Matemáticas aplicadas a las ciencias sociales
Estadística y Probabilidad
1º de bachillerato

TEMA 1: ESTADÍSTICA DESCRIPTIVA

1) LA ESTADÍSTICA: ASPECTOS GENERALES.-

La E**stadística Descriptiva** tiene por objeto la recogida, recopilación y reducción de datos, su organización en tablas y gráficos y el cálculo de unos valores que representen al conjunto de datos.

La estadística estudia las **propiedades o cualidades** que observamos en los elementos de un colectivo, que llamamos **población**. Si la población fuese muy grande tomaríamos una parte representativa de la población, que se conoce con el nombre de **muestra.**

- **Población:** Todos los elementos susceptibles de ser estudiados.
- **Muestra:** Una parte de la población.
- **Carácter Estadístico:** Propiedad que permite clasificar a los individuos de una población, por ejemplo: el color de los ojos, la altura, la edad, la profesión….

Se distinguen dos tipos de caracteres estadísticos: **Cualitativos y Cuantitativos**.

Los caracteres estadísticos **cualitativos** son aquellos que no se pueden medir, por ejemplo: el color de los ojos, el estado civil, la profesión….

Los caracteres estadísticos **cuantitativos** son aquellos que si se pueden medir, por ejemplo: el peso de una persona, la duración en horas de una lámpara, el número de días que llovió en Vecindario en el mes de Diciembre….

El conjunto de valores que puede tomar un carácter estadístico cuantitativo es lo que entenderemos por **Variables Estadística**. Las variables estadísticas se dividen en dos tipos: **Discretas y Continuas.**

Variable Estadística Discreta: Cuando puede tomar un número finito de valores o un número infinito pero numerable.

Ejemplo: el número de hijos de cada familia canaria, número de empleados de una fábrica, número de goles marcados por el Vecindario esta temporada….

Variable Estadística Continua: Cuando puede tomar, cualquier valor de un cierto intervalo de la recta real.

Ejemplo: el consumo de un coche cada 100 KM, el peso de los alumnos de una clase, presión sanguínea de una persona……

Ejercicio 1: Indica cuales de las siguientes variables estadística son cualitativas y cuales cuantitativas. En caso de ser cuantitativas, indica si son discretas o continuas.

a. Forma geométrica de las señales de tráfico
b. La altura de los alumnos del instituto.
c. Número de alumnos en cada aula del instituto.
d. Tipos de árboles plantados en los jardines de Vecindario.
e. Número de plantas en cada parque de Vecindario.
f. Litros de lluvia caídos por metro cuadrado en un determinado lugar durante un año.
g. Número de libros editados en una imprenta durante un mes.
h. Películas que han visto en el cine los alumnos en el último mes.
i. Kilómetros recorridos por un coche con 5 litros de gasolina.

2) TABLAS DE FRECUENCIAS.-

Consideremos una población de **N** individuos, de los que estudiaremos una carácter estadístico **A** que puede tomar **K** posibles valores $A_1, A_2, A_3, ..., A_K$

Designamos por n_i al número de individuos que presenta cada modalidad A_i, se conoce como **Frecuencia Absoluta** de la modalidad A_i.

Dividiendo la frecuencias absolutas, n_i, por el número total de observaciones, obtenemos la **Frecuencia Relativa** de la modalidad A_i, $f_i = \dfrac{n_i}{N}$

Propiedad: Cuando los valores que puede tomar un carácter estadístico son incompatibles, se cumple que la suma de las frecuencias absolutas es igual al total de la población estudiada, y la suma de las frecuencias relativas es igual a 1, es decir:

$$\sum_{i=1}^{K} n_i = n_1 + n_2 + ...n_K = N \qquad y \qquad \sum_{i=1}^{K} f_i = f_1 + f_2 + ... + f_K = 1$$

La **Frecuencia Absoluta Acumulada** del valor A_i, es la suma de todas las frecuencias absolutas anteriores a A_i más la frecuencia absoluta de A_i.

$N_p = n_1 + n_2 + ... + n_p$ Frecuencia Absoluta Acumulada.($p \leq k$)

La **Frecuencia Relativa Acumulada** del valor A_i, es la suma de todas las frecuencias relativas anteriores a A_i más la frecuencia relativa de A_i.

