarianza, la recta de regresión de y xobre x. Estudiar la dependencia lineal entre ambas variables.

7º)Las rectas de regresión de dos variables aleatorias x e y son :

$2x + y = 7 \qquad 2x + 3y = 13$

Calcular los valores medios, los coeficientes de regresión y la correlación.

# TEMA III. COMBINATORIA

**Factorial de un número:**

Dado un número entero $n \geq 1$, se llama factorial de *n, al siguiente valor:*

$$n! = n(n-1)(n-2)...3.2.1$$

Por convenio se establece que 0!=1.

Dados m elementos $x_1$, $x_2$, $x_3$, ...., $x_m$, podemos definir los siguientes conceptos:

**Variaciones ordinarias (sin repetición) de m elementos tomados de p en p.**

Son el conjunto formado por todas las colecciones de p elementos distintos elegidos de entre los m dados, considerando distintas dos colecciones cuando se diferencian en algún elemento o cuando teniendo los mismos elementos, difieren en el orden de colocación.

El número de las variaciones ordinarias se representará por Vm,n, y su valor es:

$Vm,p = \dfrac{m!}{(m-p)!}$ . Otro valor de este cociente es: $Vm,n = m(m-1)...(m-p+1)$

Es obvio que de la definición se desprende que $p \leq m$

**Permutaciones de m elementos**

Cuando m=p, las Variaciones anteriores resultan ser las colecciones de los m elementos (por tanto se toman todos) elegidos en las condiciones ya indicadas, es decir Vm,m. A estas colecciones se les denominan Variaciones de m elementos y su valor es:

$$Pm = Vm,m = \frac{m!}{0!} = m!$$

**Variaciones con repetición de m elementos tomados de p en p.**

Son el conjunto formado por todas las colecciones de p elementos elegidos de entre los m dados, pudiendo estos repetirse, considerando distintas dos colecciones cuando se diferencian en algún elemento o cuando teniendo los mismos elementos, difieren en el orden de colocación.

Al introducir la posibilidad de repetición de los elementos, p puede ser mayor que m.

Su número viene dado por:

$$RVm,p = m^p$$

**En calculadora.**

Las Variaciones ordinarias en la calculadora vienen determinadas en la tecla nVr. Para las permutaciones se utilizaría también nVr, considerando n=r, pero la mayoría de las marcas tiene la función n!

Por tanto, si quisiéramos obtener $V_{3,2}$ pulsaríamos 3 nVr 2 =

Si quisiéramos obtener $P_3$, pulsaríamos 5 nVr 5 = , o bien directamente 5 n!

**En Excel**

La función que en Excel determina Vm,p y Pm es **PERMUTACIÓN**, donde hay que establecer dos parámetros "número" y "tamaño", el número sería m y el tamaño p. En caso de ser una permutación ambos parámetros serían **m.**

**Permutaciones con repetición.**

Llamaremos permutaciones con repetición de m elementos entre los que hay a iguales entre sí, otros b iguales entre sí,..., siendo a+b+...+c=p, a los distintos grupos que se pueden formar con los m objetos, entre los que aparecen repetidos a, b, ...,c elementos, considerando distintos dos permutaciones cuando difieren en el orden de colocación.

Su número es:

$$P_m^{a,b,\ldots,c} = \frac{(a+b+\ldots c)!}{a!.b!\ldots c!}$$

Es destacable observar que en las variaciones con repetición, un elemento puede repetirse un número variable de veces, que llega hasta el orden de dicha variación. Esto no sucede, en cambio, en las permutaciones con repetición, donde el número de veces que aparece repetido cada elemento es siempre el mismo.

Por tanto para distinguir unas de otras, simplemente hemos de preguntarnos si la repetición de los elementos es fija o variable; en el primer caso estaríamos ante unas permutaciones con repetición y en el segundo serían variaciones con repetición.

**Combinaciones de m elementos tomadas de p en p.**

Son todos los grupos posibles que se pueden formar con p distintos tomados de los m elementos dados, de modo que dos grupos cualesquiera difieran en algún elemento, es decir que no importa el orden de colocación de los elementos como ocurría en las variaciones.

Su número viene determinado por:

$Cm,p = \dfrac{m!}{(m-p)!\,p!}$ ,a este valor se le denomina número combinatorio y se representa por $\begin{pmatrix} m \\ p \end{pmatrix}$.

Los números combinatorios se obtienen fácilmente utilizando el triángulo de Pascal-Tartaglia:

| **m** | | | | | | **p** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | | | | | | 1 | | | | | |
| **1** | | | | | 1 | | 1 | | | | |
| **2** | | | | 1 | | 2 | | 1 | | | |
| **3** | | | 1 | | 3 | | 3 | | 1 | | |
| **4** | | 1 | | 4 | | 6 | | 4 | | 1 | |
| **5** | 1 | | 5 | | 10 | | 10 | | 5 | | 1 |

Propiedades de los números combinatorios:

1.- $\begin{pmatrix} m \\ 0 \end{pmatrix} = 1$

2.- $\begin{pmatrix} m \\ p \end{pmatrix} + \begin{pmatrix} m \\ p+1 \end{pmatrix} = \begin{pmatrix} m+1 \\ p+1 \end{pmatrix}$

3.- $\binom{m}{p} = \binom{m}{m-p}$

Un caso donde intervienen los números combinatorios. El binomio de Newton:

$$(a+b)^m = \sum_{p=0}^{m} \binom{m}{p} . a^p . b^{m-p}$$

**En calculadora**

La tecla correspondiente a las combinaciones o números combinatorios es nCr y se utiliza exactamente igual que en el caso nVr.

**En Excel**

La función a usar es COMBINAT, indicando "numero" y "tamaño", es decir m y p.

**Combinaciones con repetición de m elementos tomados de p en p.**

Son las distintas agrupaciones de p elementos iguales o distintos que se pueden formar con los m elementos dados, de modo que cada dos combinaciones con repetición difieran al menos en un elemento.

Su número viene dado por:

$$RCm, p = \binom{m+p-1}{p}$$

**Problemas propuestos de Combinatoria (Nivel 1)**

1) Una carrera en la que participan seis atletas se realiza con el fin de clasificar a tres de ellos (para posteriores competiciones); los restantes corredores quedan eliminados. ¿Cuántos resultados distintos pueden haber, en principio? (Se supone que no hay posibilidad de empate).

2) Si en la carrera anterior hubieran tres premios: medalla de oro, medalla de plata y medalla de bronce, ¿cuál sería entonces el resultado?

3) ¿Cuál sería el resultado del ejercicio anterior si hubieran seis premios distintos?

4) ¿Cuántas palabras de 7 letras distintas pueden escribirse con las letras de la palabra CADAQUES?

5) En una jornada de fútbol, ¿cuántas quinielas de una apuesta, distintas, se podrían hacer?

6) En un estante de una librería capaz para 25 volúmenes, hay siete ejemplares iguales de "El Quijote", ocho ejemplares iguales de "La Celestina" y diez ejemplares iguales de "La venganza de Don Mendo". ¿De cuántas maneras diferentes (incluso en desorden) pueden colocarse dichos libros.

7) En una bolsa hay diez bolas rojas, diez azules y diez negras, todas ellas del mismo tamaño y calida. ¿De cuántas maneras pueden extraerse diez bolas de dicha bolsa:

a) Si importa el orden en que se extraigan
b) Si no importa el orden en que se extraigan.

8) ¿Cuántos productos diferentes, con cuatro factores, se pueden formar con los números primos comprendidos entre 2 y 19, ambos inclusive.

a) sin repetir ningún factor?
b) repitiendo si se desea?

9) ¿De cuántas maneras posibles pueden colocarse cuatro soldados en una fila.

10) Un camarero descandsa dos días cualesquiera por semana; ¿cuántas semanas podrán transcurrir para que no se repitan los dos días de descanso?

11) ¿Cuántas permutaciones se pueden formar con las letras de la palabra permutación? ¿Cuántas terminarán en n y cuántas empezarán por per?

12) Con las cifras del número 257836, ¿cuántos números distintos de tres cifras se pueden formar?

a) no entrando repetida ninguna de las cifras?
b) repitiendo si se desea?

13) Con seis pesas de 1, 2, 5, 10, 20, 50 kilogramos, ¿cuántas pesadas diferentes pueden obtenerse tomando aquéllas de tres en tres?

14) ¿Cuántos números distintos de ocho cifras se pueden escribir, de modo que parezcan dos unos, tres cuatros y tres ochos?

15) En unas elecciones para nombrar cinco elnlaces sindicales se presentaron 15 candidatos; cada productor, para hacer erectivo su voto escribió en un papel 5 nombres (no importaba el orden de su colocación) y se dio la extraña circunstancia de que se registro un empate total entre los 15 candidatos, porque todos los votos fueron distintos. ¿Cuántos productores trabajan en la fábrica?

16) Con las letras de la palabra valdespino, ¿cuántas "palabras" de cuatro letras distintas pueden escribirse? De éstas, ¿cuántas empezarán por d y terminarán en a?

17) ¿Cuántos números distintos de seis cifras se pueden formar con las cifras del número 375261, de modo que empiecen y terminen en 1.

18) Con 20 soldados, ¿cuántas guardias diferentes de a cuatro soldados cada una pueden formarse, y en cuántas entrará un soldado determinado J.?

19) En una clase hay 15 alumnos y 18 mesas, supongamos que cada día se distribuyen de forma distinta, on respeco a los días anteriores ¿Durante cuantos días puede mantenerse esta situación?

20) Se extrae una carta de una baraja de 52 cartas; se introduce nuevamente. Se repite esta operación tres veces más. ¿Cuántos resultados distintos puede obtenerse:

a) Si importa el orden
b) Si no importa el orden.

21) Supongamos que en el ejercicio anterior las cartas extraidas no vuelvan a la baraja ¿Cuáles son entonces los resultados?

22) ¿De cuantas maneras pueden colocarse en fila cuatro monedas de 1 euro, cuatro de 2 euros y cuatro de 50 céntimos

23) ¿Cuántos modelos de billete de tren se deben imprimir para cubrir un trayecto de diez estaciones, si en cada billete ha de figurar la estación de salida en primer lugar, y la llegada en segundo lugar?

24) Una hormiga desea ir desde el extremo inferior izquierdo de un tablero de ajedrez, hasta el extremo superior derecho, recorriendo la mínima distancia posible y con la condición de pasar unicamente por los bordes de los escaques (cuadrados), nunca en diagonal. ¿De cuántas formas distintas puede hacerlo?

**Problemas propuestos de Combinatoria (Nivel 2)**

1) Sea A = {1,2,3,4,5,6,7}. Se pide:

a) El número de subconjuntos de A con tres elementos.
b) El número de sobconjuntos de A con cuatro elementos.
c) El número de subconjuntos de A con tres elementos como máximo
d) El número de subconjuntos de A con cuatro elementos como mínimo.
e) El número de subconjuntos de A con cuatro elementos, dos pares y dos impares
f) ¿Cuántos números distintos de cuatro cifras pueden escribirse con las cifras de A de modo que empiecen y terminan en cifra impar, y que las restantes cifras sean números pares? De éstos, ¿cuántos tienen todas sus cifras distintas?

2) De 7 españoles y 4 franceses se va a elegir un comité de 6 personas. ¿De cuántas maneras pueden formarse?
a) De modo que hayan exactamente dos franceses
b) De modo que hayan dos franceses como mínimo

3) ¿Cuántas diagonales tiene un polígono convexo de n lados?

4) Con un piano de juguete de 24 notas, ¿cuántos sonidos diferentes pueden conseguirse empleando cada vez cuatro notas como máximo.

5) En un banquete la mesa de la presidencia es rectangular y tiene ocho cubiertos preparados, todos ellos a un mismo lado. ¿De cuántas maneras distintas pueden sentarse los ocho comensales?

6) Hay dos obras de tres volúmenes cada una y otras dos de dos voúmenes cada una. ¿De cuántas maneras pueden colocarse los diez libros en un estante capaz para diez libros, si deben quedar de tal modo que no se separen los volúmenes de una misma obra?

7) Un entrenador dispone, para formar un equipo de futbol, de 3 porteros, 5 defensas, 6 medios y 9 delanteros. ¿Cuántos equipos diferentes puede formar:
a) Suponiendo que los distintos puestos de defensa medio y delantero son "equivalentes" entre sí, respectivamente (es decir, sin hacer distinción entre los tres defensas: central, izquierdo y derecho; y haciendo lo mismo para los dos medios y para los cinco delanteros)?
b)No suponiéndolo?

8) Un depósito de agua tiene cinco caños de desagüe, que arrojan 1,2,4,10 y 20 litros por minuto. Abriendo indistintamente cuatro de estos caños, ¿en cuántos tiempos diferentes se puede vaciar el depósito?

9) A una clase de matemáticas asisten 24 alumnos, y todos los días resuelven problemas en la pizarra dos de ellos. El profesor desea que durante el curso no salga dos veces la misma pareja; ¿es esto posible?

# TEMA IV. ÁLGEBRA DE SUCESOS

Sea un experimento aleatorio. Definimos los siguientes conceptos:

**Suceso elemental**.- Es cada uno de los resultados directos del experimento aleatorio.

Por ejemplo si lanzamos un dado, los sucesos elementales son 1,2,3,4,5,6.

**Espacio muestr**al.- Es el conjunto formado por los sucesos elementales de un experimento aleatorio. Suele representarse con la letra $\Omega$.

**Suceso**.- Es cualquier subconjunto de un espacio muestral $\Omega$ y suelen representarse con letras mayúsculas A,B,C...etc

De este modo puedo afirmarse que si tenemos n sucesos elementales, el conjunto de sucesos, representado por P($\Omega$), tendrá $2^n$ elementos. (se deduce fácilmente por combinatoria)

**Suceso imposible**.- Es el que nunca se verifica dentro del experimento aleatorio. Se representa por $\varnothing$.

**Suceso seguro**.- Es el que siempre se verifica y coincide con $\Omega$.

**Sucesos incompatibles**.- Son aquellos que no puedes realizarse simultáneamente.

**Sucesos independientes**.- Dos sucesos se dice que son independientes, cuando la realización de uno no influye en la realización del otro.

**Operaciones con sucesos:**

**a) Unión de sucesos:**

Dados dos sucesos A y B, asociados a un espacio muestral $\Omega$ de un experimento aleatorio, se define $A \cup B$, como el suceso que se verifica cuando se verifica A, B o ambos. (Es el equivalente a lo que en lógica se denominaría una disyunción inclusiva).

**b) Intersección de sucesos:**

Dados dos sucesos A y B, asociados a un espacio muestral $\Omega$ de un experimento aleatorio, se define $A \cap B$, como el suceso que se verifica cuando se verifica Ay B simultáneamente. (Es el equivalente a lo que en lógica se denominaría una conjunción).

Según esta definición se deduce que dos sucesos A y B son incompatibles si y sólo si se verifica que $A \cap B = \varnothing$

**c) Negación o complementario:**

Se denomina complementario del suceso A, y se representa por $\rceil$A, al suceso que se verifica cuando no se verifica A.

**d) Diferencia de sucesos:**

Dados dos sucesos A y B, asociados a un espacio muestral $\Omega$ de un experimento aleatorio, se define A- B, como el suceso que se verifica cuando se verifica A y no se verifica B. Esta operación se puede definir en términos de unión intersección y negación, del siguiente modo:

$A - B = A \cap \rceil B$

**Propiedades de las operaciones:**

Para cualesquiera sucesos A, B, C de un experimento aleatorio, se tiene:

1) Commutativa : $A \cap B = B \cap A$ ; $A \cup B = B \cup A$
2) Asociativa : $A \cap (B \cap C) = (A \cap B) \cap C$ ; $A \cup (B \cup C) = (A \cup B) \cup C$
3) Distributivas : $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4) Idempotentes: $A \cap A = A$ ; $A \cup A = A$
5) Leyes de Morgan: $\rceil(A \cup B) = \rceil A \cap \rceil B$ ; $\rceil(A \cap B) = \rceil A \cup \rceil B$
6) Simplificativa o de absorción $A \cup (A \cap B) = A$ ; $A \cap (A \cup B) = A$
7) De identidad: $A \cap \varnothing = \varnothing$ ; $A \cup \varnothing = A$ ; $A \cup \Omega = \Omega$ ; $A \cap \Omega = A$
8) Del complementario:$A \cup \rceil A = \Omega$ ; $A \cap \rceil A = \varnothing$ ; $\rceil\rceil A = A$ ; $\rceil \Omega = \varnothing$ ; $\rceil \varnothing = \Omega$

**Frecuencias absolutas y relativas de un suceso**

Dado un experimento aleatorio, con un espacio muestral Ω, que realizamos N veces, y dado un suceso de dicho experimento que denotaremos por A, llamamos frecuencia absoluta del suceso A, al número de veces que dicho suceso se verifica en el experimento y llamamos frecuencia relativa de A, a la proporción con que dicho suceso se verifica.

Por tanto: $f_a(A) = n$ y $f_r(A)=n/N$

Son propiedades de la frecuencia relativa, las siguientes:

**1.** $0 \le f_r$ **(A)** $\le$ **1**

**Demostración:**

Puesto que $0 \le f_a(A) \le N$, dividiendo por N, se obtiene $0 \le f_r(A) \le 1$

**2.** $f_r(\Omega) = 1$**;** $f_r(\varnothing) = 0$

**Demostración:**

Es trivial teniendo en cuenta que $f_a(\Omega)=N$ y que $f_a(\varnothing)=0$

**3.** $f_r(A) + f\ (\neg A) = 1$

**Demostración:**

Si A aparece n veces, la frecuencia de $\rceil A$ es N-n, por lo que se tiene que

$$f_r(A) + f_r(\neg A) = \frac{n}{N} + \frac{N-n}{N} = 1$$

**4.** $f_r(A \cup B) = f_r(A) + f_r(B) - f_r(A \cap B)$

**Demostración:**

Sea $f_a(A)=n_a$ ; $f_a(B)=n_{b;}$ ; $f_a(A \cap B)=n_{ab}$. ; $f_a(A \cap \rceil B)=n_{a\rceil b}$ ; $f_a(\rceil A \cap B)=n_{\rceil ab}$

Según esta notación podemos afirmar los siguiente:

$$f_r(A \cup B) = \frac{n_{ab} + n_{\neg ab} + n_{a\neg b}}{N} \quad (1)$$

Ahora bien, dado que $n_a = n_{a\rceil b} + n_{ab}$ y $n_b = n_{\rceil ab} + n_{ab,}$ resulta que

$$(1) = \ f_r(A \cup B) = \frac{n_{ab} + n_a - n_{ab} + n_b - n_{ab}}{N} = f_r(A) + f_r(B) - f_r(A \cap B)$$

**5.** Si A y B son dos sucesos incompatibles $f_r(A \cup B) = f_r(A) + f_r(B)$
Demostración:
Es obvia basándose en la propiedad anterior y teniendo en cuenta que si A y B son incompatibles entonces A $\cap$ B = $\varnothing$

**6..** $Si \quad A \subset B, \quad f_r(A) \leq f_r(B)$
**Demostración:**
$Si \quad A \subset B, \quad B = A \cup (B - A)$ y teniendo en cuenta que A y B-A son dos sucesos incompatibles, aplicando la propiedad anterior resulta que
$f_r(B) = f_r(A) + f_r(B - A)$, dado que $P(B - A) \geq 0,$ resulta que $f_r(A) \leq f_r(B)$

**Ejercicios propuestos**

1) Si lanzamos tres monedas al aire. Se pide:

a) ¿Cuántos elementos tiene el espacio muestral?
b) ¿Cuántos sucesos tiene el experimento?
c) Describe los sucesos A = "que salgan al menos dos caras"; B = "que salga al menos una cruz"
d) ¿Cuál es el suceso complementario de B?
e) Expresar el suceso $A \cap B$

2) Sea el siguiente experimento aleatorio: En una urna hay una bola blanca y una negra. Se extrae una bola: si es blanca se devuelve a la urna y se vuelve a extraer otra bola, procediendo del mismo modo; es decir que si la segunda es blanca se vuelve a introducir en la urna y se vuelve a extraer otra bola...y asi sucesivamente hasta que salga una negra, en cuyo momento el experimento finaliza. ¿Cuál es el espacio muestral de dicho experimento? ¿Es finito o infinito? ¿Si es infinito es numerable o no?

3) Un experimento consiste en lanzar un dado. Se pide:

a) Describe el espacio muestral.
b) Sea A = "salir par"; B = "ser múltiplo de 3". Hallar $A \cap B$ y $\rceil B$

4) Si lanzamos 5 dados, ¿cuántos sucesos elementales tendremos? ¿En cuantos de ellos nos salen los cinco resultados impares?

5) Si lanzamos 6 monedas, ¿cuántos sucesos elementales tendremos?. ¿En cuantos casos distintos nos pueden salir exactamente dos caras?

6) Si extraemos dos bolas de una urna que contiene 4 rojas y 2 negras, ¿cuántos casos se pueden producir? ¿En cuántos de ellos nos saldrá una de cada color? ¿En cuantos casos nos saldrán las dos bolas rojas?

7) Lanzamos una moneda al aire. Si sale cara, lanzo un dado de 6 caras (un cubo de caras numeradas del 1 al 6) y si sale cruz lanzo un dado de cuatro caras (tetraedro de caras numeradas del 1 al 4). ¿Cuántos sucesos elementales tiene el experimento? ¿En cuántos casos nos saldrá el resultado del dado par?

8) Tenemos 6 cartas en seis sobres correspondientes. Se extraen las 6 cartas del sobre y se barajan, para introducirlas de nuevo al azar. ¿Cuántas introducciones distintas podemos realizar? ¿En cuántas de ellas obtendremos la situación de partida?

9) Cuántos casos distintos pueden producirse al lanzar al aire 4 monedas. ¿En cuántos de estos casos va a salir por lo menos una cara?

10) Pon un ejemplo de un par de sucesos que sean compatibles e independientes.

# TEMA V. PROBABILIDAD

Se comprueba empíricamente que la frecuencia relativa de un suceso de un experimento aleatorio se aproxima a un valor fijo al aumentar el número de experiencias. Esta propiedad, llamada ley del azar, fue inicialmente descubierta en los juegos de azar; al tirar una moneda al aire, la frecuencia relativa del suceso "cara" tiene, al aumentar el número de tiradas, hacia el valor constante ½. Posteriormente se observó esta propiedad en datos demográficos; así la frecuencia relativa del nacimiento de varones tiende a 0,51.

Estas experiencias condujeron en el siglo XIX a definir la probabilidad de un suceso como el valor límite de su frecuencia relativa al repetir indefinidamente el suceso. Esta definición presenta problemas importantes puesto que no es posible una experiencia indefinida; así en los años 30 del siglo XX, se definió la probabilidad como una función definida en el conjunto de sucesos que tiene las propiedades de la frecuencia relativa.

Esto conduce a la llamada definición axiomática de la probabilidad.

**Concepto de probabilidad:**

Sea Ω un espacio muestral asociado a un experimento aleatorio. Se define la probabilidad como una función que a cada suceso del conjunto de sucesos de Ω, le asigna un número real

$$P : \wp(\Omega) \longrightarrow \mathrm{R}$$

A      P(A) = "probabilidad del suceso A"

verificando los siguiente axiomas (proposiciones indemostrables dadas por válidas)

1. $0 \leq P(A) \leq 1, \quad \forall A \in \wp(\Omega)$
2. $P(\Omega) = 1$
3. Si $A \cap B = \varnothing \quad P(A \cup B) = P(A) + P(B)$

En estas condiciones, a la terna (Ω, $\wp(\Omega)$, $P$) se le denomina espacio de probabilidad.

**Propiedades de la probabilidad:**

Parecería lógico que debiera tener las mismas propiedades que la frecuencia relativa, y así va a ser, deduciéndose las mismas de los tres axiomas anteriores. Son las siguientes:

**1.-** $P(\neg A) = 1 - P(A) \quad , \forall A \in \wp(\Omega)$

**Demostración:**

$\Omega = A \cup \neg A \quad , P(\Omega) = P(A \cup \neg A) = P(A) + P(\neg A)$**, de donde** $P(\neg A) = 1 - P(A)$

**2.-** P(∅) = 0

**Demostración:**

$\varnothing = \neg\Omega$ , entonces por el axioma anterior tenemos que

P(∅) = 1-P(Ω) = 1-1 = 0

**3**.- Si $A \subset B$, $P(A) \leq P(B)$

**Demostración:**

$B = A \cup (B - A) \Rightarrow P(B) = P(A) + P(B - A) \Rightarrow P(A) \leq P(B)$

**4**.- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

**Demostración:**

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

$B = (B - A) \cup (A \cap B) \Rightarrow P(B) = P(B - A) + P(A \cap B)$

Restando las dos igualdades anteriores, se tiene

$P(A \cup B) - P(B) = P(A) - P(A \cap B)$, de donde se concluye que:

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

**Postulado de equiprobabilidad. Regla de Laplace**

Cuando un experimento aleatorio tiene n sucesos elementales y no existe ninguna razón que favorezca la realización de uno respecto de los otros, debe admitirse que todos tienen la misma probabilidad y se llaman equiprobables.

Si llamamos x a la probabilidad de uno cualquiera de ellos se debe cumplir, dado que $\Omega = \{e_1, e_2, ..., e_n\}$, $P(\Omega) = P(e_1) + P(e_2) + ... + P(e_n)$ = x+x+ ... +x = n.x, es decir que 1=n.x, de donde $\boxed{x = 1/n}$

Ahora bien, dado un suceso cualquiera A, sabemos que está constituido por p sucesos elementales (aquellos favorables a su verificación), por tanto siguiendo el razonamiento anterior, sabemos que A = $\{e_1, e_2, ..., e_p\}$ y por tanto

$P(A) = x + x + ... + x$ = x.p, pero teniendo en cuenta que x=1/n, resulta finalmente

$$\boxed{P(A) = p/n}$$

Esta es la llamada Regla de Laplace que dice: *En un espacio muestral equiprobable, la probabilidad de un suceso es la relación entre el número de casos favorables (a la aparición del suceso) y el de casos posibles (todos los del experimento).*

Es de vital importancia resaltar que esta regla tan sólo se puede aplicar cuando los sucesos elementales son equiprobables, por lo que su utilidad es restringida.

La combinatoria es muy útil a la hora del cálculo del número de casos favorables y posibles, a la hora de aplicar esta regla en múltiples problemas.

**Probabilidad condicionada:**

A veces, el conocimiento de una información complementaria o preliminar, hace variar la probabilidad de un suceso.

Veamos un ejemplo ilustrativo de esta afirmación:

Supongamos que tenemos una urna con 100 bolas distribuidas según la siguiente tabla:

| | Rojas | Blancas | Totales |
|---|---|---|---|
| Madera | 7 | 25 | 32 |
| Cristal | 32 | 36 | 68 |
| Totales | 39 | 61 | 100 |

El experimento consiste en extraer una bola al azar. Sean los sucesos siguientes:

A= “salir roja”; B= “salir de madera”

Está claro que se trata de un espacio muestral equiprobable (no hay ninguna razón para que me salga una determinada bola con más frecuencia que las demás) y por tanto se puede aplicar la regla de Laplace, en cuyo caso

P(A) = 39/100 P(B) = 32/100 y P(A∩B) = 7/100

Consideremos ahora el suceso “salir una bola de madera de entre las rojas”, o expresado de otro modo “Si sabemos que la bola que sale es roja, que sea de madera”. Evidentemente es un hecho seguro en el enunciado del suceso que la bola es roja, por tanto tenemos una información preliminar o dicho de otro modo, el suceso está condicionado por el hecho de que la bola es roja. A este tipo de sucesos se les denomina sucesos condicionados y se escriben asi:

B/A, leyéndose B condicionado a A. El suceso A es la condición y B es el condicionado.

En nuestro caso sería “salir de madera condicionado a ser roja”.

Obsérvese que la probabilidad cambia ya que el espacio muestral se reduce a las bolas rojas, por tanto P(B/A) = 7/39.

Así pues en general, podemos afirmar lo siguiente:

Dado que $P(A) = \dfrac{P(A \cap \Omega)}{P(\Omega)}$, en el caso de B/A, el espacio muestral pasa de ser Ω a ser A, por tanto podemos afirmar que:

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

que es la expresión que determina el valor de la probabilidad condicionada.

De esta expresión se deduce que $P(A \cap B) = P(A).P(B/A)$

**Sucesos independientes**

Dos sucesos A y B son independientes cuando la realización de uno no influye en la realización del otro, esto es que uno no condiciona al otro a la hora de verificarse, por tanto podemos afirmar que en este caso A/B = A o B/A = B

Según esto, si A y B son independientes se tiene que $\boxed{P(A \cap B) = P(A).P(B)}$ que es la expresión que los caracteriza.

**Teorema de la probabilidad total y Teorema de Bayes.**

Llamamos sistema completo de sucesos a un conjunto de sucesos incompatibles entre sí $\{A_1, A_2, A_3, ..., A_k\}$, de forma que $\Omega = A_1 \cup A_2 \cup ... \cup A_K$.

Consideremos un suceso X del experimento y tratemos de averiguar su probabilidad, suponiendo conocidas las de los $A_i$ y las de $X/A_i$

Obviamente X = $(X \cap A_1) \cup (X \cap A_2) \cup ... \cup (X \cap A_K)$, siendo las expresiones entre paréntesis sucesos incompatibles, por tanto:

$P(X) = \sum_{i=1}^{k} P(X \cap A_i)$, ahora bien, sabemos que $P(X \cap A_i) = P(A_i).P(X/A_i)$

Asi pues, $\boxed{P(X) = P(A_1).P(X/A_1) + P(A_2).P(X/A_2) + ... + P(A_k).P(X/A_k)}$

llamado Teorema de probabilidad total.

Supongamos ahora que lo que queremos calcular son las probabilidades de los sucesos $A_i/X$.

En este caso sabemos que $P(A_i / X) = \dfrac{P(A_i \cap X)}{P(X)} = \dfrac{P(A_i).P(X / A_i)}{P(X)}$

y aplicando el teorema de las probabilidades totales, se obtiene:

$$\boxed{P(A_i / X) = \frac{P(A_i).P(X / A_i)}{P(A_1).P(X / A_1) + P(A_2).P(X / A_2) + ... + P(A_k).P(X / A_k)}}$$

que es el llamado Teorema de Bayes.

Es importante en los problemas de probabilidad, tratar de detectar un sistema completo de sucesos para poder aplicar estos teoremas.

**Ejercicios propuestos (BLOQUE I)**

1. Una bolsa contiene tres bolas (1 roja, 1 azul, 1 blanca). Se sacan dos bolas con reemplazo, es decir, se saca una se observa y se vuelve a meter en la bolsa, a continuación se saca la segunda bola. Representar el espacio muestral: a) En diagrama de árbol. b) listado; c) en red.

2. Calcular la probabilidad de sacar al menos una cara cuando se lanzan al mismo tiempo cuatro monedas.

3. Se lanza una moneda trucada en la que se sabe que la probabilidad de obtener cara es 2/3, y la de obtener curz es 1/3. Si sale cruz se escoge al azar un número del 1 al 9; si sale cruz se escoge un número al azar del 1 al 5. Calcular la probabilidad de que se escoja un número par.

4. ¿Cuál es la probabilidad de obtener dos ases al sacar dos cartas sucesivamente de una baraja de 40 cartas?

5. Calcular la probabilidad al lanzar un dado blanco y un dado negro, de que el dado blanco salga con un número menor que tres o que la suma de ambos dados sea mayor que 9.

6. Se lanza una moneda. Si sale cara se saca una bola de una urna que tiene cuatro bolas azules, tres rojas y tres verdes. Si sale cruz se saca una bola de otra urna que tiene cinco bolas azules, dos rojas y tres verdes. Calcular la probabilidad de sacar una bola roja.

7. Tres ciclistas a,b, y c intervienen sólo ellos en una carrera. El ciclista a tiene doble probabilidad de ganar que b y b doble probabilidad de ganar que c. ¿Cuáles son las respectivas probabilidades de ganar a,b y c?

8. Se tiene un grupo de doce tornillos de los cuales cuatro son defectuosos. Se cogen dos tornillos al azar. Calcular:

   a) Probabilidad de que los dos sean defectuosos
   b) Probabilidad de que ninguno de los dos sea defectuoso
   c) Probabilidad de que por lo menos de los dos uno sea defectuoso

9. Se ha cargado un dado de tal forma que la probabilidad de salir un número cuando se lanza es proporcional al número que sale. Calcular:
   a) La probabilidad de que salga número par o primo
   b) La probabilidad de que salga número impar primo.
   c) La probabilidad de que salga número par pero no primo.

10. Calcular la probabilidad de un suceso conociendo que la suma de su cuadrado más la del cuadrado del suceso contrario es igual a 5/9

11. En un conjunto de 10000 reclutas se ha comprobado que la proporcioón de los cuatro grupos sanguíneos O, A, B, AB es:
    0 = 45%   A=40%   B=10%   AB=5%
    Cuál es la probabilidad de que eligiendo dos de ellos aleatoriamente sean:
    a) Ambos del tipo A
    b) Ninguno del tipo A
    c) Uno del tipo O y otro del tipo A
    d) Que sean de tipos diferentes

12. En un grupo de alumnos, el 25% suspendieron las Matemáticas, el 15% la Química y el 10% ambas. Se pide:

    a) Si un alumno suspendió la Química, ¿cuál es la probabilidad de que suspenda las Matemáticas
    b) ¿Cuál es la probabilidad de que suspendiera Matemáticas o Química?

13. Sabemos que del suceso A, P(A)=3/8, y del suceso B, P(B)=5/8, y P(AUB)=3/4. Calcular: a) P(A/B) b) P(B/A)

14. En un juego de dados se apuesta por el 4. Se tira el dado y antes de ver el resultado, nos dicen que ha salido par. Hallar la probabilidad de ganar.

15. Una urna contiene seis bolas blancas y ocho negras.

    a) ¿Cuál es la probabilidad de que al extraer una bola sea blanca?
    b) Se extraen dos bolas simultáneamente. ¿Cuál es la probabilidad de que las dos sean blancas?
    c) ¿Varía el resultado si se extrae primero una y luego la otra:

        1. Sin devolver la primera después de su extracción
        2. Devolviendo la primera después de su extracción?
    d) Se extraen siete bolas, ¿Cuál es la probabilidad de que sean exactamente cuatro blancas si no se devuelve la bola después de cada extracción?
    e) Se extraen cinco bolas. ¿Cuál es la probabilidad de que haya al menos una blanca.?

16. Se lanzan tres dados. ¿Cuál es la probabilidad de que salgan de forma que sumen 10?

17. Se lanza una moneda al aire dos veces consecutivas. ¿Cuál es la probabilidad de obtener al menos una cara?. Se se lanza 6 veces consecutivas. ¿Cuál es la probabilidad de obtener al menos una cara?

18. Se lanzan al aire 8 monedas, ¿cuál es la probabilidad de que salgan o todas caras o todas cruces?

19. Se tienen tres urnas conteniendo: la primera 50 bolas rojas y 50 bolas blancas, la segunda 60 amarillas y 40 blancas, la tercera 70 verdes y 30 blancas. Al sacar a la vez una de cada urna, ¿cuál es la probabilidad de que ninguna sea blanca?

20. Cuál es la probabilidad de que al tirar cinco monedas al aire queden mayoría caras.

21. ¿Cuál es la probabilidad de torpedear un barco, sabiendo que sólo pueden lanzarse tres torpedos y que la probabilidad de hacer blanco con un torpedo es 0,2?

22. Se escriben cinco cartas y sus cinco sobres correspondientes introduciéndose luego al azar las cartas en los sobres. Calcular la probabilidad de que cada carta se haya introducido en el sobre que le corresponde.

23. Dados los sucesos A y B de los que se conoce:
$P(A) = 0,4$; $P(B)=0,3$ y $P(A\cap B) = 0,1$. Calcular las siguientes probabilidades: a) $P(\bar{A}/B)$. b) $P(A/A\cup B)$ c) $P(A/A\cap B)$ d) $P(\bar{A}/B)$

24. En un grupo de un colegio han suspendido las Matemáticas el 60% de los niños, la Física el 50%, y ambas asignaturos el 20%. Calcular la probabilidad de que elegido un niño al azar, haya suspendido las Matemáticas, la Física o ambas.

25. Siendo A, B, y C, tres sucesos independientes de los que se conocen sus probabilidades: 0,2 ; 0,8 y 0,7 respectivamente. Calcular:
a) $P(A\cup B)$ b) $P(A\cup C)$ c) $P(A\cup B\cup C)$

26. La probabilidad de que en los hogares de una población tengan lavavajillas es 0,4, y de que tengan vídeo es 0,3. Calcular las siguientes probabilidades:

    a) Que tengan lavavajillas y video
    b) Que tengan lavavajillas o tengan video
    c) Que en tres hogares elegidos al azar haya lavavajillas
    d) Que en dos hogares haya dos lavavajillas o dos videos.

27. En un clínica se atienden sólo cuatro tipos de enfermedades: A, B, C, D.
La probabilidad de que un enfermo ingrese con cada una de las cuatros es:
$P(A)=0,2$ $P(B)=0,05$ $P(C)=0,6$ $P(D)=0,15$
Las probabilidades de curación de cada una de ellas es:
$P(CA)=0,8$ $P(CB)=0,75$ $P(CC)=0,2$ $P(CD)=0,3$
Calcular la probabilidad de curación de un enfermo que ingresa en la clínica sin conocer cual es su enfermedad.

28. En un estudio realizado sobre accidentes de automóviles se ha comprobado que el 10% se debe a fallos mecánicos, el 60% a fallos humanos y el 30% a defectos en las carreteras.
Si designamos por A, B, C los sucesos "tener un accidente por fallo mecánico", "tener un accidente por fallo humano" y "tener un accidente por fallo en la carretera". Calcular la probabilidad de que:
    a) Un conductor tenga un viaje con accidente
    b) Un conductor tenga un accidente debido a un fallo mecánico.

    Supuestos conocidos las probabilidades del suceso V (tener viaje con acciente), condicionado por A, B y C: $P(V/A)=0,2$ ; $P(V/B)=0,32$; $P(V/C)=0,16$

29. Dos máquinas M1 y M2 han fabricado respectivamente 100 y 200 pìezas. Se ha comprobado que M1 produce un 5% de piezas defectuosas y M2 un 6%. Tomando una de las piezas fabricadas, se pide calcular:
    a) Probabilidad de que sea defectuosa
    b) Sabiendo que es defectuosa, probabilidad de que proceda de la máquina M1

30. Se tienen tres urnas con el siguiente contenido:
    La A tiene tres bolas rojas y cinco blancas, la B dos bolas rojas y una blanca, la C dos bolas rojas y tres blancas. Se elige una urna al azar y se saca una bola. Se pide, calcular si la bola es roja la probabilidad de que proceda de la urna A.

31. Un avión realiza diariamente el mismo servicio. Estadísticamente se ha comprobado que la probabilidad de accidente en día sin niebla es 0,002 y en dia con niebla es 0,01. Cierto día de un mes que hubo 18 días sin niebla y 12 con niebla, se produjo accieente. Calcular la probabilidad de que el accidente haya ocurrido:
    a) En día sin niebla
    En día con niebla

32. Una bolsa contiene seis bolas numeradas con los números 1,2,3,4,5,6. Se pide:
    a) ¿Cuál es la probabilidad de que al tomar dos la suma de sus números sea par?
    b) Al tomar tres. ¿Qué probabilidad hay de que sumen 9?
    c) Al tomar primero una, devolverla a la bolsa, y luego otra, ¿cuál es la probabilidd de que se obtenga un número par y otro impar

33. Se lanzan tres monedas al aire simultáneamente
    a) ¿Cuál es la probabilidad de que las tres salgan cara?
    b) ¿Cuál es la probabilidad de que salgan dos caras y una cruz?

34. ¿Cuál es la probabilidad de obtener al menos un 6 al lanzar dos dados?

35. En la Facultad de Ciencias, en el primer curso, el 25% de los estudiantes suspendieron Matemáticas, el 15% suspendió Química y el 10% suspendió ambas. Eligiendo un estudiante del primer curso de esta Facultad al azar, se pide:

    a) Si suspendió Química, ¿cuál será la probabilidad de que suspendiera Matemáticas?
    b) Si suspendió Matemáticas, ¿cuál es la probabilidad de que suspendiera Química?
    c) ¿Cuál es la probabilidad de que suspendiera Matemáticas o Química?

36. En la provincia de Valencia la media de dias de sol en el mes de Julio es de 25 días. Calcular la probabilidad de que dos días consecutivos haga sol.

36. Considérese el experimento consistente en anotar el sexo de los tres primeros hijos de una familia numerosa cualquiera. Se pide:
    a) Describir el espacio muestral mediante un diagrama de árbol.
    b) Describir el suceso A={ primer hijo varón o tercero hembra}

37. De una baraja se separan cuatro cartas: un as, una sota, un caballo y un rey. De esas cuatro cartas, se eleign dos al azar, una tras otra y sin devolución. ¿Cuántos elem,entos tiene el espacio muestral? Descríbase el mismo mediante un diagrama en árbol.

38. Sea Ω = {a,b,c} un espacio muestral. Defínase una probabilidad $p:\Omega \rightarrow \Re$, tal que los sucesos elementales no sean equiprobables. Después hallar P{a,b}

39. Sea Ω = {a,b,c} un espacio muestral. Defínase una aplicación $p:\Omega \rightarrow \Re$ que cumpla p(a)+p(b)+p(c) = 1, pero que no origine una probabilidad.

40. Hállese la probabilidad de que en un lanzamiento de un dado perfecto la suma de las caras laterales y la cara superior valga 18.

41. Hállese la probabilidad de que en un lanzamiento de un dado perfecto la suma de las caras laterales sea 14. (Supóngase que son caras opuestas 1 y 6, 2 y 5, 3 y 4)

42. Si un dado perfecto se lanza dos veces, ¿qué probabilidad hay de que ambas veces resulte el mismo número?. ¿Y de que la segunda vez resulte un número mayor que la primera?

43. Se lanzan dos dados no trucados. Hállese la probabilidad de los siguientes sucesos:

    a) A={el producto de tantos obtenidos vale 17}
    b) B={la suma de tantos obtenidos es mayor que el producto}

44. Un dado ha sido trucado de tal modo que p(1) = p(2) = p(3) = p(4) = $x^2$, p(5) = -x, p(6) = -2x. Hállese x y la probabilidad de obtener impar en un lanzamiento.

45. Un jugador se construye un dado de modo que los números pares sean equiprobables, los impares también, pero la probabilidad de obtener par sea el doble de la de obtener impar. ¿Qué probabilidad hay de obtener un número menor que 4 en un lanzamiento?

46. Se elige al azar una ficha del dominó. Hállese la probabilidad de que la suma de puntos de la misma sea mayor que 5 y menor que 9. Hállese tambiénla probabilidad de que la ficha obtenida tenga un seis o un cinco pero no ambos.

47. Se elige al azar un entero entre 1 y 850 (inclusive). Hallar la probabilidad de que el número elegido sea:

    a) Múltiplo de 4 b) Impar o cuadrado perfecto
    c) No múltiplo de 5 d) Acabado en 1

48. Hállese la probabilidad de que al elegir un número de 6 cifras resulte un capicúa.

49. Se elige al azar un número natural menor que 100. Hállese la probabilidd de que el resto obtenido al:

a) Dividir por 10 valga 9
b) Dividir por 7 valga 7
c) Dividir por 8 valga 5

50. En una urna hay 50 bolas entre blancas, verdes, y negras. Suponiendo que son indistinguibles al tacto, ¿cuántas hay de cada color si sabemos que la probabilidad de sacar una verde es 2/5 y la de extraer negra es 1/10?

51. En una urna hay 50 bolas entre blancas, verdes, y negras. Suponiendo que son indistinguibles al tacto, ¿cuántas hay de cada color si sabemos que la probabilidad de sacar una blanca es 2/5 y la de extraer negra es el doble de la de obtener verde.

52. Un coche lleva una alarma que se desconecta pulsando una cierta secuencia de cuatro cifras distintas. ¿Qué probabilidad hay de conseguir desconectarla pulsando 9 cifras distintas?

**Ejercicios propuestos (Bloque II)**

1. Se considera el experimento aleatorio de lanzar un dado al aire y anotar el número de la cara superior. Hallar:
    a) El espacio muestral.
    b) El suceso A=”obtener número par”
    c) El suceso B=”obtener número primo”
    d) El suceso C=”obtener número múltiplo de 3”
    e) La unión e intersección de cada dos de los sucesos de los apartados anteriores
    f) Los sucesos: $\overline{A}, \overline{B}, \overline{C}$

2. Se lanzan al aire dos dados distintos. Determinar:
    a) El espacio muestral
    b) El suceso A=”los números de las caras suman 7”
    c) El suceso B=”el producto de los números de las caras es 12”
    d) Los sucesos $A \cup B \quad y \quad A \cap B$

3. De los individuos que componen una muestra se conocen los siguientes datos: 50 son mujeres, de ellas 30 son rubias; hay 10 varones rubios y 60 individuos morenos. ¿De cuántos individuos se compone la muestra? ¿Cuántos de los individuos son varones?

4. En una cafetería contamos los refrescos de naranja y limón que se venden de dos marcas distintas A y B. Sabemos que el 80% de los refrescos que venden son de naranja y que de ellos el 35% lo fabrica la marca B. Si de la marca A son el 70% de todos los refrescos que se venden calcula el porcentaje de refrescos de limón que se vende de la marca A y de la marca B.

5. En un avión viajan 240 personas. De ellas 110 hablan inglés, 66 francés, 58 alemán, 42 francés e inglés, 36 francés y alemán, 30 alemán e inglés y 14 los tres idiomas. Se pide:
    a) ¿cuántos no hablan ninguno de los 3 idiomas?
    b) ¿Cuántos sólo hablan alemán e inglés?
    c) ¿Cuántos hablan sólo francés?.

6. En una escuela de estudios empresariales los alumnos de 2º curso que suspenden las tres asignaturas, Matemáticas, Contabilidad y Estadística, repiten curso. El último año los resultados fueron: 6% aprobaron las 3 asignaturas; 22% aprobaron Matemáticas y Contabilidad; 16% aprobaron Matemáticas y Estadística; 28% aprobaron Contabilidad y Estadística; 37% aprobaron Matemáticas; 56% aprobaron Contabilidad y el 41% aprobaron Estadística.
    a) ¿Qué porcentaje de alumnos repitió curso?
    b) ¿Qué porcentaje aprobó solo una asignatura?

7. Dados dos sucesos A y B incompatibles que verifican P(A)=0,3 y P(B)=0,12. Calcular:

$P(\overline{A}), P(\overline{B}), P(A \cup B)$ $y$ $P(\overline{A \cup B})$.

8. Sean A y B dos sucesos tales que: P(A)=0'6, P(B)=0'4 y P( $A \cap B$ )=0'2. Calcular: $P(A \cup B)$, $P(\overline{A})$, $P(\overline{B})$, $P(\overline{A \cup B})$, $P(B - A)$

9. Dados dos sucesos que verifican $P(A \cup B) = 3/4$, $P(\overline{A}) = 2/3$, $P(A \cap B) = 1/4$. Calcular: P(A), P(B), $P(\overline{A} \cap B)$, $P(A \cap \overline{B})$

10. En una carrera participan los caballos A, B, C y D. Se estima que la probabilidad de que gane A es el doble de cada una de las probabilidades de los otros caballos. Calcular la probabilidad de ganar de cada uno de los caballos.