$F_p = f_1 + f_2 + ... + f_p$ Frecuencia Relativa Acumulada. ($p \leq k$)

Ejemplo: Supongamos que tiramos un dado 12 veces y obtenemos los siguientes resultados:
1 , 2 , 2 , 6 , 1 , 5 , 3 , 6 , 4 , 4 , 1 , 2.

Está claro que estamos trabajando con una variable estadística cuantitativa discreta, cuyos posibles valores son 1, 2, 3, 4, 5, 6; es decir 6 posibles modalidades o posibles resultados, por tanto en este ejercicio **N** = 12 y **K** = 6

Posibles Resultados	Frec. Absolutas n_i	Frec. relativas $f_i = \frac{n_i}{N}$	Frec. Abs. Acumuladas. N_i	Frec. Relativas Acumuladas. F_i
$A_1 = 1$	$n_1 = 3$	$f_1 = 3/12$	$N_1 = 3$	$F_1 = 3/12$
$A_2 = 2$	$n_2 = 3$	$f_2 = 3/12$	$N_2 = 6$	$F_2 = 6/12$
$A_3 = 3$	$n_3 = 1$	$f_3 = 1/12$	$N_3 = 7$	$F_3 = 7/12$
$A_4 = 4$	$n_4 = 2$	$f_4 = 2/12$	$N_4 = 9$	$F_4 = 9/12$
$A_5 = 5$	$n_5 = 1$	$f_5 = 1/12$	$N_5 = 10$	$F_5 = 10/12$
$A_6 = 6$	$n_6 = 2$	$f_6 = 2/12$	$N_6 = 12$	$F_6 = 12/12 = 1$

Ejercicio 2: Las notas obtenidas por 25 alumnos en un examen de matemáticas fueron las siguientes : 3, 9, 0, 6, 6, 5, 4, 3, 1, 8, 7, 2, 7, 10, 5, 4, 5, 8, 3, 5, 6, 2, 9, 6, 7.
Construye la tabla de frecuencias.

NOTA. Si la variable estadística es continua, o si a pesar de ser discreta los valores de dicha variable son muchos, entonces se suelen agrupar esos valores en intervalos.
La **Marca de Clase** es el valor que ocupa el centro del intervalo y se considera el representante del intervalo.

Ejemplo: Construir la tabla de frecuencias de las edades de las personas que acuden a un logopeda a lo largo de un mes, sabiendo que estas edades son: 3, 2, 11, 13, 4, 3, 2, 4, 5, 6, 7, 3, 4, 5, 3, 2, 5, 6, 27, 4, 15, 21, 12, 4, 3, 6, 29, 13, 6, 17, 6, 13, 6, 5, 12, 26.

Clases (Edades)	$\overline{x_i}$ = marca de clase	Fr. Abs. n_i	Fr. Relat. f_i	Fr .Abs. Ac. N_i	Fr. .Rel. Ac. F_i
[0, 5)	2,5	13	13/36	13	13/36
[5, 10)	7,5	11	11/36	24	24/36

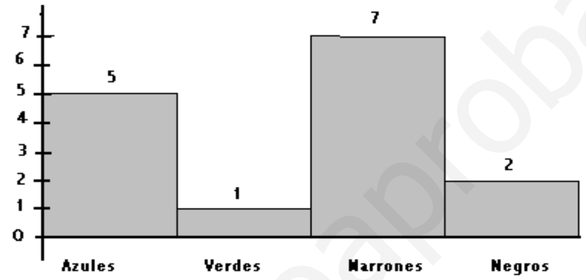
[10, 15)	12,5	6	6/36	30	30/36
[15, 20)	17,5	2	2/36	32	32/36
[20, 25)	22,5	1	1/36	33	33/36
[25, 30)	27,5	3	3/36	36	1

3) REPRESENTACIONES GRÁFICAS.-

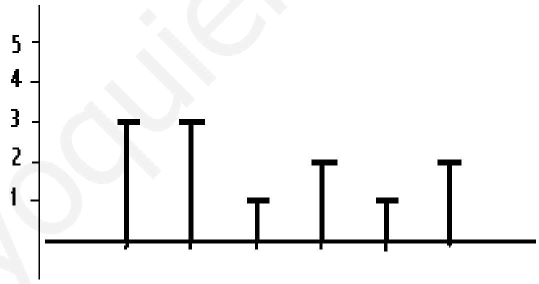
Los gráficos estadísticos nos permiten dar una idea visual y global de la distribución de los datos. Los más usados son los siguientes:

a) **Diagrama de Barras**: Tienen una base constante y una altura proporcional a la frecuencia absoluta correspondiente, las barras deben estar separadas. Son apropiados para variables cualitativas y cuantitativas discretas.

Ejemplo: Número de jugadores de un equipo de baloncesto según el color de sus ojos (V. cualitativa)

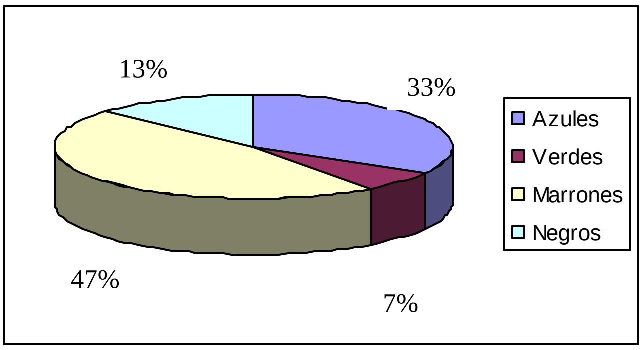


Si usamos el ejemplo anterior de la tirada del dado (V. discreta), sería



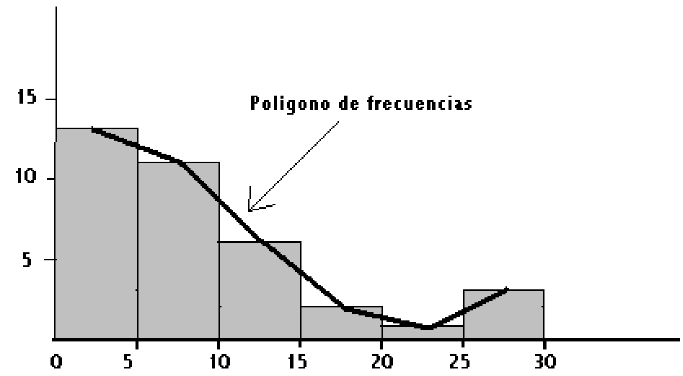
b) **Diagrama de Sectores**: Cada suceso viene representado por un sector circular de amplitud proporcional a su frecuencia. Son apropiados para variables cualitativas y cuantitativas discretas.

Para el **ejemplo** anterior sería:



c) **Histogramas.-** Está formado por rectángulos en los que las bases son las diferentes clases y las alturas son tales que las áreas resultantes equivalen a las respectivas frecuencias relativas.

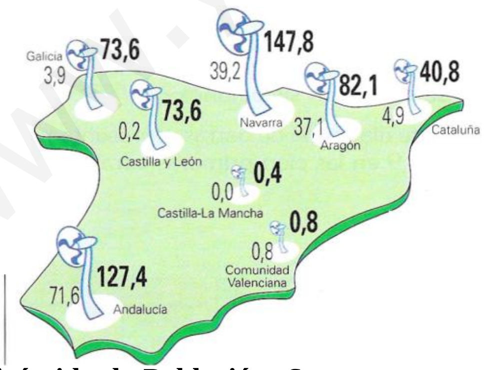
d) **Polígono de Frecuencias.-** Se construye a partir del diagrama de barras o del histograma, uniendo los puntos medios de la parte superior de cada rectángulo.



e) **Pictogramas**: Son dibujos alusivos al carácter estadístico estudiado y donde el tamaño de los mismos es proporcional a las frecuencias absolutas correspondientes.

El Pictograma muestra las potencias de las centrales eólicas españolas, en megavatios, en los años 1996 y 1997.

f) **Cartograma**: Mapa de una provincia, región o territorio nacional coloreado en distintos tonos y colores con unos cuadros al margen que indican su significado.



g) **Pirámide de Población**: Se representa gráficamente la población clasificada por edad y sexo.