11. En una bolsa hay bolas negras y blancas. La probabilidad de extraer bola blanca es de dos quintos de la probabilidad de sacar bola negra. Determinar la probabilidad de extraer bola negra y la probabilidad de sacar bola blanca.

12. En un dado trucado, cuyas caras están numeradas del 1 al 6, la probabilidad de que salga cada cara es directamente proporcional al número que aparece en la misma. Hallar la probabilidad que tiene cada cara de salir.

13. Se lanzan 3 monedas al aire. Hallar:
    a) El espacio muestral.
    b) La probabilidad de cada uno de los sucesos elementales
    c) La probabilidad de al menos una cara.

14. Se extrae al azar una carta de una baraja de 40 naipes. Calcular la probabilidad de que la carta extraída sea: a) Copa, b) As, c) Figura, d) El 3 de oros.

15. Se lanzan al aire dos dados. ¿Cuál es la probabilidad de que la suma de los puntos sea 7?

16. Queremos marcar un número de teléfono de 7 cifras y sólo sabemos las 5 primeras:
    a) Calcular la probabilidad de acertar con el número que buscamos.
    b) Calcular la probabilidad de acertar con el número, sabiendo que las dos cifras desconocidas son distintas.

17. Una clase consta de 10 hombres y 20 mujeres; la mitad de los hombres y la mitad de las mujeres tienen los ojos castaños. Determinar la probabilidad de que una persona escogida al azar sea un hombre o tenga los ojos castaños.

18. En una ciudad se editan 3 periódicos A, B, C con la siguiente distribución de lectores: de cada 100 habitantes 30 leen A, 28 leen B, 17 leen C, 15 leen A y B, 9 leen A y C, 11 leen B y C, y 6 leen los tres. Se elige una persona al azar, calcular:
    a) La probabilidad de que lea algún periódico.
    b) La probabilidad de que lea exactamente un periódico.
    c) Probabilidad de que lea B y C pero no A.

19. En un centro escolar, los alumnos de COU pueden optar por cursar, como lengua extranjera, entre inglés o francés. En un determinado curso, el 90% estudia inglés y

el resto francés. El 30% de los que estudian inglés son varones y de los que estudian francés son chicos el 40%. Elegido un alumno al azar. ¿Cuál es la probabilidad de que sea chica?

20. Sean A y B dos sucesos con P(A)=1/2, P(B)=1/3 y $P(A \cap B)$=1/4, calcular: P(A/B), P(B/A), P(AUB)

21. Sabiendo que P(A) = 0'3 , $P(\overline{B}) = 0'6$, $P(A/B) = 0{,}32$, calcular:
P(A∩B), P(AUB), P( $(A/\overline{B})$ $P(B/A)$ $P(\overline{AUB})$, $P(\overline{A \cap B})$

22. a) Sabiendo que A $\subset$B calcular P(B/A) y P(A/B)
    b) Si A y B son incompatibles, calcular P(B/A)
23. En una empresa hay 45 empleados, de los cuales 29 son hombres y 16 mujeres. De ellos, 7 hombres y 5 mujeres son fumadores. Calcula las siguientes probabilidades: P(H) , P(M), P(H∩F), P(M∩F) , P(F), P(H/F) 1 P(M/F)

24. El 20% de los empleados de una empresa son Ingenieros y otros 20% son Economistas. El 75% de los Ingenieros ocupan un puesto directivo, y el 50% de los Economistas también, mientras que de los No-Ingenieros y No-Economistas solamente el 20% ocupan un puesto Directivo. ¿Cuál es la probabilidad de que un empleado Directivo elegido al azar sea Ingeniero?

25. Una urna contiene 25 bolas blancas sin marcar, 75 bolas blancas marcas, 125 bolas negras sin marcar y 175 bolas negras marcadas. Se extrae al azar una bola. Calcula:
    a) La probabilidad de que sea blanca.
    b) Si la bola extraída está marcada, ¿Cuál es la probabilidad de que sea blanca?

26. Se sortea un viaje a Canarias entre los 200 clientes de una tienda de electrodomésticos. De ellos 125 son mujeres, 155 están casados y 95 son mujeres casadas. ¿cuál es la probabilidad de que le toque el viaje a un hombre soltero?. Si del afortunado sabemos que está casado ¿cuál es la probabilidad de que sea mujer?

27. En una clase el 40% de los alumnos aprueba Matemáticas y el 50% aprueba filosofía. Se sabe que la probabilidad de aprobar matemáticas si se aprobó filosofía es 0'6.
    a) Estudiar si los sucesos aprobar Matemáticas y aprobar Filosofía son independientes.
    b) ¿Qué porcentaje de alumnos aprobaron las dos asignaturas?
    c) De los alumnos que aprobaron Matemáticas, ¿qué porcentaje aprobó Filosofía?

28. En un Instituto el 25% de los alumnos han suspendido Matemáticas, el 15% han suspendido Física y el 10% han suspendido las 2 asignaturas. Se selecciona un individuo al azar. Calcular:
    a) La probabilidad de que haya suspendido matemáticas.
    b) Si ha suspendido Matemáticas, cuál es la probabilidad de que haya suspendido Física
    c) ¿Cuál es la probabilidad de que haya suspendido Matemáticas ó Física.

d) Si ha aprobado Matemáticas, ¿cuál es la probabilidad de que haya suspendido la Física?

29. Un juego consiste en lanzar tres monedas al aire. Si las tres monedas aparecen de igual modo (tres caras o tres cruces), se gana. En caso contrario, se vuelve a tirar. Se pide:
    a) ¿Cuál es la probabilidad de ganar en la primera tirada?
    b) ¿Cuál, la de perder las dos primeras y ganar la tercera?

30. Una urna contiene 8 bolas rojas, 3 blancas y 9 azules. Si se extraen 2 bolas aleatoriamente sin reemplazamiento. Halla la probabilidad de que:
    a) Las 2 bolas sean rojas.
    b) Las bolas sean extraídas en este orden: roja, azul
    c) Una bola sea roja y una sea blanca.
    d) Al menos una bola sea blanca.

31. En una caja hay 15 bombillas de las cuales 5 están fundidas. Si cogemos 3 de ellas al azar, cual es la probabilidad de que:
    a) Ninguna esté fundida
    b) Exactamente una esté fundida.
    c) Por lo menos una esté fundida.

32. Las probabilidades de acertarle a un blanco de tres tiradores, A, B y C son respectivamente, 1/6, ¼ y 1/3. Si cada uno de ellos dispara una sola vez al blanco, calcular:
    a) La probabilidad de que uno exactamente acierte en el blanco
    b) Si sólo uno acierta en el blanco, cual es la probabilidad de que sea A
    c) Hallar la probabilidad de que alguno acierte en el blanco.

33. La Compañía "Minisegadoras" fabrica los motores, las hojas y las cubiertas de sus productos. El porcentaje de los motores defectuosos es del 5%, el de hojas defectuosas es el 1% y el de cubiertas el 3%. ¿Cuál es la probabilidad de que una segadora montada no tenga defectos?
34. La probabilidad de que un torpedo hunda un barco es 0'2. Un submarino dispara 3 torpedos, ¿Cuál es la probabilidad de que hunda a un barco?

35. En una bolsa hay 12 bolas blancas y 20 verdes. Al sacar 4 bolas sucesivamente calcular la probabilidad de que las 4 sean blancas. (Con y sin devolución)

36. Probabilidad de que al sacar sucesivamente sin devolución 5 cartas de una baraja las 5 sean del mismo palo.

37. Urna I: Contiene 6 bolas rojas y 4 bolas blancas. Urna II: contiene 4 bolas rojas y 8 bolas blancas. Se lanza un dado. Si aparece un número menor que 3, nos vamos a la urna I; si el resultado es 3 o más nos vamos a la urna II. A continuación extraemos una bola. Se pide:
    a) Probabilidad de que la bola sea roja y de la urna II
    b) Probabilidad de que la bola sea blanca.

38. Tres cofres idénticos contienen: El primero, 3 lingotes de oro y 2 de plata; el segundo, 2 de oro y 5 de plata; y el tercero, 6 de oro y 7 de plata. ¿Cuál es la probabilidad de que al extraer un lingote al azar de un cofre sea de plata?

39. En una casa hay dos tarros que contienen caramelos. En el primer tarro hay 8 caramelos de naranja y 12 de limón. En el segundo tarro hay 15 caramelos de naranja y 5 de limón. Un niño que viene de visita elige uno de los tarros y en él un caramelo. Si al comerlo nota que es de naranja, ¿qué probabilidad tiene de haber elegido el segundo tarro?

40. Tres maquinas A, B y C fabrican tornillos del mismo tipo. Los porcentajes de defectuosos en cada máquina son respectivamente 1%, 2%, 3%. Se mezclan 120 tornillos: 20 de la máquina A, 40 de la B y 60 de la C. Elegido uno al azar resulta defectuoso. ¿Cuál es la probabilidad de que haya sido fabricado por la máquina B?

41. Dos medicamentos A y B son eficaces para tratar una enfermedad. El medicamento A produce mejoría en el 74% de los caos y el B en el 80% de los casos. En una clínica tienen 3 tubos del medicamento A y 2 del medicamento B.Eelegimos un tubo al azar y le damos de él una pastilla al enfermo.
    a) Calcula la probabilidad de que el enfermo tenga mejoría en la enfermedad.
    b) Si sabemos que el enfermo mejora, calcula la probabilidad de que se le suministrase el medicamento B.

42. Se lanza una moneda hasta que el resultado sea cara. Halla la probabilidad de que esto suceda: **a)** en el primer lanzamiento. **b)** en el segundo lanzamiento. **c)** En el lanzamiento n-ésimo.

43. Sean A y B dos sucesos independientes de un experimento aleatorio, tales que la probabilidd de que ocurran simultáneamente ers 1/3 y la de que no ocurra ninguno de los dos es 1/6. halla P(A) y P(B).

44. Suponiendo que todos los meses del año son de 30 días, hallar la probabilidad de que los cumpleaños de tres hermanos sean:
    a) El mismo día del año ; b) Los tres en días distintos c) Los tres en el mismo mes
    d) Cada uno en un mes distintos e) Los tres en marzo f) Ninguno en mayo.

45. Un banco partiendo de la información sobre el comportamiento de sus clientes referida a los errores cometidos al cubrir cheques obtiene las siguientes conclusiones: - De 850 clientes con fondos, 25 cometieron algún error. El 98% de los clientes tiene fondos. De 50 cheques sin fondos, 45 tenían algún error. Calcular la probabilidad de que un cheque con algún error no tenga fondos.

46. En un lote de 40 pastillas de jabón hay 5 premiadas. Compramos 3 pastillas. Calcular la probabilidad de que: a) Las tres tengan premio; b) Exactamente 2 tengan premio; c) Alguna tenga premio.

47. Ante un examen, un alumno sólo ha estudiado 15 de los 25 temas correspondientes a la materia del mismo. Este se realiza extrayendo al azar tres temas y dejando que

el alumno escoja uno de ellos para ser examinado del mismo. Halla la probabilidad de que el alumno pueda elegir en el examen uno de los temas estudiados.

48. Una clase tiene 6 niñas y 10 niños. Si se escoge un comité de 3 al azar, hallar la probabilidad de:
    a) Seleccionar 3 niños.
    b) Seleccionar exactamente dos niños y una niña.
    c) Seleccionar por lo menos un niño.
    d) Seleccionar exactamente dos niñas y un niño.

49. En una bolsa hay 4 bolas verdes y 8 bolas rojas. Se saca una bola de la bolsa y se devuelve acompañada de otra del mismo color. Se saca entonces una segunda bola. Calcular:
    a) La probabilidad de que la segunda bola sea verde.
    b) Probabilidad de que las dos bolas sean rojas.
    c) Probabilidad de que sean la primera roja y la segunda verde.

50. Un joyero compra los relojes a dos casas proveedoras. La primera le sirve el 60% de los relojes, de los cuales el 0,4% son defectuosos. La segunda le proporciona el resto, siendo defectuosos el 1,5%. Un día el joyero, al vender un reloj, observa que este no funciona. Hallar la probabilidad de que el reloj proceda de la primera casa proveedora.

51. Dos personas comparten el mismo número de teléfono. De las llamadas que llegan 2/5 son para A y 3/5 para B. Sus ocupaciones los tienen alejados del teléfono de modo que A está fuera el 50% del tiempo y B el 25% del tiempo. Calcula la probabilidad de que al recibir una llamada no haya nadie para coger el teléfono y la probabilidad de que al recibir una llamada esté presente la persona a la que llaman.

52. Un 70% de los clientes de una compañía de seguros de automóviles tienen más de 25 años. Un 5% de los clientes de ese grupo tienen un accidente a lo largo del año. En el caso de clientes menores de 25 años este porcentaje es del 20%.
    a) Si elegimos un asegurado al azar, calcular la probabilidad de que tenga un accidente ese año.
    b) Si una persona tuvo un accidente, calcular la probabilidad de que sea menor de 25 años.

53. Los expertos afirman que la probabilidad de que la bolsa suba es 0’4. Por otra parte la probabilidad de que el dólar se mantenga estable es 0’5 y la probabilidad de que la bolsa no suba cuando el dólar permanece estable es 0’9. Calcula:
    a) La probabilidad de que la bolsa suba si el dólar permanece estable.
    b) La probabilidad de que el dólar se mantenga estable o suba la bolsa.
    c) La probabilidad de que el dólar se mantenga estable si sube la bolsa.

54. Consideramos tres dados de los cuales dos son correctos y uno está trucado de forma que el 6 aparece en la mitad de las tiradas y las otras caras aparecen con la misma probabilidad. Se elige un dado al azar y se lanza: a) ¿Cuál es la probabilidad de obtener un cinco? b) Si ha salido un seis, ¿cuál es la probabilidad de haber elegido el dado trucado?

55. Un coche frena bruscamente y provoca un accidente. Tres testigos estaban presentes: A, B y C. la probabilidad de que A haya apreciado la brusquedad de la frenada es 90%, y las correspondientes a B y C son 85% y 80%. Supuesto que los testimonios que se presten sean independientes unos de otros, ¿qué probabilidad hay de que los tres testimonien que la frenada ha sido brusca? ¿Qué probabilidad hay de que lo testimonien al menos dos de los testigos?.

56. En una fábrica de autocares se descubrió que 1 de cada 100 tenía problemas con el cierre de la puerta. Como medida de precaución, antes de la venta, a cada autocar se le hace un test de verificación y se obsequia a los compradores con un cinturón multiusos para una reparación de emergencia. El test no es totalmente fiable, pues, si el coche tiene problemas con la puerta se lo detecta en un 95% de los casos, mientras que si no lo tiene, en un 2% de las veces indica que si.
    a) ¿Cuál es la probabilidad de que un autocar tenga problemas con la puerta y no lo detecte el test.
    b) Si el test indica problemas en la puerta, ¿cuál es la probabilidad de que no lo tenga?

57. Una secretaria escribe 5 cartas diferentes a 5 personas y mete cada carta en un sobre sin fijarse. ¿Cuál es la probabilidad de que todas las personas reciban la suya?

58. Un test que detecta la presencia de cierto tipo T de bacterias en el agua da positivo con una probabilidad de 0’9 en caso de que las haya. Si no las hay, la probabilidad de que sea positiva es de 0,2. Se dispone de 100 muestras de las que 25 tienen bacterias de tipo T. Si elegimos una muestra al azar:
    a) ¿Cuál es la probabilidad de que la muestra tenga bacterias de tipo T el test sea positivo?
    b) ¿Cuál es la probabilidad de que la muestra no tenga bacterias tipo T y el test sea positivo?
    c) ¿Cuál es la probabilidad de que el test dé negativo si la muestra tiene bacterias de tipo T?

59. Una caja A contiene 9 cartas numeradas del 1 al 9 y otra caja B contiene 5 cartas numeradas del 1 al 5. Se elige una caja al azar y se toma una carta, si está numerada con un número par se toma otra carta de la misma caja, y si está numerada con un número impar se toma de la otra caja.
    a) ¿Calcula la probabilidad de que ambas cartas estén numeradas con números impares.
    b) Si ambas cartas tienen números pares, calcula la probabilidad de que sean de la caja A.

60. En una bolsa hay 2 bolas blancas y 3 negras, y en otra bolsa hay 4 blancas y 1 negra. Elegimos al azar una bolsa y en ella una bola y resulta que es negra. A continuación vamos a la otra bolsa y elegimos una bola, ¿cuál es la probabilidad de que también sea negra?

# TEMA VI. CADENAS DE MARKOV

## Definición

Una cadena de Markov es un experimento que se realiza en un número infinito de etapas, bajo las siguientes condiciones:

1º) En cada etapa solo pueden producirse un número finito de sucesos que denominaremos **estados** y representaremos por $\{e_1, e_2, ... ,e_k\}$

2º) Si llamamos $R_n$ al resultado del experimento en la etapa n ($R_n$ puede ser cualquiera de los k estados), se verifica:

$$P(R_{n+1}=e_i / R_n, R_{n-1}, \dots , R_2, R_1) = P(R_{n+1}=e_i / R_n)$$

Esto quiere decir que la probabilidad de que en una etapa se produzca como resultado un estado determinado, sabiendo el resultado de las etapas anteriores, solamente depende del resultado de la etapa inmediatamente anterior. Esta propiedad se denomina propiedad de Markov o de *carencia histórica*.

Si la probabilidad de una etapa, no depende de la anterior, la cadena de Markov se dice que es homogénea.

Ejemplo: En un pueblo, al 90% de los días soleados le siguen días soleados, y al 80% de los días nublados le siguen días nublados. Con esta información modelar el clima del pueblo como una cadena de Markov.

Cada etapa es un día. Los estados son dos {N, S}, siendo N =”día nublado” y S=”día soleado”.

Esta claro, según el enunciado anterior que $P(R_{n+1}=S / R_n, R_{n-1},...,R_1)$, esto es que el tiempo que haga mañana (etapa n+1) va a depender solamente del que haga hoy (etapa n) y para nada necesitamos conocer lo que ocurrió antes de ayer.

Sabemos que $P(R_{n+1}=S / R_n=S)=0,9$ ; $P(R_{n+1}=S / R_n=N)=0,2$;

$P(R_{n+1}=N / R_n=S)=0,1$ y $P(R_{n+1}=N / R_n=N)=0,8$.

Estamos ante una cadena de Markov, pues el experimento se puede realizar en infinitas etapas (días), ocurriendo en cada etapa dos (nº finito) sucesos S y N con probabilidades que solo dependen de lo ocurrido en la etapa anterior.

## Probabilidades de transición

Según lo visto anteriormente, nos van a interesar las probabilidades de que ocurra cierto estado en la etapa n+1, sabiendo lo ocurrido en la etapa n, por tanto estamos hablando de la probabilidad siguiente:

$$P(R_{n+1}=e_j / R_n=e_i)$$

Es decir, la probabilidad de que en la etapa n+1 ocurra $e_j$, sabiendo que en la etapa anterior n ha ocurrido $e_i$. A estas probabilidades, cuando $e_j$ y $e_i$, toman todos los valores posibles $\{e_1, e_2, ... ,e_k\}$, se les denomina probabilidades de transición, y se representan por $p_{ij}$. (Obsérvese que el primer subíndice corresponde al estado n y el segundo al estado n+1). Hay tantas probabilidades de transición como parejas ordenadas de estados podamos formar, es decir kxk.

Es cómodo representar las probabilidades de transición en forma de matriz cuadrada, donde las filas indican los estados en la etapa n y las columnas los estados en la etapa n+1.

En nuestro ejemplo anterior del clima, la matriz de transición sería:

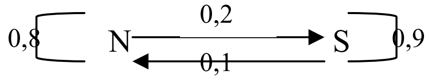
$P = \begin{pmatrix} 0,9 & 0,1 \\ 0,2 & 0,8 \end{pmatrix}$, donde la primera fila indica que en la etapa n hacía sol y la segunda indica que hacía nublado. Por el contrario la primera columna indica que en la etapa n+1 hacía sol y la segunda columna indica que hacía nublado.

Obsérvese que la suma de todos los elementos de una fila, suma 1. Esto va a ocurrir en todas las matrices de transición de una cadena de Markov.

**Grafos asociados a una cadena de Markov**

Otra forma de representar los estados y las probabilidades de transición, suele realizarse mediante los llamados grafos, que consisten en unos nodos que son los estados y unas flechas entre éstos indicando sobre ellas las probabilidades de transición. Todas las flechas que salen de un nodo han de sumar 1.

El grafo de nuestro ejemplo sería:




Otra forma de representación gráfica de una cadena de Markov es el conocido árbol, pero solamente es operativo de la etapa 0 (inicial) a la etapa 1, y se complica cuanto mayor sea el número de estados.

**Transiciones a más de una etapa:**

Hasta ahora podemos conocer la probabilidad de un estado en una etapa conociendo el anterior, pero ¿podemos averiguar la probabilidad de los estados en la etapa n+r, a partir de la etapa n?.

Veámoslo en el caso más sencillo, para r=2

$$P(R_{n+2}=e_j \,/\, R_n=e_i)$$

Si planteamos el árbol de la situación, veremos que:

$P(R_{n+2}=e_j \,/\, R_n=e_i)=P(R_{n+2}=e_j \,/\, R_{n-1}=e_1).\ P(R_{n-1}=e_1 \,/\, R_n=e_i)+ P(R_{n+2}=e_j \,/\, R_{n-1}=e_2) .P(R_{n-1}=e_2 \,/\, R_n=e_i)+...+ P(R_{n+2}=e_j \,/\, R_{n-1}=e_k).\ P(R_{n-1}=e_k \,/\, R_n=e_i)= p_{1j}.p_{i1}+p_{2j}.p_{i2}+\ldots+p_{kj}.p_{ik}$

Este es el valor que resultaría en la posición ij de la matriz P.P.

En general se tiene que las probabilidades de transición para r=2, son los coeficientes de la matriz $P^2$.

En general las probabilidades de transición de la etapa n a n+r, vendrá dado por $P^r$

En nuestro ejemplo anterior, si quisiéramos saber las probabilidades de transición al cabo de tres días, por ejemplo, haríamos $P^3$, cuyo resultado es:

$P^3 = \begin{pmatrix} 0,781 & 0,219 \\ 0,438 & 0,562 \end{pmatrix}$ ¿Qué quiere decir esto?. Pues, por ejemplo, que la probabilidad de que dentro de tres días haga sol, sabiendo que hoy hace nublado es 0.438, ya que se sigue el mismo criterio para filas y columnas que en P.

(Nota.- Las potencias de P o los productos de matrices los realizamos en Excel con la función MMULT., dentro de las funciones Matemáticas y trigonométricas. Recuérdese que la salida de datos de una matriz se efectúa con la combinación de teclas siguiente: ctrl.+Mayúsculas+Enter)

**Tipos de estados.-**

Dados dos estados $e_i$ y $e_j$, se dice que $e_i$ comunica con $e_j$, si existen n,m>0, tales que $p_{ij}^{(n)}>0$ y $p_{ji}^{(m)}>0$, estos es de que exista alguna probabilidad de que, partiendo de $e_i$, llegar al estado $e_j$ en cierto número de etapas y viceversa

Si en una cadena de Markov todos los estados se comunican entre sí, se llama irreducible y no se puede descomponer en subcadenas más pequeñas.

Un estado se llama absorbente cuando $p_{ii}=1$. En este caso $p_{ii}^{(n)}=1$ para todo valor de n, por lo que el estado $e_i$ no comunica con ningún otro estado, con lo cual por si mismo es una subcadena.

**Distribución estacionaria y distribución límite.**

Partimos de una etapa inicial que representaremos por 0. En ese momento los estados tienen una probabilidad determinada por el vector $q^{(0)} =(q_1^{(0)},q_2^{(0)},...,q_k^{(0)})$. llamado distribución inicial de la cadena. Estas componentes representan las probabilidades de que se verifiquen los estados $\{e_1, e_2, ... ,e_k\}$ en el instante inicial.

Gracias a las propiedades de las cadenas de Markov, se puede demostrar fácilmente que la probabilidad de los estados al cabo de n etapas es $q^{(0)}.P^n$, obteniéndose otro vector de distribución de probabilidad $q^{(n)}= q^{(0)}.P^n = (q_1^{(n)},q_2^{(n)},...,q_k^{(n)})$.

En muchas ocasiones ocurre que existe un valor de n a partir del cual se verifica que $q^{(n)} = q^{(n+1)}$ en cuyo caso la distribución de probabilidad se estaciona, es decir que siempre vale lo mismo a partir de esa etapa n. A dicha distribución se le denomina estacionaria o situación estable de la cadena, y se le representa por $q=(q_1,q_2,...,q_k)$. Para averiguar q, tenemos en cuenta que si $q^{(n)} = q^{(n+1)} =q$ , puesto que $q^{(n+1)}= q^{(n)}.P$, resulta q = q.P, de donde se obtiene un sistema de ecuaciones, que unidas, a la ecuación $\sum_{i=1}^{k} q_i = 1$, resulta un sistema compatible determinado (solución única) cuya solución determina el vector de probabilidad estacionaria q.

Se llama distribución límite, al límite de los vectores $q^{(n)}$, cuando n tiende a infinito siempre que este límite exista. Es obvio que si una cadena posee distribución estacionaria, ésta ha de coincidir con la distribución límite.

Ahora bien, si la cadena es irreducible, se puede afirmar que existe distribución estacionaria y por tanto coincidirá con la distribución límite.

**Ejercicios propuestos de Cadenas de Markov**

1º) El ascensor de un edificio con bajo y dos pisos realiza viajes de uno a otro piso. El piso en el que finaliza el viaje n-ésimo del ascensor sigue una cadena de Markov. Se sabe que la mitad de los viajes que parten del bajo se dirigen a cada uno de los otros dos pisos, mientras que si un viaje comienza en el primer piso, sólo el 25% de las veces finaliza en el segundo. Por último, si un trayecto comienza en el segundo piso, siempre finaliza en el bajo. Se pide:

a) Calcular la matriz de probabilidades de transición de la cadena

b) Dibujar el grafo asociado

c) ¿Cuál es la probabilidad de que, a largo plazo, el ascensor se encuentre en cada uno de los tres pisos.

2º) Un agente comercial realiza su trabajo en tres ciudades A, B y C. Para evitar desplazamientos innecesarios está todo el día en la misma ciudad y allí pernocta, desplazándose a otra ciudad al día siguiente, si no tiene suficiente trabajo. Despues de estar trabajando un día en C, la probabilidad de tener que seguir trabajando en ella al día siguiente es 0,4, la de tener que viajar a B es 0,4 y la de tener que ier a A es 0,2. Si el viajante duerme un día en B, con probabilidad de un 20% tendrá que seguir trabajando en la misma ciudad al día siguiente, en el 60% de los casos viajará a C, mientras que irá a A con probabilidad 0,2. Por último si el agente comercial trabaja todo un día en A, permanecerá en esa misma ciudad, al día siguiente, con una probabilidad 0,1, irá a B con una probabilidad de 0,3 y a C con una probabilidad de 0,6.

a) Si hoy el viajante está en C, ¿cuál es la probabilidad de que también tenga que trabajar en C al cabo de cuatro días?

b) ¿Cuales son los porcentajes de días en los que el agente comercial está en cada una de las tres ciudades?

3º) Suponga que toda la industria de refresco produce dos colas: Coca Cola y Pepsi Cola. Cuando una persona ha comprado Coca Cola hay una probabilidad de 90% de que siga comprándola la vez siguiente. Si una persona compró Pepsi, hay 80% de que repita la vez siguiente. Se pide:

a) Si una persona actualmente es comprador de Pepsi. ¿Cuál es la probabilidad de que compre Coca Cola pasadas dos compras a partir de hoy?

b) Si en la actualidad una persona es comprador de Coca Cola. ¿Cuál es la probabilidad de que compre Coca Cola pasadas tres compras a partir de ahora?

c) Supongamos que el 60% de toda la gente toma hoy Coca Cola y el 40% Pepsi. A tres compras a partir de ahora, ¿Qué fracción de los compradores estará tomando Coca Cola.

Determinar el estado estable.

4º) La cervecería más importante del mundo (Guiness) ha contratado a un analista de investigación de operaciones para analizar su posición en el mercado. Están preocupados en especial por su mayor competidor (Heineken). El analista piensa que el cambio de marca se puede modelar como una cadena de Markov incluyendo tres estados, los estados G y H representan a los clientes que beben cerveza producida por las mencionadas cervecerías y el estado I representa todas las demás marcas. Los datos se toman cada mes y el analista ha construido la siguiente matriz de transición de los datos históricos.

| | G | H | I |
|---|---|---|---|
| G | 0,7 | 0,2 | 0,1 |
| H | 0,2 | 0,75 | 0,05 |
| I | 0,1 | 0,1 | 0,8 |

¿Cuáles son los porcentajes de mercado en el estado estable para las dos cervecerías grandes.?

5º) En una comunidad hay 3 supermercados (S1, S2, S3) existe la movilidad de un cliente de uno a otro. El 1 de septiembre, ¼ de los clientes va al S1, 1/3 al S2 y 5/12 al S3 de un total de 10.000 personas. Cada mes el S1 retiene el 90% de sus clientes y pierde el 10% que se va al S2. Se averiguó que el S2 solo retiene el 5% y pierde el 85% que va a S1 y el resto se va a S3, el S3 retiene solo el 40%, pierde el 50% que va al S1 y el 10% va al S2.

a) Establecer la matriz de transición
b) ¿Cuál es la proporción de clientes para los supermercados el 1 de noviembre?
c) Hallar el vector de probabilidad estable.

6º) Los consumidores de café en el área de Pontevedra usan tres marcas A, B, C. En marzo de 1995 se hizo una encuesta en lo que entrevistó a las 8450 personas que compran café y los resultados fueron:

| | Compra en el siguiente mes | | | TOTALES |
|---|---|---|---|---|
| Compra actual | Marca A | Marca B | Marca C | |
| Marca A = 1690 | 507 | 845 | 338 | 1690 |
| Marca B = 3380 | 676 | 2028 | 676 | 3380 |
| Marca C = 3380 | 845 | 845 | 1690 | 3380 |
| TOTALES | 2028 | 3718 | 2704 | **8450** |

a) Si las compras se hacen mensualmente, ¿cuál será la distribución del mercado de café en Pontevedra en el mes de junio?
b) A la larga, ¿cómo se distribuirán los clientes de café?
c) En junio, cual es la proporción de clientes leales a sus marcas de café?

# TEMA VII. VARIABLE ALEATORIA

**Definición.-** Un conjunto se dice numerable cuando es finito o cuando se puede poner en correspondencia biunívoca con el conjunto de los números naturales, en cuyo caso se dirá que es infinito numerable.

En caso contrario se denomina conjunto no numerable.

Son conjuntos numerables, por ejemplo, los números enteros, los racionales. No lo son los números reales o cualquier intervalo real.

**Variable aleatoria.-** Se denomina así a una función definida sobre un espacio muestral asociado a un experimento aleatorio, que alcanza valores en R, es decir que lleva sucesos elementales en números reales.

$X : \Omega \rightarrow R$ , de forma que $X(e_i)=x_i$.

La variable aleatoria no es única y se define siempre en función de los sucesos que se desean estudiar.

Ejemplo de variable aleatoria: En el experimento lanzar una moneda, una variable aleatoria sería asignarle 0 al suceso cara y 1 al suceso cruz,

En el experimento aleatorio lanzar cinco monedas, se podría definir la siguiente variable aleatoria: X="número de caras". Es obvio que los valores posibles que podría tomar X serían 0,1,2,3,4,5. A este conjunto de valores numéricos se le denomina Recorrido de la variable aleatoria.

Si el recorrido de una variable aleatoria es finito o infinito numerable, la variable aleatoria se denomina discreta. Si por el contrario el recorrido es un intervalo de R le llamaremos variable aleatoria continua.

Un ejemplo de variable aleatoria continua es este: Lanzamos un dardo a una diana y definimos la variable aleatoria X="distancia del impacto al centro de la diana). Es obvio que el recorrido de esta variable aleatoria es el intervalo $[0,\infty]$ que constituye un conjunto no numerable.

**Función de masa de probabilidad de una variable aleatoria discreta X.-**

Se trata de una función que toma valores en el recorrido de la variable y los alcanza en R, concretamente en el intervalo [0,1].

$f : R \rightarrow [0,1]$,

de tal modo que $f(x_i) = P(e_i / X(e_i)=x_i)$, o abreviadamente $P(X=x_i)$

Propiedades de la función de masa de distribución:

a) $0 \leq f(x_i) \leq 1,\ \forall x_i \in \mathrm{Re}c(X)$

b) $\sum_i x_i = 1$

Ejemplo:

Supongamos el experimento aleatorio, arrojar cinco monedas al aire. Sobre el espacio muestral definimos la variable aleatoria X="número de caras". El recorrido de X es

{0,1,2,3,4,5}. Si f es la función de masa de probabilidad de la v.a. X, f(2), por ejemplo, toma el valor de la probabilidad del suceso elemental asociado al valor 2 mediante X, esto es la probabilidad de que salgan dos caras (y tres cruces), cuyo valor es 10/32, de este modo tendremos que:

f(0) = 1/32; f(1)=5/32; f(2)=10/32; f(3)=10/32; f(4)=5/32; f(5)=1/32.

Cualquier valor que no esté en el recorrido de X, tendrá como masa de probabilidad 0.

La función de masa de probabilidad solo es aplicable a variable aleatorias discretas, y para v.a. finitas poco numerosas, suele venir representada mediante una tabla.

Si tomamos como ejemplo, lanzar tres monedas y definimos la v.a. X="número de caras", el recorrido de X sería 0,1,2,3, obtenido así:

**Tabla 1**

| $e_i$ | ccc | cc+ | c+c | +cc | c++ | +c+ | ++c | +++ |
|---|---|---|---|---|---|---|---|---|
| $x_i$ | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 0 |

La función de mása de probabilidad será:

| $x_i$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(x_i)$ | 1/8 | 3/8 | 3/8 | 1/8 |

La función de masa de probabilidad puede representarse mediante gráficos por medio de los diagramas de barras en las variables estadísticas unidimensionales.

**Función de distribución de una variable aleatoria discreta X.-** Esta función puede definirse como la función de masa de probabilidad acumulada hasta el valor correspondiente, incluido éste. La representaremos mediante F.

Por tanto:

$F : R \to R \,/\, F(x_i) = \sum_{k=1}^{i} f(x_k)$ ; En general se define:

$F(x) = \sum_{i=1}^{p} f(x_i),$ siendo $x_p$ el mayor valor del recorrido de X, menor que x.

Propiedades de la función de distribución:

a) Es una función escalonada monótona creciente

b) F(x) = 0, si x es un valor inferior al minimo de los valores del recorrido de X

c) F(x)=1, si x es un valor superior al máximo de los valores del recorrido de X.

d) Su gráfica es escalonada.

e) Es continua en x/ f(x)=0 y discontinua en x / f(x) =/0

**Medidas características de una variable aleatoria discreta.-**

Media o esperanza matemática

$$E[X] = \sum_{i=1}^{p} x_i . f(x_i)$$

Varianza

$$\sigma^2 = \sum_{i=1}^{p} (x_i - E[X])^2 . f(x_i) = E[X^2] - (E[X])^2$$

Desviación típica

$$\sigma = \sqrt{\sum_{i=1}^{p} (x_i - E[X])^2 . f(x_i)}$$

Se pueden definir los momentos respecto al origen y con respecto a la media, de orden p, resultando sus valores iguales que en el caso de una variable estadística; basta sustituir las frecuencias relativas por los valores de la función de masa de probabilidad.

**Operaciones con variables aleatorias discretas:**

***OPERACIONES CON VARIABLES ALEATORIAS DISCRETAS***

*SUMA*

Sean X, Y dos variables aleatorias discretas definidas sobre el mismo espacio muestral Ω, de modo que:

$X : \Omega \rightarrow R$ con $X(e_i) = x_i$ ; $Y : \Omega \rightarrow R$ con $Y(e_i) = y_i$

Definimos la variable $X + Y : \Omega \rightarrow R$ de modo que $X+Y\,(e_i) = x_i + y_i$

En estas condiciones se verifica:

a) $E[X + Y] = E[X] + E[Y]$

b) $\sigma_{x+y}{}^2 = \sigma_x{}^2 + \sigma_y{}^2 + 2\sigma_{xy}$

Demostración a)

$$E[X + Y] = \sum_{i=1}^{n} (x_i + y_i) f_i = \sum_{i=1}^{n} x_i f_i + y_i f_i = \sum_{i=1}^{n} x_i f_i + \sum_{i=1}^{n} y_i f_i = E[X] + E[Y]$$

Demostración b)

$$\sigma_{x+y}{}^2 =$$

$$\sum_{i=1}^{n} \left[(x_i + y_i) - (\bar{x} + \bar{y})\right]^2 f_i = \sum_{i=1}^{n} \left[(x_i - \bar{x}) + (y_i - \bar{y})\right]^2 f_i = \sum_{i=1}^{n} (x_i - \bar{x})^2 f_i + \sum_{i=1}^{n} (y_i - \bar{y})^2 f_i +$$

$$2\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) f_i = \sigma_x{}^2 + \sigma_y{}^2 + 2\sigma_{xy}$$

Análogamente se demuestra que $E[X + Y] = E[X] - E[Y]$ y $\sigma_{x-y}{}^2 = \sigma_x{}^2 + \sigma_y{}^2 - 2\sigma_{xy}$

*PRODUCTO POR UNA CONSTANTE*

Sea X una variable aleatoria discreta $X:\Omega\rightarrow R$ con $X(e_i) = x_i$. Sea a un número real. Defino la siguiente variable aleatoria:

$a.X:\Omega\rightarrow R$ con $a.X(e_i) = ax_i$

Resulta que: a) E[a.X] = a.E[ X]
b) $\sigma_{aX}{}^2 = a^2\sigma_X{}^2$

La demostración de ambos resultados es trivial.

*SUMA DE UNA CONSTANTE*

Sea X una variable aleatoria discreta $X:\Omega\rightarrow R$ con $X(e_i) = x_i$. Sea a un número real. Defino la siguiente variable aleatoria: $.X+a:\Omega\rightarrow R$ con $X+a(e_i) = x_i+a$. Podemos considerar a como una variable aleatoria constante. Es obvio que a = a y que $\sigma_a = 0$, por lo que teniendo en cuenta las anteriores proposiciones, resulta que E[X+a] =E[ X]+a y que $\sigma_{X+a}{}^2 = \sigma_x{}^2$.

De este último resultado se deduce que si tenemos una variable aleatoria bidimensional, donde una de ellas es constante, la covarianza es 0

*VARIABLE ALEATORIA POLINÓMICA DE GRADO 1*

Sea X una variable aleatoria discreta $X:\Omega\rightarrow R$ con $X(e_i) = x_i$. Sean a y b dos números reales. Sea Y = a.X + b. Teniendo en cuenta los resultados anteriores se tiene que :

a) E[Y] = a.E[X] + b b) $\sigma_y{}^2 = a^2\sigma_x^2$

**Variable aleatoria continua. Función de distribución**

Ya vimos que la diferencia existente entre una v.a. discreta y continua venía determinada por la naturaleza del recorrido de la función. Si este recorrido era finito o infinito numerable, la v.a. era discreta. Si, por el contrario, el recorrido era un intervalo de R, la v.a. se llama continua.

Según esta definición, carece de sentido hablar de función de masa de probabilidad y por tanto se considera como valor teórico que la probabilidad de cualquier suceso elemental del espacio muestral, vale 0 (Puede razonarse utilizando la regla de Laplace).

No obstante, si puede definirse la función de distribución, esto es:

$F : R \to R$, de modo que $F(x) = P(X \leq x)$, $\forall x \in R$

Si el recorrido de la variable es el intervalo [a,b] se dan las siguientes propiedades:
**a)** $F(x) = 0, \forall x \leq a$
**b)** $F(x) = 1, \forall x \geq b$
**c)** $Si\ x_1 < x_2 \quad F(x_2) - F(x_1) = P(x_1 < X < x_2)$
**d)** Su representación gráfica es la de una función monótona creciente y continua.

**Función de densidad de una variable aleatoria continua:**

Se define dicha función, como la derivada de la función de distribución en aquellos puntos en que ésta última sea derivable.
$f(x) = F'(x)$

Teniendo en cuenta esta definición, se desprenden las siguientes propiedades:

a) $f(x) \geq 0$ (Es una función no negativa. Su gráfica siempre está por encima del eje X.

b) Por el teorema fundamental del cálculo diferencial $F(x) = \int_a^x f(t)dt$ = P(X<x)

c) Por la regla de Barrow $\int_a^b f(x)dx = F(b) - F(a) = 1$

**Medidas características de una variable continua:**

Media o esperanza $\mu = \int_{-\infty}^{+\infty} x.f(x)dx$, siendo f la función de densidad.

Varianza $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$

Desviación típica $\sigma = \sqrt{\sigma^2}$

La varianza puede hallarse más fácilmente así:

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2$$

## EJERCICIOS Y PROBLEMAS PROPUESTOS DE V.A. DISCRETA

1) Un experimento aleatorio consiste en lanzar una moneda al aire 3 veces. Sea x la variable aleatoria que indica el número de caras obtenidas.
   a) Hallar el espacio muestral y definir la variable aleatoria X.
   b) Hallar la función de masa de probabilidad de X y la función de distribución.
   c) Calcular la probabilidad de los sucesos $(X>1)$ $(1<X<3)$

2) Una urna contiene cuatro bolas negras y dos blancas. Consideremos el experimento aleatorio que consiste en extraer tres bolas (sin devolución) de la urna. Definir la variable aleatoria X "número de bolas negras obtenidas".
   a) Hallar la función de masa de probabilidad y de distribución de la v.a. X.
   b) Calcular la probabilidad de los sucesos $( X < 2 )$ $( X > 2)$

3) Dada la variable aleatoria X y la función de masa de probabilidad

| x | - 2 | - 1 | 0 | 2 | 4 |
|---|---|---|---|---|---|
| f(x) | 1 / 8 | 1 / 6 | 3 / 8 | 1 / 4 | 1 / 12 |

   a) Calcular la esperanza, varianza y desviación típica de X
   b) La función de distribución.

4) Hallar el valor de a en las siguientes distribuciones de probabilidad

| x | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| f(x) | 0,2 | a | 3a | 0,4 |

| x | 0 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| g(x) | a | 3 a | 2 a | 3 a | a |

   En ambos casos hallar la función de distribución y los parámetros

**5)** Una variable aleatoria X toma los valores 1, 2, 3 y 4, y su función de masa de probabilidad está dada por la tabla siguiente:

| $x_i$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $f(x_i)$ | 1/8 | ¼ | ¼ | 3/8 |

Se pide:
   a) La función de distribución y su gráfica.
   b) La esperanza matemática y desviación típica.

**6)** Considerese la variable aleatoria cuya función de masa de probabilidad es la siguiente:

| $x_i$ | -2 | 1 | 3 | 5 |
|---|---|---|---|---|
| $f(x_i)$ | 1/7 | 1/8 | 1/9 | a |

   a) Hállese el valor de a y la esperanza matemática.
   b) Suponiendo que F es la función de distribución, hállese F(2)

**7)** Hállese la función de distribución, la media, y la varianza, de una variable aleatoria X cuya función de masa de probabilidad es la siguiente:

| $x_i$ | -1 | 0,5 | 1 | 2 |
|---|---|---|---|---|
| $f(x_i)$ | 0,6 | 0,2 | 0,1 | 0,1 |

**8)** La función de distribución de una variable aleatoria es:

$$F(x)=\begin{cases} 0 & si \quad x<-1 \\ 1/2 & si \quad -1\le x<2 \\ 3/5 & si \quad 2\le x<3 \\ 6/7 & si \quad 3\le x<4 \\ 1 & si \quad x\ge 4 \end{cases}$$

¿Qué valores toma dicha variable aleatoria?. Hállese su función de masa de probabilidad.

**9)** La función de distribución de una variable aleatoria X es:

$$F(x)=\begin{cases} 0 & si \quad x<1 \\ 1/4 & si \quad 1\le x<2 \\ 1/3 & si \quad 2\le x<3 \\ 1 & si \quad 3\le x \end{cases}$$

Calcúlese:

a) $P(x\le 3)$ b) $P(1<X<3)$ c) $P(x>\frac{3}{2})$ d) $P(1<x\le 3)$ e) $P(1\le x<3)$

**10)** La función de distribución de una variable aleatoria X es:

$$F(x)=\begin{cases} 0 & si \quad x<2 \\ 1/3 & si \quad 2\le x<4 \\ 2/5 & si \quad 4\le x<6 \\ 1 & si \quad 6\le x \end{cases}$$

¿Qué valores toma X?. Hállese su función de masa de probabilidad, y calcúlense $P(2'5<X\le 5) \quad y \quad P(2<X5'5)$

**11)** Se efectúan tres lanzamientos de una moneda, y se pide:

a) Definir la variable aleatoria X que describe el número de caras obtenidas.
b) Hallar la función de masa de probabilidad de X
c) Representar gráficamente la función de distribución de X.

**12)** Un jugador lanza dos dados, y cobra tantos euros como veces aparezca el 1. Descríbase ese juego mediante una variable aleatoria. ¿Resulta rentable participar en el juego si para ello hay que pagar 3 euros por tirada?

**13)** Para participar en un juego se exige pagar 50 euros por tirada. El juego consiste en lanzar dos dados y formar el número de dos cifras más grande posible con los resultados obtenidos. Se cobran tantos euros como indica ese número. Se pide:

a) Describir el juego mediante una variable aleatoria.
b) Hallar la correspondiente función de masa de probabilidad.
c) ¿Resulta favorable al jugador participar en el juego?

**14)** Se elige al azar una ficha del dominó, y se considera la variable aleatoria que describe la suma de puntos que aparecen en la ficha. Se pide:

a) Hallar la correspondiente función de masa de probabilidad.
b) Calcular la media, la varianza, y la desviación típica.

**15)** Se lanza una moneda, tantas veces como sea necesario, hasta que aparezca una cara o hayan salido cinco cruces consecutivas, en cuyo caso se acaba el experimento.

a) ¿Qué valores toma la variable aleatoria "número de cruces"?. Hállese la correspondiente función de masa de probabilidad.

b) Hállese, también, la función de distribución, y dígase cuál es la probabilidad de obtener un máximo de tres cruces en el experimento.

**16)** En una lotería hay premios de 500000 euros, premios de 50000 euros y premios de 10000 euros. Las probabilidades de obtener cada uno de ellos son, respectivamente, 0,0001, 0,002, y 0,007. Suponiendo que se vendan todos los billetes, cada billete habrá dejado 180 euros de ganancia bruta a los organizadores. ¿Qué cuesta un billete?

**17)** Sea X la variable aleatoria que describe el menor de los números obtenidos en el lanzamiento de dos dados. Hállese la correspondiente función de distribución y hágase una representación gráfica de la misma.

**18)** Un dado ha sido trucado de modo que la probabilidad de cada cara sea inversamente proporcional al número que aparece en la misma, es decir, la función de masa de probabilidad para un lanzamiento es:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $f(x_i)$ | k/1 | k/2 | k/3 | k/4 | k/5 | k/6 |

a) Hállese el valor de k
b) Un jugador lanza el dado y cobra tantos euros como indique el número obtenido, ¿Cuánto debe pagar por tirada para que el juego sea equitativo?

**19)** En una urna hay cinco bolas señaladas con el número 1, dos señaladas con el nº 2, y cuatro con el 3. Se considera el experimento consistente en extraer una bola para mirar el número. Hállese la función de masa de probabilidad, la media, y la desviación típica asociadas a la variable aleatoria que describe el experimento.

**20)** Una persona trabaja en una empresa en cuyos alrededores no puede aparcar. Si lleva el coche al garaje, tiene que pagar 1,80 euros, mientras que si aparca en lugar prohibido, le pueden poner una multa de 3 euros, con probabilidad ½. ¿Qué le resulta más ventajoso?

**21)** Se exigen x euros por participar en este juego: Se lanza un dado y se cobran tantos euros como indique el número obtenido elevado al cuadrado.
a) Descríbase el juego mediante una variable aleatoria y determínese su función de masa de probabilidad.
b) ¿Para qué valor de x el juego es equitativo?

**22)** En una bolsa hay 30 bolas iguales, quince con el número 1, ocho con el 5, cinco con el 25, y dos con el 50. Se saca una bola al azar y se cobran tantos euros como indique el número que hay en ella.
a) Determinar la función de masa de probabilidad correspondiente.
b) ¿Es rentable participar en el juego pagando 5 euros por jugada?
c) ¿Cuál es la probabilidad de lograr 20 euros de ganancia neta en una jugada?

**23)** Un jugador lanza dos monedas. Gana 5 euros si aparecen dos caras, 2 euros si aparece una cara, y paga 10 euros si salen dos cruces. Suponiendo que no se paga entrada ninguna por jugar, ¿resulta rentable hacerlo?

**24)** Considérese el experimento consistente en lanzar cuatro veces una moneda. Sea X la variable aleatoria "número de caras obtenidas". Se pide:
a) La función de masa de probabilidad de X
b) La esperanza y la varianza de X

**25)** Se lanzan dos dados, y se llama X a la variable aleatoria que describe el número total de puntos conseguidos. Se pide:
a) La función de masa de probabilidad de X

b) La función de distribución de X, y la probabilidad de obtener un máximo de 9 puntos.

**26)** Una variable aleatoria X tiene por valores los números naturales del 1 al 100, y su función de masa de probabilidad f(x) está dada por:

$$f(x)=\begin{cases} ax & si \quad x\in\{1,2,3,...,100\} \\ 0 & si \quad x\notin\{1,2,3,...,100\} \end{cases}$$

Hallar la media de la variable X

**27)** Un alumno de un IES dice haber inventado el siguiente juego para dos: Cada jugador lanza dos monedas y el que obtenga mayor número de caras cobra un euro del otro por cada cara de más obtenida (si hay empate, nadie paga). Éste afirma que es el campeón indiscutible en ese juego, y reta a todos los alumnos de su instituto a que cada uno juegue con él diez partidas. El reto es aceptado por todos, aunque nadie sabe que una de las monedas con que jugará el tramposo está trucada de tal forma que la probabilidad de lograr cara con la misma es el doble de la de obtener cruz.
Suponiendo que hay en el instituto 300 alumnos, y que todos van a jugar con monedas no trucadas, ¿qué cantidad de dinero es de esperar que se lleve este alumno a sus arcas?

**28)** Una ruleta contiene 37 números, de 0 a 36. Hay muchas maneras de apostar, pero nos fijamos en un jugador que hace apuestas a números únicos: "Si deposita una ficha sobre un número y sale éste, entonces cobrará 36 veces la apuesta".
Sabemos que el jugador sólo tiene tres fichas y que tiene una predilección especial por el 12, razón por la que está empeñado en hacer una de las siguientes apuestas:

a) Apostar las tres fichas al 12
b) Apostar al 12 tres veces
c) Apostar cada ficha a un múltiplo de 12

¿Qué opción deberíamos aconsejarle?

## EJERCICIOS Y PROBLEMAS PROPUESTOS DE V.A. CONTINUA

1) Una función de densidad de una variable aleatoria es:

$$f(x) = \begin{cases} 1/5 & si \quad 0 < x < 5 \\ 0 & en\ otro\ caso \end{cases}$$

a) Calcula P ( 2 < X < 4 ), P ( X > 3 ), P ( X < 0 )
c) Representa gráficamente la función de distribución.

2) La función de densidad de una variable aleatoria x es $f(x) = \begin{cases} kx & si \quad x \in [0 \quad 5] \\ 0 & si \quad x \notin [0 \quad 5] \end{cases}$

a) Hallar k y la función de distribución.
b) Hallar la media y la varianza de x
c) Calcular P(0 < X < k )

3) La función de distribución de una variable aleatoria es

$$F(x) = \begin{cases} 0 & si \quad x < -1 \\ \dfrac{(x+1)^2}{4} & si \quad x \in [-1 \quad 1] \\ 1 & si \quad x \geq 1 \end{cases}$$

Calcular la función de densidad, la media y la varianza de x.

4) La función de densidad de una variable aleatoria X es $f(x) = \begin{cases} 0 & si \quad x < 0 \\ k & si \quad 0 \leq x < 1 \\ k' & si \quad 1 \leq x < 2 \\ 0 & si \quad x \geq 2 \end{cases}$

Hallar las constantes k y k' sabiendo que P(0 < X < 1,5) = 0,7. Luego hallar la función de distribución y dibujar su gráfica.

5) Obtener el valor de c sabiendo que es un número real positivo, para que la función

$$f(x) = \frac{e^x}{c}$$

pueda ser función de densidad de una variable aleatoria continua X, definida en el intervalo [ 0, 1 ]. Calcular la función de distribución, así como P ( X > ½) ; esperanza de X.

6) La función de densidad de una variable aleatoria es

$$f(x) = \begin{cases} kx + \dfrac{1}{2} & si \quad x \in [0 \quad 3] \\ 0 \quad si & x \notin [0 \quad 3] \end{cases}$$

Calcular la probabilidad de los sucesos A, A B, B U C, C siendo
A= ( X > 3/2), B = ( 1 < X < 3 ), C = ( 3/2 < X < 2 )

7) Un jugador tira al blanco. La distribución de los impactos en torno a la diana viene dada por la función de densidad $f(x) = e^{-kx}$ donde x representa la distancia del impacto a la diana. Hallar el valor de k y la función de distribución.

## TEMA VIII. DISTRIBUCIÓN BINOMIAL (Caso particular de v.a. discreta)

### Experimento de Bernouilli:

Se llama así a un experimento aleatorio con las siguientes características:

a) En cada prueba estudiamos sólo la realización de un suceso A (éxito) y su contrario ContA (fracaso). Se realizan n pruebas.

b) La proporción de éxitos y fracasos es constante en la población y no se modifica cualquiera que sea la cantidad de elementos de la población observada. Llamamos p = P(A) , probabilidad de éxito y q = P(contA) = 1-p, probabilidad de fracaso.

c) Las n pruebas son independientes; es decir, el resultado de una prueba no depende de las precedentes.

Este experimento genera un espacio muestral del tipo:

$\Omega = \{AA...A, \quad \neg AA...A, \quad ... \quad \neg A\neg A...\neg A\}$, que tiene exactamente $2^n$ elementos ya que como podemos observar son las variaciones con repetición de 2 elementos (A y su contrario) tomados de n en n.

Sobre este espacio muestral definimos la siguiente variable aleatoria

X="número de éxitos"

Es obvio que el recorrido de X es {0,1,2,3,...,n} y dado que es finito estamos ante una variable aleatoria discreta.

Bajo las circunstancias anteriores, se dice que X es una variable aleatoria que sigue una distribución binomial de parámetros n y p, representándose así:

B(n,p)

### Función de masa de probabilidad de una binomial:

Dado que el recorrido de X es {0,1,2,3,...,n}, la función de masa de probabilidad es:

f: $\{0,1,2,3,...,n\} \rightarrow [0,1]$,

siendo f(r) = P(X=r) o dicho de otro modo, la probabilidad de que al realizar n pruebas, se obtengan r éxitos (r toma valores de 0 a n)

Si aplicamos el hecho de que las pruebas son independientes, la probabilidad pedida es el producto de las probabilidades en cada prueba, pero en cada prueba sólo puede salir A o contA, y como sabemos que A aparece r veces y contA n-r veces, resulta que la probabilidad de cada caso es $p^r q^{n-r}$, ahora bien...¿en cuántos casos salen exactamente r éxitos? La combinatoria nos dice que son $RP_r^{r,n-r} = \dfrac{r!}{r!(n-r)!}$, pero obsérvese que esto es lo mismo que $C_n^r = \dbinom{n}{r}$, por lo tanto la probabilidad buscada es:

$$f(r) = \binom{n}{r} p^r q^{n-r}, donde\ r = 0,1,2,...,n$$

Dada la complejidad en los cálculos, estos valores vienen ya determinados en una tabla.

En la siguiente página se adjunta una tabla para el cálculo de la función de masa de probabilidad para n desde 2 hasta 9.

| n | k | p | 0,01 | 0,05 | 0,10 | 0,15 | 0,20 | 0,25 | 0,30 | 0,33 | 0,35 | 0,40 | 0,45 | 0,49 | 0,50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0 | | 0,9801 | 0,9025 | 0,8100 | 0,7225 | 0,6400 | 0,5625 | 0,4900 | 0,4449 | 0,4225 | 0,3600 | 0,3025 | 0,2601 | 0,2500 |
| | 1 | | 0,0198 | 0,0950 | 0,1800 | 0,2550 | 0,3200 | 0,3750 | 0,4200 | 0,4442 | 0,4550 | 0,4800 | 0,4950 | 0,4998 | 0,5000 |
| | 2 | | 0,0001 | 0,0025 | 0,0100 | 0,0225 | 0,0400 | 0,0625 | 0,0900 | 0,1109 | 0,1225 | 0,1600 | 0,2025 | 0,2401 | 0,2500 |
| 3 | 0 | | 0,9703 | 0,8574 | 0,7290 | 0,6141 | 0,5120 | 0,4219 | 0,3430 | 0,2967 | 0,2746 | 0,2160 | 0,1664 | 0,1327 | 0,1250 |
| | 1 | | 0,0294 | 0,1354 | 0,2430 | 0,3251 | 0,3840 | 0,4219 | 0,4410 | 0,4444 | 0,4436 | 0,4320 | 0,4084 | 0,3823 | 0,3750 |
| | 2 | | 0,0003 | 0,0071 | 0,0270 | 0,0574 | 0,0960 | 0,1406 | 0,1890 | 0,2219 | 0,2389 | 0,2880 | 0,3341 | 0,3674 | 0,3750 |
| | 3 | | 0,0000 | 0,0001 | 0,0010 | 0,0034 | 0,0080 | 0,0156 | 0,0270 | 0,0369 | 0,0429 | 0,0640 | 0,0911 | 0,1176 | 0,1250 |
| 4 | 0 | | 0,9606 | 0,8145 | 0,6561 | 0,5220 | 0,4096 | 0,3164 | 0,2401 | 0,1979 | 0,1785 | 0,1296 | 0,0915 | 0,0677 | 0,0625 |
| | 1 | | 0,0388 | 0,1715 | 0,2916 | 0,3685 | 0,4096 | 0,4219 | 0,4116 | 0,3953 | 0,3845 | 0,3456 | 0,2995 | 0,2600 | 0,2500 |
| | 2 | | 0,0006 | 0,0135 | 0,0486 | 0,0975 | 0,1536 | 0,2109 | 0,2646 | 0,2960 | 0,3105 | 0,3456 | 0,3675 | 0,3747 | 0,3750 |
| | 3 | | 0,0000 | 0,0005 | 0,0036 | 0,0115 | 0,0256 | 0,0469 | 0,0756 | 0,0985 | 0,1115 | 0,1536 | 0,2005 | 0,2400 | 0,2500 |
| | 4 | | 0,0000 | 0,0000 | 0,0001 | 0,0005 | 0,0016 | 0,0039 | 0,0081 | 0,0123 | 0,0150 | 0,0256 | 0,0410 | 0,0576 | 0,0625 |
| 5 | 0 | | 0,9510 | 0,7738 | 0,5905 | 0,4437 | 0,3277 | 0,2373 | 0,1681 | 0,1320 | 0,1160 | 0,0778 | 0,0503 | 0,0345 | 0,0313 |
| | 1 | | 0,0480 | 0,2036 | 0,3281 | 0,3915 | 0,4096 | 0,3955 | 0,3602 | 0,3295 | 0,3124 | 0,2592 | 0,2059 | 0,1657 | 0,1563 |
| | 2 | | 0,0010 | 0,0214 | 0,0729 | 0,1382 | 0,2048 | 0,2637 | 0,3087 | 0,3291 | 0,3364 | 0,3456 | 0,3369 | 0,3185 | 0,3125 |
| | 3 | | 0,0000 | 0,0011 | 0,0081 | 0,0244 | 0,0512 | 0,0879 | 0,1323 | 0,1643 | 0,1811 | 0,2304 | 0,2757 | 0,3060 | 0,3125 |
| | 4 | | 0,0000 | 0,0000 | 0,0005 | 0,0022 | 0,0064 | 0,0146 | 0,0284 | 0,0410 | 0,0488 | 0,0768 | 0,1128 | 0,1470 | 0,1563 |
| | 5 | | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0003 | 0,0010 | 0,0024 | 0,0041 | 0,0053 | 0,0102 | 0,0185 | 0,0282 | 0,0313 |
| 6 | 0 | | 0,9415 | 0,7351 | 0,5314 | 0,3771 | 0,2621 | 0,1780 | 0,1176 | 0,0881 | 0,0754 | 0,0467 | 0,0277 | 0,0176 | 0,0156 |
| | 1 | | 0,0571 | 0,2321 | 0,3543 | 0,3993 | 0,3932 | 0,3560 | 0,3025 | 0,2638 | 0,2437 | 0,1866 | 0,1359 | 0,1014 | 0,0938 |
| | 2 | | 0,0014 | 0,0305 | 0,0984 | 0,1762 | 0,2458 | 0,2966 | 0,3241 | 0,3292 | 0,3280 | 0,3110 | 0,2780 | 0,2436 | 0,2344 |
| | 3 | | 0,0000 | 0,0021 | 0,0146 | 0,0415 | 0,0819 | 0,1318 | 0,1852 | 0,2191 | 0,2355 | 0,2765 | 0,3032 | 0,3121 | 0,3125 |
| | 4 | | 0,0000 | 0,0001 | 0,0012 | 0,0055 | 0,0154 | 0,0330 | 0,0595 | 0,0821 | 0,0951 | 0,1382 | 0,1861 | 0,2249 | 0,2344 |
| | 5 | | 0,0000 | 0,0000 | 0,0001 | 0,0004 | 0,0015 | 0,0044 | 0,0102 | 0,0164 | 0,0205 | 0,0369 | 0,0609 | 0,0864 | 0,0938 |
| | 6 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0002 | 0,0007 | 0,0014 | 0,0018 | 0,0041 | 0,0083 | 0,0138 | 0,0156 |
| 7 | 0 | | 0,9321 | 0,6983 | 0,4783 | 0,3206 | 0,2097 | 0,1335 | 0,0824 | 0,0587 | 0,0490 | 0,0280 | 0,0152 | 0,0090 | 0,0078 |
| | 1 | | 0,0659 | 0,2573 | 0,3720 | 0,3960 | 0,3670 | 0,3115 | 0,2471 | 0,2053 | 0,1848 | 0,1306 | 0,0872 | 0,0604 | 0,0547 |
| | 2 | | 0,0020 | 0,0406 | 0,1240 | 0,2097 | 0,2753 | 0,3115 | 0,3177 | 0,3074 | 0,2985 | 0,2613 | 0,2140 | 0,1740 | 0,1641 |
| | 3 | | 0,0000 | 0,0036 | 0,0230 | 0,0617 | 0,1147 | 0,1730 | 0,2269 | 0,2558 | 0,2679 | 0,2903 | 0,2918 | 0,2786 | 0,2734 |
| | 4 | | 0,0000 | 0,0002 | 0,0026 | 0,0109 | 0,0287 | 0,0577 | 0,0972 | 0,1277 | 0,1442 | 0,1935 | 0,2388 | 0,2676 | 0,2734 |
| | 5 | | 0,0000 | 0,0000 | 0,0002 | 0,0012 | 0,0043 | 0,0115 | 0,0250 | 0,0383 | 0,0466 | 0,0774 | 0,1172 | 0,1543 | 0,1641 |
| | 6 | | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0004 | 0,0013 | 0,0036 | 0,0064 | 0,0084 | 0,0172 | 0,0320 | 0,0494 | 0,0547 |
| | 7 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0002 | 0,0005 | 0,0006 | 0,0016 | 0,0037 | 0,0068 | 0,0078 |
| 8 | 0 | | 0,9227 | 0,6634 | 0,4305 | 0,2725 | 0,1678 | 0,1001 | 0,0576 | 0,0392 | 0,0319 | 0,0168 | 0,0084 | 0,0046 | 0,0039 |
| | 1 | | 0,0746 | 0,2793 | 0,3826 | 0,3847 | 0,3355 | 0,2670 | 0,1977 | 0,1565 | 0,1373 | 0,0896 | 0,0548 | 0,0352 | 0,0313 |
| | 2 | | 0,0026 | 0,0515 | 0,1488 | 0,2376 | 0,2936 | 0,3115 | 0,2965 | 0,2734 | 0,2587 | 0,2090 | 0,1569 | 0,1183 | 0,1094 |
| | 3 | | 0,0001 | 0,0054 | 0,0331 | 0,0839 | 0,1468 | 0,2076 | 0,2541 | 0,2730 | 0,2786 | 0,2787 | 0,2568 | 0,2273 | 0,2188 |
| | 4 | | 0,0000 | 0,0004 | 0,0046 | 0,0185 | 0,0459 | 0,0865 | 0,1361 | 0,1704 | 0,1875 | 0,2322 | 0,2627 | 0,2730 | 0,2734 |
| | 5 | | 0,0000 | 0,0000 | 0,0004 | 0,0026 | 0,0092 | 0,0231 | 0,0467 | 0,0680 | 0,0808 | 0,1239 | 0,1719 | 0,2098 | 0,2188 |
| | 6 | | 0,0000 | 0,0000 | 0,0000 | 0,0002 | 0,0011 | 0,0038 | 0,0100 | 0,0170 | 0,0217 | 0,0413 | 0,0703 | 0,1008 | 0,1094 |
| | 7 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0004 | 0,0012 | 0,0024 | 0,0033 | 0,0079 | 0,0164 | 0,0277 | 0,0313 |
| | 8 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0002 | 0,0002 | 0,0007 | 0,0017 | 0,0033 | 0,0039 |
| 9 | 0 | | 0,9135 | 0,6302 | 0,3874 | 0,2316 | 0,1342 | 0,0751 | 0,0404 | 0,0261 | 0,0207 | 0,0101 | 0,0046 | 0,0023 | 0,0020 |
| | 1 | | 0,0830 | 0,2985 | 0,3874 | 0,3679 | 0,3020 | 0,2253 | 0,1556 | 0,1174 | 0,1004 | 0,0605 | 0,0339 | 0,0202 | 0,0176 |
| | 2 | | 0,0034 | 0,0629 | 0,1722 | 0,2597 | 0,3020 | 0,3003 | 0,2668 | 0,2345 | 0,2162 | 0,1612 | 0,1110 | 0,0776 | 0,0703 |
| | 3 | | 0,0001 | 0,0077 | 0,0446 | 0,1069 | 0,1762 | 0,2336 | 0,2668 | 0,2731 | 0,2716 | 0,2508 | 0,2119 | 0,1739 | 0,1641 |
| | 4 | | 0,0000 | 0,0006 | 0,0074 | 0,0283 | 0,0661 | 0,1168 | 0,1715 | 0,2045 | 0,2194 | 0,2508 | 0,2600 | 0,2506 | 0,2461 |
| | 5 | | 0,0000 | 0,0000 | 0,0008 | 0,0050 | 0,0165 | 0,0389 | 0,0735 | 0,1021 | 0,1181 | 0,1672 | 0,2128 | 0,2408 | 0,2461 |
| | 6 | | 0,0000 | 0,0000 | 0,0001 | 0,0006 | 0,0028 | 0,0087 | 0,0210 | 0,0340 | 0,0424 | 0,0743 | 0,1160 | 0,1542 | 0,1641 |
| | 7 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0003 | 0,0012 | 0,0039 | 0,0073 | 0,0098 | 0,0212 | 0,0407 | 0,0635 | 0,0703 |
| | 8 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0004 | 0,0009 | 0,0013 | 0,0035 | 0,0083 | 0,0153 | 0,0176 |
| | 9 | | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0001 | 0,0003 | 0,0008 | 0,0016 | 0,0020 |

**Función de distribución de una B(n,p):**

Se trata de la función de mása de probabilidad acumulada, por tanto será una función del tipo:

$F : R \rightarrow R$, de tal modo que: $F(x) = P(X \leq x) = \sum_{r=0}^{q} f(r)$, siendo q el mayor entero tal que $q \leq x$.

Verifica todas las propiedades de las funciones de distribución de una variable aleatoria discreta.

Medidas características de una distribución binomial:

Media o esperanza:

$$\mu = \sum_{r=0}^{n} r.f(r) = np$$

Desviación típica:

$$\sigma = \sqrt{\sum_{r=0}^{n} (r-\mu)^2 f(r)} = \sqrt{npq}$$

**Como calcular los valores de la función de masa de probabilidad y la función de distribución de una binomial n, p en EXCEL:**

La función que lo determina es, dentro de las funciones estadísticas, DISTR.BINOM. y los parámetros que pide son:

Núm_éxito (número); Ensayos (número); Prob_éxito (número); Acumulado (valor lógico)

Núm éxito es el valor de r; Ensayos es el valor de n o número de pruebas; Prob_éxito es p; Acumulado puede ser Verdadero o Falso. Si es verdadero da la el valor de la función de distribución F(r), mientras que si es falsa da el valor de la función de masa de probabilidad f(r)

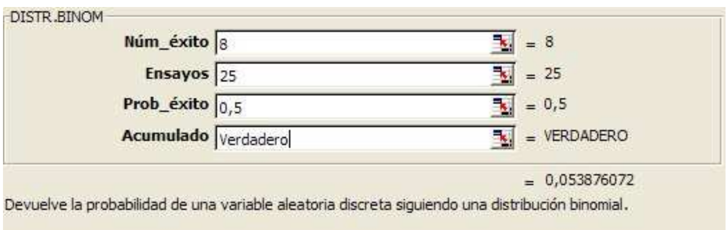



**Ilustración 1**

Supongamos que lanzamos 25 veces una moneda y queremos saber cual es la probabilidad de obtener 8 caras como máximo.

En Excel introduciríamos en la función DISTR.BINOM., los parámetros de la ilustración 1, es decir, hallaríamos F(8) cuyo valor sería 0,053. Por el contrario si quisiéramos hallar f(8), es decir la probabilidad de obtener 8 caras exactamente, solamente tendríamos que modificar el parámetro "Acumulado", donde tendríamos que consignar "Falso".

La función BINOM.CRIT.

Actúa de modo inverso que la anterior, esto es: Si conocemos el valor de probabilidad de un suceso, averigua el valor de r de forma que la probabilidad hasta dicho valor, acumulada, coincida con la dada.

Por ejemplo: Supongamos que en el ejemplo anterior queremos conocer cuántas caras como máximo dan probabilidad 0,7.

El resultado viene dado en la siguiente ilustración:

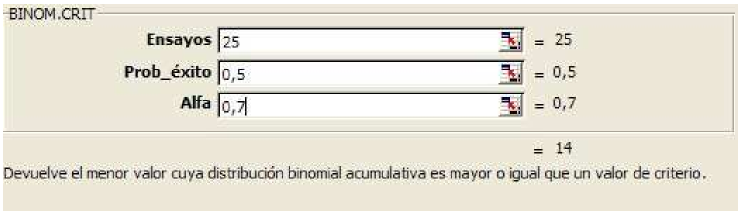



**Ilustración 2**

## PROBLEMAS PROPUESTOS

1) En un taller hay 10 máquinas iguales. Se ha visto que una máquina determinada un día de cada cinco está averiada. ¿ Calcula la probabilidad de que un cierto día haya más de 7 máquinas averiadas ?. Si es 5000 pesetas la pérdida diaria ocasionada por tener una máquina averiada, calcular la pérdida media diaria.
2) De la producción diaria de una cierta pieza se examinan 10 de dichas piezas durante 23 días, dando la siguiente tabla de piezas defectuosas:

| dias | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.d. | 0 | 1 | 1 | 2 | 1 | 3 | 2 | 2 | 1 | 0 | 1 | 2 | 1 | 2 | 0 | 2 | 3 | 0 | 2 | 1 | 0 | 1 | 1 |

Suponiendo que la probabilidad de fabricar una pieza defectuosa es fija, ajustar una distribución binomial a las observaciones.

3) Si la distribución hallada en el problema anterior es la verdadera ley del proceso, ¿Cuál es la probabilidad de que en las 10 piezas observadas, de un día determinado, haya más de 2 defectuosas.
4) Se sabe que, en un nacimiento, no se tiene la misma probabilidad de que sea niño que niña, pues la experiencia nos dice que nacen más niños que niñas. Si suponemos que de cada 100 recién nacidos 55 son varones y 45 mujeres. a) ¿ Cuál es la probabilidad que tiene un recién nacido de ser mujer ?
   b) ¿ Cuál es la probabilidad de que los 5 primeros recién nacidos del año en un hospital sean niñas?
   c) ¿ Cuál es la probabilidad de que haya exactamente 2 niñas entre los cinco primeros?

5) En los exámenes de selectividad del curso 1997/98 aprobaron en Galicia el 85% de los alumnos presentados. Calcular la probabilidad de que al coger 7 alumnos a) aprueben 3. b) aprueben más de uno.
6) Para participar en un concurso de tiro al plato hay que abonar una cuota. Cada participante realiza 10 disparos: Si acierta 5 recupera el importe pagado abandonando la competición y si acierta más se clasifica para la siguiente ronda. Un competidor muy regular acostumbra a acertar el 40% de sus disparos. a) ¿Cuál es la probabilidad de que acierte exactamente 5 disparos?. b) ¿Cuál es la probabilidad de que se clasifique?
7) Al transmitir una comunicación, la probabilidad de distorsionar un signo es igual a 1/10. ¿Cuáles son las probabilidades de que en una comunicación de 10 signos a) No sea distorsionada. b) Contenga exactamente tres distorsiones. c) Contenga tres distorsiones como máximo.
8) Cada miembro de un comité de 9 personas acude a las reuniones con una probabilidad igual a ½. ¿Cuál es la probabilidad de que, como mucho, se reunan 2/3 de los miembros.
9) Se lanza una moneda a) 4 veces, b) 5 veces, c) 6 veces. ¿Cuál es la probabilidad en cada caso de obtener un número impar de caras?.¿ Y para n veces?
10) Hallar la probabilidad de obtener un total de 11 a) una vez, (b) dos veces, en dos lanzamientos de un par de dados
11) ¿Cuál es la probabilidad de obtener 9 una vez en tres lanzamientos de un par de dados?

12) Un vendedor de seguros vende pólizas a 5 hombres, todos de la misma edad y con buena salud. Se sabe que la probabilidad de que un hombre viva 30 años o más es 2/3. Hallar la probabilidad de que a los 30 años vivan a) los 5 hombres. b) al menos 3. c) Solamente 2. d) al menos 1.
13) Se lanzan 6 veces una moneda ¿Cuál es la probabilidad de que el resultado cruz no salga nunca más veces que el resultado cara?
14) Un jugador propone a un amigo el siguiente juego: se lanza 20 veces una moneda. El amigo gana si aparece cara 9, 10 o 11 veces y pierde en caso contrario. Este juego ¿ es favorable al amigo ?
15) Se ha estudiado que 1/3 de los alumnos de COU no leen nunca la prensa diaria. Tomando una muestra al azar de 10 alumnos estudiar las probabilidades siguientes:
    a) Encontrar dos alumnos que no leen la prensa. b) Más de 3 alumnos que no leen la prensa. c) Por lo menos cinco alumnos que no leen la prensa.

**SOLUCIONES**: **1)** 0,0035 ; 10.000 ptas. // **2)** B (10, 0,126) // **3)** 0,1219 // **4)** a. 0,45; b.0,0185; c.0,3369 // **5)** a. 0,0109; b. 0,9999 // **6)** a. 0,2007; b. 0,1663 // **7)** a. 0,3487; b. 0,0574; c. 0,9872 // **8)** 0,9101 // **9)** a. 0,5; b. 0,5; c. 0,5 // **10)** a. 17/162; b. 1/324 // **11)** 64/243 // **12)** a. 0,1317; b. 0,7901; c. 0,1646; d. 0,9959 // **13)** 0,6563 // **14)** No es favorable // **15)** a. 0.1951; b. 0,4408; c. 0,2131 //

**TEMA IX. DISTRIBUCIÓN NORMAL** (Caso particular de v.a. continua)

Una variable aleatoria continua X se dice que sigue una distribución ormal de media $\mu$ y desviación típica $\sigma$, simbolizándose por N($\mu$,$\sigma$), si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \;;\; -\infty < x < +\infty$$

La distribución normal aparece espontáneamente en multitud de problemas, medidas físicas del cuerpo humano, características psíquicas, medidas de calidad,...,etc.

**Gráfica de f(x)**

1) f tiene dominio en R y es continua.
2) f(x)>0, para todo x real.
3) f es simétrica respecto a la recta $x = \mu$
4) f tiene un máximo en el punto $(\mu, \frac{1}{\sigma\sqrt{2\pi}})$
5) f tiene una asíntota en el eje OX
6) f presenta puntos de inflexión en las abscisas x-$\mu$ , x+$\mu$

Este es el aspecto que presenta:

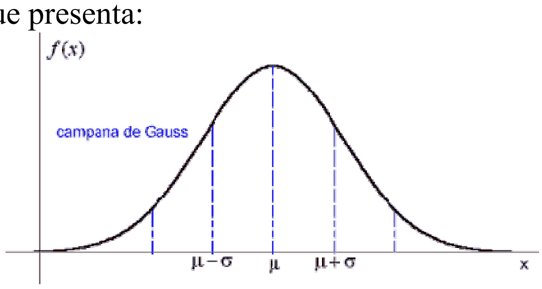



**Ilustración 3**

**Función de distribución de una variable aleatoria normal:**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2} dt$$

**Distribución normal estándar o tipificada:**

**Cuando** $\mu$=0 y $\sigma$=1, entonces la variable aleatoria normal N(0,1) se llama distribución normal tipificada o estándar. La función de densidad correspondiente es:

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

**TEOREMA (Tipificación de una variable aleatoria normal)**

Si X es una v.a. normal N($\mu$ ,$\sigma$ ), entonces la variable Z=$\frac{X-\mu}{\sigma}$ es una normal tipificada.

En efecto, recurriendo a las propiedades de la esperanza y la varianza, se tiene:

$$E\left[\frac{X-\mu}{\sigma}\right]=\frac{1}{\sigma}E[X-\mu]=\frac{1}{\sigma}(E[X]-\mu)=0$$

$$Var\left[\frac{X-\mu}{\sigma}\right]=\frac{1}{\sigma^2}Var[X-\mu]=\frac{1}{\sigma^2}Var[X]=\frac{1}{\sigma^2}\sigma^2=1$$

En base a este resultado, como toda normal puede, mediante un cambio de variable, convertirse en una tipificada, cualquier resultado puede ser estudiado en las tablas de la variable Z y luego volver a deshacer el cambio de variable para obtener el resultado en la variable original X.

Por ello, la función de distribución de la variable normal tipificada viene desarrollada en la siguiente tabla:

**TABLA DE DISTRIBUCION NORMAL ACUMULADA**

| | | | | | SEGUNDO DECIMAL DE Z | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Z** | **0.00** | **0.01** | **0.02** | **0.03** | **0.04** | **0.05** | **0.06** | **0.07** | **0.08** | **0.09** |
| **0.0** | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| **0.1** | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| **0.2** | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| **0.3** | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| **0.4** | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| **0.5** | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| **0.6** | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| **0.7** | .2580 | .2611 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| **0.8** | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| **0.9** | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| **1.0** | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| **1.1** | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| **1.2** | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| **1.3** | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| **1.4** | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| **1.5** | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| **1.6** | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| **1.7** | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| **1.8** | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| **1.9** | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| **2.0** | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| **2.1** | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| **2.2** | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| **2.3** | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| **2.4** | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| **2.5** | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| **2.6** | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| **2.7** | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| **2.8** | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| **2.9** | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| **3.0** | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| **3.1** | .4990 | .4991 | .4991 | .4991 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| **3.2** | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| **3.3** | .4995 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| **3.4** | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 |
| **3.5** | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 |
| **3.6** | .4998 | .4998 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| **3.7** | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| **3.8** | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| **3.9** | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 |

Obsérvese que F(z) viene tabulada para valores de 0 en adelante, asi pues se hace necesario averiguar F(-z) cuando z>0. El resultado es: F(-z) = 1-F(z)

La demostración es trivial.

También se tiene que $P(z_1 \leq Z \leq z_2) = F(z_2) - F(z_1)$

**Funciones en EXCEL para trabajar con la distribución normal:**

La primera es DISTR.NORMAL que devuelve el valor de probabilidad de un valor x de la variable (en realidad toma un pequeño intervalo entorno al mismo) o el valor de probabilidad acumulada hasta el mismo. La ilustración siguiente lo refleja:

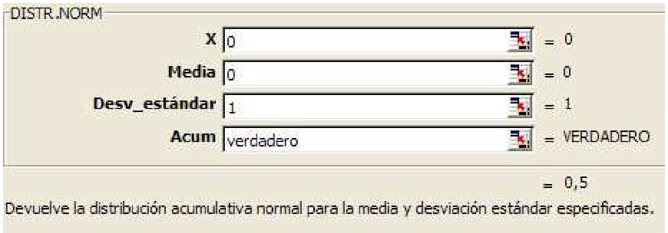



**Ilustración 4**

Donde x es el valor de la variable objeto de estudio, Media es el valor de la media de la variable, Desv_estándar es el valor de la desviación típica y acumulado es un campo lógico que puede tomar dos valores: verdadero (se obtiene F(x)) y falso (se obtiene P(0,5-x<X<0,5+x))

La inversa de la anterior es DISTR.NORM.INV. donde, una vez conocido el valor de probabilidad acumulado, nos devuelve el valor de la variable donde se alcanza dicha probabilidad. Veamos el ejemplo en una N(3,0.5) para probabilidad 0,3. En la ilustración 5 observamos que el resultado es 2,73. Esto quiere decir que P(X<2,73)=0,3

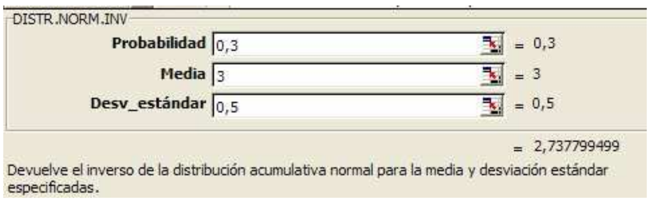



**Ilustración 5**

Análogamente a la anterior, tenemos la DISTR.NORM.ESTAND. que en trabaja directamente con la N(0,1) y los únicos parámetros a introducir son el valor de z . Obteniéndose como resultado F(z).

La función DISTR.NORM.ESTAND.INV., actúa de forma inversa que la anterior, es decir, conocemos la probabilidad y deseamos averiguar el valor de z con ese valor de probabilidad acumulado.

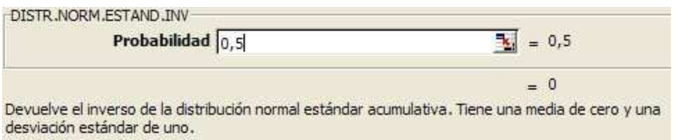



**Ilustración 6**

## APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL

Cuando estamos ante una distribución binomial con un número de pruebas muy elevado, se hace muy laborioso el cálculo de las probabilidades ya que las tablas publicadas normalmente no van más allá de 10 pruebas.

Si no disponemos de un programa (EXCEL u otros) que nos permiten averiguar el resultado exacto, en este caso se utiliza el siguiente teorema que permite calcular probabilidades en una distribución binomial, utilizando la normal. Esta aproximación será tanto mejor como el valor de p se aproxime a 0,5 y n sea muy grande.

**Si X sigue una distribución B(n,p), entonces sus valores pueden ser aproximados mediante una distribución normal** $N(np, \sqrt{npq})$**.**

Aquí hay que hacer la siguiente matización:

Dado que en una distribución normal, la probabilidad de un valor concreto de la variable era 0, ¿cómo se efectúa la aproximación anterior toda vez que en la distribución binomial ese valor no es nulo?.

La respuesta es la siguiente:

$$P_B(X = a) \cong P_N(a - 0{,}5 \leq a \leq a + 0{,}5)$$

Entendemos por $P_B$ cuando trabajamos con la binomial y $P_N$ cuando se aplica la normal.

Este procedimiento de tomar media unidad antes o después de valor de la variable, se extiende también a las probabilidades de los intervalos, es decir:

$$P_B(a < X < b) \cong P_N(a + 0{,}5 \leq X \leq b - 0{,}5)$$

$$P_B(a \leq X \leq b) \cong P_N(a - 0{,}5 \leq X \leq b + 0{,}5)$$

Recordemos que en cualquier caso, cuando se trata de trabajar con la distribución normal, es indiferente, a efectos de cálculo de probabilidades en intervalos, que se tomen o no los extremos, es decir que:

$P_N(a < X < b) = P_N(a \leq X \leq b)$, debido precisamente al carácter continuo de la variable.

**PROBLEMAS PROPUESTOS**

1) Si los diametros de los cojinetes de bolas se distribuyen normalmente con media 0,6140 pulgadas y desviación típica 0,0025 pulgadas, determinar el porcentaje de cojinetes de bolas con diametro:
   a) Entre 0,610 y 0,618 pulgadas
   b) Mayor de 0,617 pulgadas
   c) Menor de 0,608 pulgadas.

2) La puntuación media en un examen final fue 72 y la desviación típica 9. El 10% de los alumnos tuvieron la calificación de sobresaliente (nota máxima). ¿ Cuál es la puntuación mínima para recibir un sobresaliente sabiendo que la distribución de las notas es normal.
3) Una variable aleatoria X tiene una distribución normal con una esperanza matemática igual a 0 y una varianza igual a 1. ¿ Cuál de los dos sucesos tiene mayor probabilidad ?

$$( |x| < 0,7) \quad \text{ó} \quad ( |x| > 0,7)$$

4) Cierta categoría de individuos tiene un peso medio de 60 Kg. y una desviación típica del peso igual a
3 Kg. Determinar la probabilidad de que el peso de un individuo tomado al azar se distinga de la me
dia no más que en 5 Kg, sabiendo que el peso se distribuye normalmente.

4) La talla de los hombres en edad militar en España sigue una distribución normal de media 169 cm. y desviación típica 6 cm. Si no se admiten para el servicio militar los individuos de talla inferior a 150 cm., ¿ qué proporción se rechaza ?

5) Supongamos que en cierto curso las notas de matemáticas se distribuyen normalmente con una media de 6 (sobre 10) y una desviación típica de 3.

   a) Si se elige un alumno al azar, ¿ cuál es la probabilidad de que su nota esté comprendida entre 7 y 8 ambos inclusive?
   b) ¿ Y la de que tenga una nota igual o superior a 9 ?

6) Una población de mazorcas de maiz se distribuye normalmente respecto al carácter longitud con una media de 25 cm. y una desviación típica de 5. Calcular la probabilidad de que una mazorca elegida al azar mida a) Entre 22 y 27 cm. b) 27 o más de 27 cm. c) entre 26 y 30. d) a lo sumo 20 cm.

7) Entre 2000 estudiantes la media del peso resultó ser 70 Kg. con una desviación típica de 8,5 Kg. Determinar el peso mínimo del conjunto formado por los 200 estudiantes más pesados.

8) Los errores aleatorios que se cometen en las pesadas de una balanza siguen una normal de media 0 y desviación típica 2 decigramos. Hallar el error máximo en una pesada con una probabilidad del 0,95.

9) La media de una v.a. X normal es el quíntuplo de la desviación típica. Sabiendo que P (X<6) = 0,8413, calcular la media y la desviación típica.

10) Hallar la probabilidad de que entre 100.000 cifras al azar la cifra 6 salga menos de 9.971 veces.

11) Hallar el valor aproximado de la probabilidad de que la cantidad de “nueves” entre 10.000 números aleatorios quede comprendido entre 940 y 1060. ( Se llaman números aleatorios al número resultantes de elegir al azar un dígito entre 0 y 9 ambos inclusive)

12) Sea A un suceso de probabilidad 0,4. Suponiendo que se hacen 900 pruebas del experimento, calcular la probabilidad de que A se verifique entre 360 y 390 veces. ¿Cuál es la probabilidad de que A se verifique exactamente 380 veces.?

13) Hallar la probabilidad de obtener tantas caras como cruces en 100 lanzamientos de una moneda.

**SOLUCIONES**: **1)** a. 89% ; b. 11,5 % ; c. 0,8 % // **2)** 83,5 // **3)** 0,5178 ; 0,4822 // **4)** 0,9030 // **4)** 8 de cada 10.000 // **5)** a. 0,1161 ; b. 0,1587 // **6)** a. 0,3811 ; b. 0,3446 ; c. 0,2620 ; d) 0,1587 // **7)** 80,8 // **8)** 3,3 dg. // **9)** x = 3 ; $\sigma = 0,6$ // **10)** 0,1814 // **11)** 0,9998 // **12)** 0,4652 ; 0,0111 // **13)** 0,07958

## PROBLEMAS RESUELTOS DE VARIABLE ALEATORIA

**5) Un experimento aleatorio consiste en lanzar una moneda al aire 3 veces. Sea x la variable aleatoria que indica el número de caras obtenidas.**
   **d) Hallar el espacio muestral y definir la variable aleatoria X.**
   **e) Hallar la función de masa de probabilidad de X y la función de distribución.**
   **f) Calcular la probabilidad de los sucesos (X>1) (1<X<3)**

SOLUCIÓN:
   a) El espacio muestral es {ccc, cc+,c+c,+cc,++c,+c+,c++,+++}. Siendo el recorrido de la variable aleatoria {0,1,2,3}.
   b) función de masa de probabilidad viene dada por la siguiente tabla:

| $x_i$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(x_i)$ | 1/8 | 3/8 | 3/8 | 1/8 |

la función de distribución es:

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 1/8 & si \quad 0 \le x < 1 \\ 4/8 & si \quad 1 \le x < 2 \\ 7/8 & si \quad 2 \le x < 3 \\ 1 & si \quad x \ge 3 \end{cases}$$

c) P(X>1) = 4/8 ; P(1<X<3) = 3/8

**6) Una urna contiene cuatro bolas negras y dos blancas. Consideremos el experimento aleatorio que consiste en extraer tres bolas (sin devolución) de la urna. Definir la variable aleatoria X "número de bolas negras obtenidas".**
**a) Hallar la función de masa de probabilidad y de distribución de la v.a. X.**
**b) Calcular la probabilidad de los sucesos ( X < 2 ) ( X > 2)**

SOLUCIÓN:
   a) La función de masa de probabilidad viene dada por la tabla:

| $x_i$ | 1 | 2 | 3 |
|---|---|---|---|
| $f(x_i)$ | 24/120 | 72/120 | 24/120 |

b) P(X<2) = 24/120 ; P(X>2) = 24/120

**7) Dada la variable aleatoria X y la función de masa de probabilidad**

| **x** | **- 2** | **- 1** | **0** | **2** | **4** |
|---|---|---|---|---|---|
| **f(x)** | **1 / 8** | **1 / 6** | **3 / 8** | **1 / 4** | **1 / 12** |

**a) Calcular la esperanza, varianza y desviación típica de X**
**b) La función de distribución.**

SOLUCIÓN:

a) E[X] = -2.1/8 – 1.1/6 + 2.1/4 + 4. 1/12 = 5/12

b) $$F(x)=\begin{cases} 0 \quad si \quad x<-2 \\ 1/8 \quad si \quad -2\le x<-1 \\ 7/24 \quad -1\le x<0 \\ 16/24 \quad si \quad 0\le x<2 \\ 11/12 \quad si \quad 2\le x<4 \\ 1 \quad si \quad x\ge 4 \end{cases}$$

**8) Hallar el valor de a en las siguientes distribuciones de probabilidad**

| x | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| f(x) | 0,2 | a | 3a | 0,4 |

| x | 0 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| g(x) | a | 3 a | 2 a | 3 a | a |

**En ambos casos hallar la función de distribución y los parámetros**

SOLUCIÓN:

Para la primera se tiene que: 4 a + 0,6 =1, de donde a = 0,1
Para la segunda se tiene que 10 a = 1, de donde a = 0,1

$$F(x)=\begin{cases} 0 \quad si \quad x<1 \\ 0{,}2 \quad si \quad 1\le x<2 \\ 0{,}3 \quad si \quad 2\le x<3 \\ 0{,}6 \quad si \quad 3\le x<4 \\ 1 \quad si \quad x\ge 4 \end{cases} \qquad G(x)=\begin{cases} 0 \quad si \quad x<0 \\ 0{,}1 \quad si \quad 0\le x<2 \\ 0{,}4 \quad si \quad 2\le x<3 \\ 0{,}6 \quad si \quad 3\le x<4 \\ 0{,}9 \quad si \quad 4\le x<5 \\ 1 \quad si \quad x\ge 5 \end{cases}$$

**9) Una función de densidad de una variable aleatoria es:**

$$f(x)=\begin{cases} 1/5 \quad si \quad 0<x<5 \\ 0 \quad en\ otro\ caso \end{cases}$$

**Calcula P ( 2 < X < 4 ), P ( X > 3 ), P ( X < 0 ) y representa gráficamente la función de distribución.**

SOLUCIÓN

$$F(x)=\begin{cases} 0 \quad si \quad x<0 \\ \frac{1}{5}x \quad si \quad 0\le x<5 \\ 1 \quad si \quad x\ge 5 \end{cases}$$ de donde P(2<X<4) = F(4)-F(2) = 2/5 ;

P(X>3)=1-F(3) = 2/5 ; P(X<0) = 0

6) **La función de densidad de una variable aleatoria x es**

$$f(x)=\begin{cases} kx & si \quad x\in[0 \quad 5] \\ 0 & si \quad x\notin[0 \quad 5]\end{cases}$$

**Hallar k y la función de distribución.**
**Hallar la media y la varianza de x**
**Calcular P(0 < X < k )**

SOLUCIÓN:

$$\int_0^5 Kxdx=\left[k\frac{x^2}{2}\right]_0^5=\frac{25}{2}k=1;$$ de donde k = 2/25

La media o esperanza de X es $\int_0^5 \frac{2}{25}x^2\,dx = \frac{2}{25}\left[\frac{x^3}{3}\right]_0^5=\frac{10}{3}$

7) **La función de distribución de una variable aleatoria es**

$$F(x)=\begin{cases} 0 \quad si \quad x<-1 \\ \dfrac{(x+1)^2}{4} \quad si \quad x\in[-1,1] \\ 1 \quad si \quad x\geq 1\end{cases}$$

**Calcular la función de densidad y la media**

SOLUCIÓN:

$$f(x)=\begin{cases} \dfrac{x+1}{2} & si \quad x\in[-1,1) \\ 0 \quad si & x\notin[-1,1]\end{cases}$$

La media o esperanza es: $\int_{-1}^{1} xf(x)dx=\left[\frac{x^3}{6}+\frac{x^2}{4}\right]_{-1}^{1}=1/3$

**8) La función de densidad de una variable aleatoria X es**

$$f(x)=\begin{cases} 0 \quad si \quad x<0 \\ k \quad si \quad 0<x<1 \\ k' \quad si \quad 1<x<2 \\ 0 \quad si \quad x>2\end{cases}$$

**Hallar las constantes k y k' sabiendo que P(0 < X < 1,5) = 0,7. Luego hallar la función de distribución**.

SOLUCIÓN:

Por ser f función de densidad se tiene: $\int_0^1 kdx + \int_1^2 k'dx = 1 \Rightarrow k + k' = 1$

Por otra parte y dado que P(0 < X < 1,5) = 0,7, resulta que

$\int_0^1 kdx + \int_1^{1,5} k'dx = 0,7 \Rightarrow k + 0,5k' = 0,7$ , de ambas ecuaciones se obtiene k = 0,4 y k'=0,6

La función de distribución es:

$$F(x) = \begin{cases} 0 \quad si \quad x < 0 \\ 0,4x \quad si \quad 0 \le x < 1 \\ 0,6x - 0,2 \quad si \quad 1 \le x < 2 \\ 1 \quad si \quad x \ge 2 \end{cases}$$

**9) Obtener el valor de c sabiendo que es un número real positivo, para que la función f(t) = $e^t$ / c pueda ser función de densidad de una variable aleatoria continua X, definida en el intervalo [ 0, 1 ]. Calcular la función de distribución, así como P ( X > ½) ; esperanza de X.**

SOLUCIÓN:

$$\int_0^1 \frac{e^t}{c} dt = 1 \Rightarrow \frac{e}{c} - \frac{1}{c} = 1 \Rightarrow c = e - 1$$

La función de distribución de esta variable aleatoria viene determinada por:

$$F(x) = \begin{cases} 0 \quad si \quad x < 0 \\ \dfrac{e^x - 1}{e - 1} \quad si \quad 0 \le x < 1 \\ 1 \quad si \quad x \ge 1 \end{cases}$$

El valor de P ( X > ½) = 1-F(1/2) = 0,6224 y la esperanza es 1/(e-1) = 0,5819

**10) La función de densidad de una variable aleatoria es**

$$f(x) = \begin{cases} kx + 1/2 \quad si \quad x \in [0,3] \\ 0 \quad si \quad x \notin [0,3] \end{cases}$$

**Calcular la probabilidad de los sucesos A, B, B U C, C siendo**
**A= ( X > 3/2), B = ( 1 < X < 3 ), C = ( 3/2 < X < 2 )**

SOLUCIÓN:

En primer lugar hay que hallar k, cuyo valor se obtiene del siguiente modo:

$$\int_0^3 kx+\frac{1}{2}dx=\left[\frac{kx^2}{2}+\frac{kx}{2}\right]_0^3=1\Rightarrow 6k=1\Rightarrow k=\frac{1}{6}$$

Calculemos ahora las probabilidades de A, B y C respectivamente:

P(A) = $\int_0^{\frac{3}{2}}\frac{1}{6}x+\frac{1}{2}dx=\left[\frac{x^2}{12}+\frac{x}{12}\right]_0^{\frac{3}{2}}=5/16$, del mismo modo se tiene que P(B)=5/6 y P(C) = 3/16.

P(BUC)=P(B)+P(C)-P(ByC), ahora bien, ByC es el suceso C, por tanto P(BUC)=5/6

**11) Un jugador tira al blanco. La distribución de los impactos en torno a la diana viene dada por la función de densidad f(x) = $e^{-kx}$ donde x representa la distancia del impacto a la diana. Hallar el valor de k y la función de distribución.**

SOLUCIÓN:

$$\int_0^\infty e^{-kx}dx=1\Rightarrow\left[\lim_{t\to+\infty}\frac{e^{-kt}}{-k}\right]_0^t=1\Rightarrow \text{k=1}$$

La función de distribución es $F(x)=\begin{cases}0 & si \quad x<0\\ 1-e^{-x} & si \quad x\geq 0\end{cases}$

**12) La probabilidad de tener éxito en una prueba es p = 0,6. Cuál es la probabilidad de conseguir 4 o 5 éxitos en 10 pruebas**

SOLUCIÓN:

Se trata de una distribución binomial para n=10 y p=0,4 ya que 0,6 no viene en las tablas. Por tanto la probabilidad pedida es: P(X=6)+P(X=5) siendo X="número de fracasos" El valor pedido es: 0,3122

**13) Un tirador tiene una probabilidad de acertar igual a 0,8. ¿ Cuál es la probabilidad de que en cinco disparos acierte al menos 4 ?**

SOLUCIÓN:

Se trata de una distribución binomial para n=5 y p=0,2 ya que 0,8 no viene en las tablas. Por tanto la probabilidad pedida es: P(X<2), siendo X="número de fallos". La probabilidad pedida es: 0,7373

**14) Supongamos que un sistema de 8 componentes independientes requiere para su funcionamiento que al menos 6 estén en buen estado. Si la probabilidad de funcionamiento de cada componente es 0,95, calcular la probabilidad del sistema definida por la probabilidad de que funcione. Calcular la esperanza y la desviación típica.**

SOLUCIÓN:

Se trata de una distribución binomial para n=8 y p=0,05 ya que 0,95 no viene en las tablas. Por tanto la probabilidad pedida es: P(X<3), siendo X="número de componentes en mal estado". La probabilidad pedida es: 0,9942.

La esperanza es 8.0,95= 7,6 (es decir entre 7 y 8 componentes que funcionen) ya que si tomo 8.0,05=0,4 (sería entre 0 y 1 componente que fallen).

La desviación tipica de la variable "Numero de componentes que funcionan como que fallen" será: 0,6164

**15) La probabilidad de que un equipo de futbol gane a otro equipo determinado es 1 / 4. Se juegan 6 partidos. Se pide:**

**a) La probabilidad de que gane por lo menos dos veces.**

**b) Calcular el número de partidos que tiene que jugar para que la probabilidad de ganar por lo menos una vez sea mayor que 3/5.**

SOLUCIÓN

a) Se trata de una distribución binomial para n=6 y p=0,25. Por tanto la probabilidad pedida es: P(X>1), siendo X="número de partidos ganados". La probabilidad pedida es: 0,4660

b) P(X>0)>3/5 1-P(X=0)>3/5 P(X=0)<2/5. Solución 4 partidos como mínimo ya que en las tablas el primer valor de n que hace esa probabilidad menor que 2/5 es 4.

**16) Un alumno debe realizar un examen de 8 preguntas con respuestas de sí o no. ¿ Qué probabilidad tiene de contestar correctamente por lo menos a la mitad de ellas si desconoce por completo la materia del examen** ?

SOLUCIÓN:

Se trata de una distribución binomial para n=8 y p=0,5. Por tanto la probabilidad pedida es: P(X>3), siendo X="número de aciertos". La probabilidad pedida es: 0,6367

**17) Calcular la probabilidad de que al lanzar al aire 5 veces una moneda se obtengan al menos dos caras.**

SOLUCIÓN:

Se trata de una distribución binomial para n=5 y p=0,5. Por tanto la probabilidad pedida es: P(X>1), siendo X="número de caras". La probabilidad pedida es: 0,1874

**18) Suponiendo que Z es la distribución normal tipificada, calcular**

**a) P ( z > 0,7 ) P ( 1 < z < 1,37 ) P ( | z | > 1,05 )**

SOLUCIÓN:

a) 0,2420 b) 0,0733 c) 0,2938

**18) Las alturas de 300 estudiantes se distribuyen normalmente con una media igual a 172 cm. y una desviación típica de 7 cm. ¿ Cuántos estudiantes tienen altura: mayor que 182 cm. ; menor que 163 cm. ; entre 165 y 181cm. ; igual a 172 cm.**

SOLUCIÓN:

Nos piden las siguientes valores: 300.P(X>182) ; 300.P(X<163); 300.P(165<X<181); 300.P(X=172).

Tipificando los valores de X, resulta:

| | | | | | |
|---|---|---|---|---|---|
| X | 182 | 163 | 165 | 181 | 172 |
| Z | 1,43 | -1,28 | -1 | 1,28 | 0 |

Por tanto 300.P(X>182) = 300. P(Z>1,43)=300.[1-P(Z<=1,43)]=300.(1-0,9236)=23

300.P(X<163)= 300.P(Z<-1,28)=300.[1- P(Z<1,28)]=300.(1-0,8997)=30

300.P(165<X<181)= 300.P(-1<Z<1,28)=300.(0,8997-1+0,8413)=222;

300.P(X=172)=0.

**19) Los pesos de los habitantes de una población se distribuyen normalmente con una media de 65 Kg. y una desviación típica de 6 Kg. Calcular la probabilidad de que un individuo de dicha población pese:**

**a) a lo sumo 65 Kg. b) entre 60 y 75 Kg. c) más de 75 Kg.**

SOLUCIÓN:

a) el valor tipificado de 65 es 0. Así pues la respuesta es 0,5.

b) el valor tipificado de 60 y 75 es respectivamente –0,83 y 1,67, de donde P(-0,83<Z<1,67)=0,9525-1+0,7967=0,7492

c) el valor tipificado de 75 es 1,67, entonces P(Z>1,67)=1-0,9525=0,0475

**20) Sea X una variable aleatoria normal tal que P ( X < 15 ) = 0,1003 y P ( X < 20 ) = 0,9605. Calcular:**

**a) Calcular la media y la desviación típica de x.**

**b) P ( 16,5 < X < 17,8 )**

**c) Calcular el número k tal que P( X > k ) = 0,5**

SOLUCIÓN:

a) $P(Z < \frac{15-\mu}{\sigma}) = 0{,}1003;\quad P`(Z < \frac{\mu-15}{\sigma}) = 0{,}8997;\quad \frac{\mu-15}{\sigma} = 1{,}28$

$P(Z < \frac{20-\mu}{\sigma}) = 0{,}9605;\quad \frac{20-\mu}{\sigma} = 1{,}81$, obteniéndose un sistema de dos ecuaciones con dos incógnitas, cuyas soluciones son 17,07 para la media y 1,62 para la desviación típica.

c) Los valores tipificados de 16,5 y 17,8 son respectivamente –0,35 y 0,45, por tanto el resultado pedido es 0,6736-1+0,6368=0,3104

d) $P(Z > \frac{k-17{,}07}{1{,}62}) = 0{,}5$; de donde k=17,07

21) La talla de los hombres en edad militar en España sigue una distribución normal de media 169 cm. y desviación típica 6 cm. Si no se admiten para el servicio militar los individuos de talla inferior a 160 cm. ¿ Qué proporción se rechaza ?

22) Sea A un suceso de probabilidad 0,4. Suponiendo que se hacen 900 pruebas del experimento, calcular la probabilidad de que A se verifique exactamente 380 veces ?

23) Hallar la probabilidad de obtener tantas caras como cruces en 100 lanzamientos de una moneda.

24) Entre 2000 estudiantes la media del peso resultó ser 70 kg. con una desviación tíica de 8,5 Kg. Determinar el peso mínimo del conjunto formado por los 200 estudiantes más pesados.

25) Hallar la probabilidad de que entre 100000 cifras al azar la cifra 6 salga menos de 9971 veces.

26) La media de una variable aleatoria X normal es el quíntuplo de la desviación típica. Sabiendo que P(X < 6) = 0,8413, calcular la media y la desviación típica.

# TEMA X: ESTIMACIÓN PUNTUAL

**Población**

Es un conjunto homogéneo de elementos de los que se quiere estudiar alguna característica determinada, determinada por una variable aleatoria X. Sus elementos se llaman *individuos* y el número de éstos es el *tamaño* de la población que nosostros representaremos por **N**.

Generalmente no es posible estudiar todos los elementos de la población porque no es viable física o económicamente, porque la evolución temporal de la característica es demasiado rápida y no da tliempo a elaborar el censo, ,porque el ensayo requerido es destructivo (supongamos que queremos saber cuanta presion soporta un huevo antes de reventar), etc.

Para evitar este problema tomamos una **muestra** de la población.

Así pues, una **muestra** es un conjunto representativo de los elementos de una población. El tamaño de la muestra lo representaremos por **n**.

**Formas de elección de la muestra**:

Entre otras tenemos las siguiente:

***Muestra aleatoria simple***: Cuando todos los invidivuos de la población tienen la misma probabilidad de salir elegidos en la muestra. Es decir que sobre una lista ordenada, o no, de los elementos de la población, tomamos al azar los elegidos de la muestra.

Un método para generar números aleatorios de forma *limpia* es utilizando la tecla RAN de la calculadora, que nos genera números aleatorios entre 0 y 1, por lo que si quiero generar números enteros aleatorios entre p y q, ambos inclusive, se utilizará la siguiente fórmula:

$$\mathbf{Ent\ [\ (q-p+1).RAN] + p}$$

***Muestra aleatoria sistemática:***

Es la que consiste en obtener el primer individuo de la muestra como en el caso aleatorio simple y después tomar los siguientes a "saltos" de magnitud igual dentro de la lista, de modo que se complete un ciclo de la lista. El primer elemento se llama **origen** y la magnitud del salto es el llamado **coeficiente de elevación**. Este coeficiente se obtiene dividiendo el tamaño de la población N, entre el tamaño de la muestra, n.

***Muestra estratificada:***

Es una técnica muy susada cuando la población puede no ser tan homogénea como se desearía y existen distintos grupos (llamados estratos) dentro de ella en los que el comportamiento de la variable en estudio puede diferir bastante de unos a otros.

Lo normal a la hora de obtener muestras estratificadas es que en la muestra se respeten las proporciones poblacionales de cada estrato. Así, si $N_1$, $N_2$….$N_k$ son los tamaños de los estratos de la población, y n es el tamaño deseado de la muestra, el número de elementos de los estratos en la muestra i, debe ser:

$$\mathbf{n_i = (N_i /N).\ n}$$

**Estimación puntual**

El objetivo de la estimación puntual es obtener un valor numérico para cierto parámetro de la población a estudio.

Sea una población en la que estamos estudiando una característica determinada por la variable aleatoria X, con función de distribución conocida o no F(x;θ) conocida salvo por los valores de uno o más parámetros θ. Sea una muestra caracterizada por las variables X1, X2, …, Xn.

Se llama estimador a una función T(X1, X2, …, Xn) cuyo valor numérico, $\hat{\theta}$, llamado parámetro muestral, obtenido a partir de los datos muestrales permite estimar el parámetro poblacional θ.

Puesto que T es función de variables aleatorias, es a su vez una variable aleatoria, por lo que tiene perfecto sentido hablar de su esperanza y su desviación típica.

Es importante recordar que si X e Y son dos variables aleatorias, se tiene que:

$$E[X+Y] = E[x]+E[Y] \quad ; \quad E[k.X] = k.E[X]$$

Mientras que para la varianza se obtiene:

$$Var[X+Y] = Var[x]+VAr[Y] \quad ; \quad Var[k.X] = k^2.E[X]$$

Para medir la bondad de un estimador, esto es para determinar si el estimador es mejor o peor, se utilizan los siguientes conceptos

Un estimador se dice **insesgado o centrado** si la esperanza de dicho estimador coincide con el parámetro estimado. Es decir $E(T)=\theta$. En caso contrario el estimador se dice sesgado. En principio es deseable que el estimador sea lo menos sesgado posible como es lógico.

Se llama **sesgo** del estimador al valor $E(T)-\theta$ y dicho valor representa lo que por término medio se desvía el estimador del valor teórico.

Un ejemplo de estimador centrado es la media muestral

$T=\sum_{i=1}^{n}\frac{1}{n}X_i$, sin embargo la varianza muestral $S^2=\sum_{i=1}^{n}\frac{1}{n}(X_i-\mu)^2$ no es un estimador centrado de la varianza poblacional, ya que su esperanza no coincide con $\sigma^2$, sino que es igual a $\frac{n-1}{n}\sigma^2$. Para buscar un estimador insesgado o centrado de la varianza se utiliza la llamada cuasivarianza muestral que utilizaremos más adelante y viene dada por:

$S_{n-1}{}^2=\sum_{i=1}^{n}\frac{1}{n-1}(X_i-\mu)^2=\frac{n}{n-1}S^2$. Este valor se obtiene en la calculadora.

Por ejemplo consideremos en una población donde se está estudiando una variable X, una muestra de tamaño 4, dada por los siguientes valores de la variable $X_1$, $X_2$, $X_3$ y $X_4$

Y consideremos dos estimadores para la media aritmética de X, μ

T = $\frac{X_1+2X_2-X_3+X_4}{4}$ y T' = $\frac{X_1+X_4}{2}$, resulta fácil comprobar que ambos estimadores son centrados para el parametro que estamos estudiando, en este caso la media μ. Pero ¿cómo podemos saber cuál de ellos es el mejor?

Esto vendrá determinado por los siguientes conceptos:

Error cuadrático medio o ECM de un estimador T, es la esperanza de los cuadrados de las desvliaciones del mismo con respecto al parámetro que se está estimando θ.

Es decir:

$$\mathbf{ECM\ (T) = \ (E(T)- \theta)^2 + Var\ (T)}$$

Si el estimador es centrado resulta que ECM(T) = Var(T)

Diremos que el estimador T es más eficiente que el estimador T' cuando ECM(T) < ECM(T')

Se llama eficiencia de un estimador T, a **Efic (T) = 1/ECM(T)**.

Dados dos estimadores será mejor el de mayor eficiencia.

Un requisito mínimo que debe cumplir un etimador para que sea aceptable es que, a medida que aumente el tamaño de la muestra, el estimador se aproxime al parámetro estimado. Esto se puede expersar en términos de un nuevo concepto denominado consistencia, que definiremos matemáticamente a continuación. Un estimador se dirá consistentes si, a medida que aumenta el tamaño muestral, el error cuadrático medio del estimador tiene a cero, es decir:

$$\lim_{n \to \infty} ECM(T) = 0$$

**Ejercicios propuestos:**

1) En una población normal con media μ desconocida y desviación típica 5, tomamos muestras aleatorias de tamaño 3, $X_1$, $X_2$, $X_3$, considerando como estimadores de la media de la población las siguientes expersiones:

$T = 2X_2 – X_1$ y $T' = \dfrac{X_1 + X_2 + X_3}{3}$

¿Cuál de estos dos estimadores es mejor?

2) Sea $X_1$, $X_2$, $X_3$, $X_4$, $X_5$ una muestra aleatoria simiple de una población de media μ y varianza $\sigma^2$ **.** Consideremos como estimadores de la media de la población μ, las siguientes funciones:

$T_1 = \dfrac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$ Y $T_2 = \dfrac{X_1 + X_2 + 2X_3 + X_4 + X_5}{6}$

¿son insesgados? ¿Cuál tiene menor varianza? ¿Cuál es más eficiente?

# TEMA XI: ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Consiste en buscar un intervalo numérico donde es muy probable que se encuentre el parámetro de la población que queremos estimar.
Siempre intentaremos dar un intervalo lo mas pequeño posible.
Por ejemplo, si queremos dar un intervalo para la estatura media de los alumnos de un Instituto y decimos que la media está en el intervalo (149, 180), evidentemente tendríamos una probabilidad casi del 100% de que en efecto la media se encuentre ahí, pero el objetivo es dar precisamente un intervalo de radio inferior. Por ejemplo una buena estimación para una media de la población que fuese igual a 41, sería afirmar que se encuentra en el intervalo (41.2,41.5).

En definitiva, queremos construir un intervalo de forma que la probabilidad de que el parámetro a estimar esté dentro de este intervalo sea previamente conocida. A esta probabilidad la denominaremos **Nivel de Confianza** y la denotaremos por 1-α, siendo ésta generalmente una probabilidad muy pequeña (0,01 ó 0,05). Entonces se pretende construir un intervalo $(\hat{a},\hat{b})$ de forma que $P(\vartheta \in (\hat{a},\hat{b})) = 1-\alpha$, siendo θ el parámetro de la población a estimar.
Para ello se utiliza el método del pivote que consiste en buscar una función de la muestra $\vec{X} = (X_1, X_2,..., X_n)$ y del parámetro θ, que llamaremos estadístico pivote $g(\vec{X},\vartheta)$ de forma que siga una distribución conocida, de forma que dicha distribución no dependa del parámetro. Al tener una distribución conocida (normal, chi-cuadrado...etc) buscamos el intervalo de probabilidad de forma que $P(g(\vec{X},\theta)) = 1-\alpha$, considerandolo centrado en la distribución, esto es que a la izquierda del intervalo y a la derecha queda un valor de probabilidad α/2.
Veamos un caso:

**Estimación de intervalo de confianza de la media μ de una población que sigue una distribución normal, conociendo la desviación típica y** tomando como estimador la media muestral de una muestra $(X_1, X_2,..., X_n)$, que denotaremos por $\overline{X}_n = \frac{X_1 + ... + X_n}{n}$. En primer lugar hemos de advertir que la distribución en la muestra de la media muestral, es una variable aleatoria normal, de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. La demostración de esto es fácil de ver ya que calculando la media y varianza de $\overline{X}_n$ se tiene: $E[\overline{X}_n] = \frac{\sum_{i=1}^{n} E[X_i]}{n} = \frac{n\mu}{n} = \mu$, mientras que para la varianza si tiene:

$$Var[\overline{X}_n] = \frac{\sum_{i=1}^{n} Var[X_i]}{n^2} = \frac{n\sigma}{n^2} = \frac{\sigma}{n}$$, de donde la desviación típica es como dijimos $\frac{\sigma}{\sqrt{n}}$.

En consecuencia tenemos que $\overline{X}_n$ es N(μ, $\frac{\sigma}{\sqrt{n}}$), de donde se obtiene que:

$\dfrac{\overline{X}_n - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ es N(0,1). Este es por tanto el estadístico pivote que necesitamos pues involucra a los elementos del problema en juego y sigue una distribución conocida (normal tipificada).

Como queremos que la probabilidad de que dicho estadístico se encuentre en un intervalo centrado respecto a la media de la distribución de forma que deje $\alpha/2$ a ambos lados de dicho intervalo, consideraremos $z_{\alpha/2}$ y - $z_{\alpha/2}$ como los valores extremos de dicho intervalo, resultando entonces:

$P(-z_{\alpha/2} < \dfrac{\overline{X}_n - \mu}{\dfrac{\sigma}{\sqrt{n}}} < z_{\alpha/2}) = 1 - \alpha$, despejando el valor del parámetro que es el que queremos estimar, μ, resulta : $P(\overline{X}_n - \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \overline{X}_n + \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}) = 1-\alpha$.

Hemos encontrado el intervalo que estima la media de la población con un nivel de confianza $1-\alpha$

Nota.- Para poblaciones finitas, se suele tomarse $\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{n_p - n}{n_p - 1}}$ , en lugar de $\dfrac{\sigma}{\sqrt{n}}$ donde $n_p$ es el tamaño de la población, pero nosotros no lo vamos a considerar en nuestro estudio.

**Distribuciones asociadas a la normal**

**Chi-Cuadrado $\chi^2$**

Dadas n variables aleatorias N(0,1) e independientes entre sí, la variable aleatoria, dada por $X_n^2 = X_1^2 + X_2^2 + \ldots + X_n^2$, sigue una distribución que se denomina chi-cuadrado con n grados de libertad y se representa por $\chi^2$

¿Qué tiene de particular esta distribución?. La respuesta es en condiciones de normalidad de una población X, y para una muestra de n individuos, resulta que el estadístico pivote dado por $\dfrac{(n-1)S_{n-1}^2}{\sigma^2}$ sigue precisamente una distribución $\chi^2$ con n-1 grados de libertad, y teniendo en cuenta que dicha variable involucra a la cuasivarianza muestral y a la varianza poblacional, es un perfecto estadístico pivote para **obtener un intervalo de confianza de la varianza muestral, desconociendo la media en una distribución normal**. Para ello se procede como en el caso anterior teniendo en cuenta que la distribución ahora a considerar no es normal sino que es una Chi-cuadrado.

El intervalo que se obtiene es:

$$\left( \frac{(n-1)S_{n-1}^2}{\chi_{\alpha/2}^2} \quad \frac{(n-1)S_{n-1}^2}{\chi_{1-\alpha/2}^2} \right)$$

**t de Student**

Dadas n variables aleatorias con distribución N(0,1) e independientes y dada otra variable independiente de las anteriores con distribución N(0,1), la variable aleatoria determinada por $t_n = \dfrac{Y}{\sqrt{\dfrac{\sum_{i=1}^{n} X_i^2}{n}}}$ sigue una distribución llamada t de Student con n grados de libertad. A nosotros nos interesa puesto que en condiciones de normalidad se tiene que la variable dada por $\dfrac{\overline{X}_n - \mu}{\dfrac{S_{n-1}}{\sqrt{n}}}$ es una distribución t de Student con n-1 grados de libertad. Este pivote estadístico sirve para dar un **intervalo de confianza de la media desconociendo la desviación típica**. Procediendo como en los casos anteriores y aplicando la distribución t de Student, se tiene que dicho intervalo viene determinado por:

$$\left( \overline{X}_n - \frac{S_{n-1}}{\sqrt{n}} t_{n-1,\alpha/2} \quad , \quad \overline{X}_n + \frac{S_{n-1}}{\sqrt{n}} t_{n-1,\alpha/2} \right)$$

Si el tamaño muestral n es tan grande que el valor de la t no se encuentra en la tabla, se suele aproximar por el correspondiente, al mismo nivel de confianza, a la distribución normal.

**Intervalo de confianza para una proporción**

Supongamos una población donde los individuos se pueden clasificar según una determinada característica.
Queremos calcular un intervalo de confianza para la proporción de individuos que poseen dicha característica.
Elegimos una muestra aleatoria de tamaño n. Sea X la variable aleatoria que determina el número de individuos de la muestra que cumplen dicha característica. Definimos como proporción muestral, $\widehat{p} = \dfrac{X}{n}$. Es obvio que X es una binomial de parámetros n y p. Sabemos que para valores de n grandes y p próximos a 0,5, una binomial se puede aproximar por una normal, de parámetros np y $\sqrt{np(1-p)}$ , de donde podemos inferir que $\dfrac{X - np}{\sqrt{np(1-p)}}$ es una N(0,1), dividiendo por n, el numerador y el denominador, se tiene $\dfrac{\widehat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ es N(0,1), que será nuestro pivote estadístico, de donde se obtiene el siguiente intervalo

$$\left( \widehat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \quad , \quad \widehat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right)$$

Obsérvese que este intervalo no se puede calcular al desconocer el valor de la proporción poblacional p. Esto podemos resolverlo utilizando su estimador, quedando el intervalo de la forma:

$$\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \quad , \quad \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\hat{p})}{n}}\right)$$

Otra posibilidad para el calculo del intervalo, sin utilizar el estimador anterior, es utilizar el máximo de los valores posibles que puede tomar el producto p(1-p), que como se puede demostrar fácilmente es menor que 1/4. La demostración de esto puede ser la siguiente: Puesto que p(1-p) = p-$p^2$. y la derivada de esta función respecto de p es 1-2p se anula en p=1/2, donde la función presenta un máximo, por tanto el valor máximo se da para p=1/2, asi pues el valor máximo de la expresión es 1/4.
quedando pues el intervalo de la siguiente manera:

$$\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{1}{4n}} \quad , \quad \widehat{p} + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right)$$

**Selección del tamaño muestral**

Si queremos aumentar la precisión de un intervalo de confianza, lo que hacemos es reducir su longitud para lo cual debemos aumentar el tamaño muestral.
En situaciones podemos controlar el tamaño muestral en el sentido de que es posible elegir un valor de n de forma que el error cometido al estimar el parámetro sea menor que un valor concreto y especificado llamado margen de error.
Se trata en todo caso de acotar el radio del intervalo de confianza correspondiente y obtener el valor de n.

## EJERCICIOS PROPUESTOS

1)*Un nadador obtiene los siguientes tiempos, en minutos, en 10 pruebas cronometradas por su entrenador: 41,48 42,34 41,95 41,86 41,60 42,04 41,81 42,18 41,72 42,26.*
*Obtener un intervalo de confianza para la marca promedio de esta prueba con un 95% de confianza, suponiendo que se conoce por otras pruebas que la desviación típica para este nadador es de 0,3 minutos. Si el entrenador quiere obtener un error en la estimación de la media de este nadador inferior a tres segundos, ¿cuántas pruebas debería cronometrar?*

2*) La puntuación promedio de una muestra de 20 jueces de gimnasia rítmica, elegidos al azar, para una misma prueba presentó una media de 9,8525 y una cuasi desviación típica muestral de 0,0965. Calcular un intervalo de confianza con un 95% para la nota media.*
*(Se sobreentiende que la puntuación de la prueba sigue una distribución normal)*

3) *Un entrenador de fútbol está interesado en estimar, con un 99% de confianza, la fuerza máxima de los músculos cuadriceps de los futbolistas. Admitiendo que dicha fuerza sigue una distribución normal, selecciona al azar una muestra de 25 futbolistas, para la que obtuvo una media de 85 Nw y una cuasivarianza de 144. Determinar un intervalo de confianza para la media y otro para la varianza de la fuerza máxima de estos músculos.*

4) *En una encuesta hecha por los alumnos y alumnas de un Instituto a un total de 100 votantes elegidos al azar en su Ayuntamiento, se indica que el 55% volvería a votar por el alcalde actual. Calcular un intervalo de confianza al 99% e otro al 99,73% para la proporción de votantes favorables al alcalde actual.*

5) *¿Cuáles deben ser los tamaños muestrales en el sondeo del problema anterior para tener, con los mismos niveles de confianza, la certeza de que el alcalde actual salga reelegido por mayoría absoluta, en el caso de arrojar la encuesta los mismos resultados?*

6) *En una encuesta a 360 alumnos de un centro, elegidos al azar, resultaron 190 a favor de la política del actual equipo directivo. ¿Cuál es el intervalo de confianza, con nivel del 95%, para la proporción de alumnos que apoyan a esta dirección?*

7) *Se lanza una moneda 100 veces y se obtienen 62 cruces. ¿Cuál es el intervalo de confianza para la proporción de cruces con un 99% de nivel de confianza?*

8) *Para estimar el número de ranas que hay en un estanque procedemos a pescar cierta cantidad, 30, y las marcamos con un anillo, devolviéndolas al estanque. Transcurridos unos días volvemos a pescar otro montón y observamos qué proporción están marcadas con la anilla. Es esta última pesca obtenemos 100 ranas de las que 7 están marcadas. Calcular un intervalo al 99% de confianza para la proporción de ranas marcadas.*

9) *Calcula un intervalo de confianza, con un 90%, para el número total, N, de ranas del estanque del problema anterior, teniendo en cuenta que la proporción de ranas marcadas es p= 30/N*

10) *De una muestra elegida al azar de 10 alumnos de la clase, se obtuvieron los siguientes datos para el peso (en Kg) y la estatura (en cm.)*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Peso* | *74* | *79* | *85* | *49* | *83* | *78* | *74* | *54* | *63* | *68* |
| *Estatura* | *176* | *178* | *180* | *165* | *182* | *177* | *179* | *165* | *172* | *170* |

*Calcular, suponiendo que las variables peso y estatura se adecúan a una distribución normal, un intervalo de confianza para cada variable, con un nivel de confianza del 95%, tanto para las medias como para las varianzas.*

# TEMA XII: CONTRASTE DE HIPÓTESIS

Es otra técnica de inferencia que sirve para corroborar o rechazar alguna información básica sobre la población a estudio, analizando hasta qué punto puede mantenerse una afirmación sobre los parámetros de la población. La esencia de probar estadísticamente una hipótesis es decidir si la afirmación está apoyada por la evidencia experimental, comprobar si lo que afirmamos está en sintonía con lo que nos proporciona la muestra.

En la presentación general de un contraste de hipóteis se trata de poner a prueba una hipótesis, $\mathbf{H_0}$, denominada **nula**, frente a otra hipótesis, $\mathbf{H_1}$, denominada **alternativa**. El modo de contrastar será decidir si debemos aceptar o rechazar la hipótesis propuesta con unos márgenes de error previamente fijados. Si los valores muestrales difieren mucho de los teóricos, que se obtendrían de ser $H_0$ cierta, rechazamos la hipótesis, diciendo que las diferencias son **estadísticamente significativas**.

Así pues definimos:

**Hipótesis estadística**: cualquier conjetura sobre una característica de interés de una población.

**Hipótesis nula ($\mathbf{H_0}$):** es la hipótesis que se contrasta y, salvo que los datos nos demuestren su falsedad, la mantendremos como verdadera. Rechazar esta hipóteis significa asumir una complementaria llamada **hipótesis alternativa ($\mathbf{H_1}$)**, que representa una negación de la nula.

Nosotros vamos a contrastar valores paramétricos, por lo cual la hipótesis nula siempre será un valor numérico correspondiente al parámetro a estimar, es decir $H_0 = a$; mientras que la alternativa puede ser:

Bilateral $H_1 \neq a$

Unilateral $H_1 < a$ (por la izquierda) $H_1 > a$ (por la derecha)

Las hipótesis nulas y alternativas no tienen la misma importancia. La hipótesis nula se mantendrá como cierta hasta que se demuestre lo contrario.

Como el hecho de realizar un contraste de hipótesis conlleva una toma de decisiones a favor de una u otra hipótesis, es evidente que esto implica un margen de error. El error cometido al rechazar la hipótesis nula cuando en realidad es cierta, se llama **error de tipo I.** el error cometido al no rechazar la hipótesis nula cuando es falsa se llama **error de tipo II.**

Estos errores se miden por la probabilidad de ser cometidos. Así llamaremos $P(Error\ tipoI) = \alpha \quad P(Error\ tipoII) = \beta$

Como afirmamos que en el contraste de hipótesis se debe conceder mayor confianza a la hipótesis nula, rechazándola solamente si la evidencia en contra es muy fuerte, sse contruirá el contraste fijando de antemano el error de tipo I, llamado también **nivel de significación** del contraste estadística. A **potencia** del contraste se define como la probabilidad de rechazar la hipotesis nula cuando esta es falsa, es decir $1-\beta$

## Pasos para la construcción de un contraste de hipótesis

La forma de resolver el contraste (decantarse por una de las hipótesis) es tomando una muestra y, a partir de ella, obtener un valor numérico para el que se usará una función de la muestra y del parámetro a contrastar llamado **estadístico**. Se llama

**estadístico pivote** cuando su distribución muestral es conocida y no depende de la población.

Teniendo en cuenta el error de tipo I fijado, nivel de significación del contraste, se divide la recta real en una **región de aceptación** y en otra de **rechazo** de la hipótesis nula. Finalmente, observando el valor del estadístico, nos decantaremos por rechazar o no la hipótesis nula si este valor cae en la región de rechazo o en la de aceptación. Por este motivo el estadístico pivote usado para los contrastes también recibe el nombre de **medida de discrepancia**.

**1º paso**: especificar, sin ambigüedad, las hipótesis nula y alternativa.
**2º paso**: fijar el nivel de signirficación $\alpha$ o probabilidad del error de tipo I como un valor pequeño: normalmente se usa 0.05, 0.01 o 0.1.
**3º paso**: elegir el estadístico de contraste o medida de discrepancia
**4º paso**: determinar las regiones de aceptación y rechazo.
**5º paso**: tomar una muestra de la población a contrastar y calcular el valor del estadístico de contraste, elegido en el paso 3, para esta muestra concreta.
**6º paso**: tomar la decisión de rechazar o no la hipótesis nula, en función de que el valor del estadístico valorado enla muestra observada se encuentre en la región de rechazo o de aceptación.

**Contrastes de hipótesis paramétricos clásicos**

Parten de la bases de que se conoce la distribución de la población a estudio, salvo el valor de determinados parametros desconocidos. Analizaremos los contrastes para la media y la varianza en poblaciones normales, así para la proporción en el caso de una binomial. Las posibilidades a estudiar son:

$H_0$: $\theta = \theta_0$ (contraste simple) frente a las alternativas:
$H_1$: $\theta \neq \theta_0$ (contraste bilateral o de dos colas)
$H_1$: $\theta < \theta_0$ (contraste unilateral de cola izquierda)
$H_1$: $\theta > \theta_0$ (contraste unilateral de cola derecha)

La hipótesis nula $H_0$ de un contraste siempre se toma simple.

**Ejemplo de elecciónes de hipótesis:**

Supongamos que al responsable de Sanidad le llega un informe que apunta a que determinado producto de consumo contiene una cantidad de un elemento cancerígeno superior a la tolerada. Si ello es cierto, o si no existe la completa seguridad de que el producto es inocuo, debe retirarlo inmeaitamente del mercado.

Para comprobarlo decide hacer un contraste. Sea p la cantidad del elemento cancefígenos y $p_0$ la máxima cantidad tolerada. Las hipótesis del contraste deben ser en este caso:
$H_0 : p = p_0$ y $H_1$: $p < p_0$

De esta manera, si el resultado del contraste es aceptar $H_0$ no hay evidencia de que la cantidad sea tolerable y se debe retirar el productor del mercado. Si, por el contrario, el resultado es rachazar $H_0$ **tenemos suficiente evidencia** para garantizar que la cantidad de producto está por debajo de los límites y admitir su consumo.

Si el contraste se hubiese hecho al revés, es decir: $H_0 : p = p_0$ y $H_1: p > p_0$

Rechazar $H_0$ significaría que el producto es inaceptable para el consumo y deberíamos retirarlo, pero, aceptar la hipótesis, sólo concluye que no hay evidencia suficiente para asegurar que el producto sea cancerífeno, lo que de ningún modo quiere decir que no lo sea. Ante la duda, debería retirarse el lproducto. Este contraste, mal construido, siempre conduce a la decisión de retirar el producto o a un cierto riesgo de que admitamos el consumo de un producto cancerígeno.

**Contraste para una media de una distribución normal:**

Supongamos que tenemos una muestra aleatoria $X_1, \ldots X_n$, procedente de una población normal, X, con media desconocida $\mu$ y deseamos contrastar la hipótesis nula $H_0$: $\mu = \mu_0$

Si esta hipótesis es cierta (como ya vimos en los intervalos de confianza) se cumple que:

**$d_1$**$=\dfrac{\overline{X}_n - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ es N(0,1). Este es por tanto el estadístico pivote que necesitamos pues involucra a los elementos del problema en juego y sigue una distribución conocida (normal tipificada), si conocemos la desviación típica

**$d_2$**$=\dfrac{\overline{X}_n - \mu}{\dfrac{S_{n-1}}{\sqrt{n}}}$ es t-Student con n-1 grados de libertad. Este es por tanto el estadístico pivote que necesitamos pues involucra a los elementos del problema en juego y sigue una distribución conocida (t-Student), si desconocemos la desviación típica.

$d_1$ y $d_2$ mide la discrepancia observada a través de la muestra con el valor dado por la hipótesis nula, presentándose las siguientes posibilidades:

| Contraste | Hipótesis nula | Hipótesis alternativa | Region de aceptación | |
|---|---|---|---|---|
| | | | $\sigma$ conocida | $\sigma$ desconocida |
| Bilateral | $H_0$: $\mu = \mu_0$ | $H_1$: $\mu \neq \mu_0$ | $(-z_{\alpha/2}, z_{\alpha/2})$ | $(-t_{\alpha/2}, t_{\alpha/2})$ |
| Unilateral izquierda | $H_0$: $\mu = \mu_0$ | $H_1$: $\mu < \mu_0$ | $(-z_{\alpha}, +\infty)$ | $(-t_{\alpha}, +\infty)$ |
| Unilateral derecha | $H_0$: $\mu = \mu_0$ | $H_1$: $\mu > \mu_0$ | $(-\infty, z_{\alpha})$ | $(-\infty, t_{\alpha})$ |

**Contraste para la varianza en poblaciones normales**

Supongamos que tenemos una muestra aleatoria $X_1, \ldots X_n$, procedente de una población normal, X, con varianza desconocida $\sigma^2$ y deseamos contrastar la hipótesis nula $H_0$: $\sigma^2 = \sigma_0^2$

Si esta hipótesis es cierta (como ya vimos en los intervalos de confianza) se cumple que:

$d_3 = \dfrac{(n-1)S_{n-1}^2}{\sigma^2}$ sigue una Chi-cuadrado con n-1 grados de libertad.

Los posibles contrastes son:

| Contraste | Hipótesis nula | Hipótesis alternativa | Region de aceptación |
|---|---|---|---|
| Bilateral | $H_0$: $\sigma^2 = \sigma_0^2$ | $H_1$: $\sigma^2 \neq \sigma_0^2$ | $\left(\chi^2_{1-\alpha/2}, \chi^2_{\alpha/2}\right)$ |
| Unilateral izquierda | $H_0$: $\sigma^2 = \sigma_0^2$ | $H_1$: $\sigma^2 < \sigma_0^2$ | $(\chi^2_{1-\alpha}, +\infty)$ |
| Unilateral derecha | $H_0$: $\sigma^2 = \sigma_0^2$ | $H_1$: $\sigma^2 > \sigma_0^2$ | $(-\infty, \chi^2_{\alpha})$ |

**Contraste de hipótesis para una proporción**

Supongamos que los individuos de la población a estudio pueden ser clasificados según un atributo que pueden poseer o no y sea p la proporción de individuos de la pobalción que posee dicho atributo. Elegida una muestra aleatoria de la población queremos construir un contraste para comparar el verdadero valor de p con un valor hipotético $p_0$

Sea p' la frecuencia relativa de los elementos de la muetras que tienen el atributo considerado. Si es cierta la hipótesis nula $p=p_0$ se verifica que

$D_4 = \dfrac{\hat{p} - p_0}{\sqrt{np_0(1-p_0)}}$ es N(0,1), que será nuestro pivote estadístico, obteniéndose los posibles contrastes:

| Contraste | Hipótesis nula | Hipótesis alternativa | Región de aceptación |
|---|---|---|---|
| Bilateral | $H_0$: $p=p_0$ | $H_1$: $p \neq p_0$ | $(-z_{\alpha/2}, z_{\alpha/2})$ |
| Unilateral izquierda | $H_0$: $p=p_0$ | $H_1$: $p < p_0$ | $(-z_{\alpha}, +\infty)$ |
| Unilateral derecha | $H_0$: $p=p_0$ | $H_1$: $p > p_0$ | $(-\infty, z_{\alpha})$ |

## EJERCICIOS PROPUESTOS

1) *Un investigador quiere contrastar si el peso medio de ciertas hortalizas está en los 1,90 Kg. que le aseguran en el mercado. Suponiendo que este peso se distribuye de forma normal con desviación típica de 100 g, selecciona al azar 10 hortalizas observando los siguientes pesos (en Kg): 1,92 1,98 1,87 2,05 2,02 1,87 1,92 1,95 1,87 2,04. Contrastar, con un nivel de significación de 0,05, que stos datos se adecuan a una distribución con media 1,9 Kg.*

2) *Con los mismos datos del ejercicio anterior, contrastar que la desviación típica del peso de las hortalizas es de 100 gramos.*

3) *Una empresa de teléfonos asegura que por termino medio realiza una instalación estándar en una casa en menos de 15 días con una desviación de dos días. Se seleccionan un total de 20 instalaciones realizadas por dicha empresa, resultando un tiempo medio de 14,2 días. Contrastar con un nivel de significación de 0,05 que el tiempo medio de cada instalación es inferior a los 15 días.*

4) *El encargado de personal de una joyería sospecha que el peso de los torques de oro que fabrica un determinado trabajador fluctúa más de lo aconsejable, que corresponde a una desviación típica de 0,9 gramos. Para contrastar sus sospechas, observa 10 días elegidos al azar a dicho trabajador, obteniendo los siguientes pesos: 10,1 9,4 8,2 11,5 10,2 9,8 11,1 10,4 9,8 y 10,2. ¿Tiene apoyo la sospecha del encargado con un nivel de significación de 0,05?*

5) *Una máquina produce balas de calibre 9 mm, las fabrica con una desviación típica de 0,007. Para saber si el proceso funciona correctamente se taman muestras de 5 balas cada dos horas. Encontrar las condiciones para contrastar, con un nivel de significación de 0,05, que la desviación típica es menor de 0,007 y la media es de 9 mm. (lo que equivale, en terminos de control de calidad, a que la producción esté bajo control)*

6) *Un hospital sostiene que el número de infectados en intervenciones quirúrgicas en sus quirófanos no sobrepasa el 12%. Se realiza un control sobre 300 enfermos intervenidos en dicho hospital de los cuales 45 si fueron infectados. ¿Es coherente la afirmación del hospital con 1% de significación?*

7) *Un proveedor de planchas para la construcción de embarcaciones suministra material a un astillero asegurando que la resistencia a tracción de este material cumple la normativa de calidad, que establece una media de 120 Mpa (Megapascales) con una desviación típica de 3,7 Mpa. Se toma una muestra al azar de 9 planchas de este proveedor y se someten a una prueba de resistencia resultando los siguientes valores: 115, 120, 105, 130, 114, 118, 109, 107, 130.*
*Contrástense las hipótesis oportunas con un nivel de confianza del 99%*

8) *Un investigador asegura tener encontrado un medicamenteo que reduce el nivel de colesterol total. Aplicado a 9 pacientes se obtuvieron los siguientes resultados, medido en mg/100ml, para antes y después de ser aplicado el medicamento:*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Antes | 215 | 220 | 205 | 230 | 214 | 218 | 209 | 207 | 230 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Desp. | 218 | 217 | 209 | 217 | 221 | 224 | 219 | 221 | 207 |

*Contrastar si las diferencias son significativas con un nivel de confianza del 99%*

9*) El responsable de una campaña electoral piensa que su candidato está en desventaja frente a su rival político, por lo que decide hacer una encuesta a 1500 electores, resultando que 720 votarían a su candidato y el resto al otro. ¿Existen razones para pensar que este candidato está en desventaja frente a su oponente con un nivel de significación del 1%?*

10) *Queremos contrastar si el porcentaje de desempleo en la comarca de Ferrol es mayor que en la del resto de Galicia que están en el 15%. Tomando una muestra al azar de 500 personas de dicha comarca se obtuvieron 95 desempleados. ¿Existen razones para sospechar que hay diferencia entre esta comarca y el resto de Galicia?*

11) *Un profesor de matemáticas quiere demostrar que la nota media de selectividad del alumnado de su materia es superior a la media gallega, que estaba en 5,5 en la última prueba. Elige una muestra al azar de 25 alumnos presentados a selectividad en la última convocatoria resultando una media de 5,95 con una cuasi desviación muestral de 0,92. Verificar la afirmación de este profesor conun nivel de confianza de un 95%*

12) *Una maternidad está interesada en contrastar si la estatura media de los niños al nacer es de 50 cm, fijando como criterio de aceptación de esta afirmación que la media muestral se desvía de los 50 cm en menos de 1,5 cm. Suponiendo que se toma como desviación típica poblacional de la estatura 2,5 cm y una muestra de tamaño 10, calcular la probabilidad de cometer un error de tipo I.*

13) *Con los datos del problema anterior y suponiendo que se quiere rechazar la hipótesis nula si la media fuese de 52 cm. ¿Cuál será la probabilidad de cometer un error de tipo II?*