4) MEDIDAS CARACTERÍSTICAS.-

Para estudiar una distribución de frecuencias, conviene dar algunas medidas objetiva que describan el comportamiento de los datos. Se agrupan en las siguientes categorías.

a) **MEDIDAS DE CENTRALIZACIÓN.-** Aquellos valores que representan a la distribución de frecuencias en su conjunto.

a.1) MEDIA ARITMÉTICA: La media de una variable estadística es la suma de todos sus valores dividido por el número de ellos. La media de la variable X se representa por $\bar{x}$.

$$\bar{x} = \sum_{i=1}^{K} f_i \bullet x_i = \frac{\sum_{i=1}^{K} n_i \bullet x_i}{N}$$, donde x_i son los valores de la variable.

Ejercicio 3: Calcular la nota media de un alumno que ha obtenido en sus exámenes: 7, 6, 3, 3, 1, 5, 9, 3, 4, 6.

Nota: En el caso de variable agrupada por intervalos, la media se calcula usando las marcas de clase de cada intervalo.

a.2) MEDIANA: La mediana es el valor de la variable estadística que divide en dos partes iguales los individuos de la población, una vez ordenados estos de menor a mayor.

Si el número de individuos es impar se toma el valor central.

Si el número de individuos es par se hace la media de los dos valores centrales.

Ejemplo: En el ejemplo anterior tenemos 10 notas con lo cual habrá dos valores centrales. Ordenemos primero las notas: 1, 3, 3, 3, 4, 5, 6, 6, 7, 9.

En este caso los valores centrales son el 4 y el 5, por tanto la mediana será $Me = \frac{4+5}{2} = 4{,}5$

Ejemplo: Supongamos que las notas de un alumno son: 5, 3, 7, 2, 6. En este caso tenemos un número impar.
Ordenamos las notas: 2, 3, 5, 6, 7; y la que ocupa la posición central es el 5, por tanto *Me = 5*.

(**IMPORTANTE**: Los ejemplos que acabamos de ver son para el caso de variables discretas.)

Para datos agrupados en intervalos, la mediana se calcula de la siguiente forma:

1. Se busca el intervalo en el que se encuentra la mediana, es decir es valor que divide a la distribución en dos partes iguales. Ese intervalo se denota por $[L_i, \ L_{i+1})$
2. $Me = L_i + \left(\frac{\frac{N}{2} - N_{i-1}}{n_i} \right) \bullet a$, donde ***a*** = amplitud del intervalo donde está la mediana.

Ejemplo. En el caso del logopeda sería:

1. el intervalo donde esta la mediana es [5, 10)
2. $Me = 5 + \left(\frac{18-13}{11} \right) \bullet 5 = 7{,}27$ años.

a.3) MODA: La moda de una variable estadística es el valor que más se repite, es decir el más frecuente, y por tanto el que corresponde al máximo en el diagrama de barras o en el histograma. A veces puede haber más de una moda.

Ejemplo: En el caso de las 10 notas de un alumnos, $Mo = 3$.

Para datos agrupados en intervalos (variable continua) la Moda se calcula de la siguiente forma:

1. Se busca el intervalo modal, es decir aquel donde la frecuencia absoluta es mayor. Llamaremos $[L_i, \ L_{i+1})$ a dicho intervalo.
2. $Mo = L_i + \left(\dfrac{\Delta_1}{\Delta_1 + \Delta_2}\right) \bullet a$ donde: a = amplitud del intervalo.

$$\Delta_1 = n_i - n_{i-1}$$
$$\Delta_2 = n_i - n_{i+1}$$

Ejemplo: El peso de 50 recién nacidos en una semana en un hospital viene dado por la siguiente tabla:

Peso en Kg.	[2.5 , 3)	[3, 3.5)	[3.5 , 4)	[4, 4.5)
Nº de niños	6	23	12	9

Está claro que el intervalo modal es este caso es el intervalo [3, 3.5) y su amplitud ***a*** = 0.5

En este caso $\Delta_1 = 23 - 6 = 17$ y $\Delta_2 = 23 - 12 = 11$ Por tanto:

$$Mo = 3 + \left(\frac{17}{17+11}\right) \bullet 0.5 = 3{,}3$$

Ejercicio 4: La distribución de las edades de los jugadores que intervienen en una determinada competición deportiva es:

Edad	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	Total
Nº	3	6	11	19	32	36	26	18	8	10	3	5	1	0	2	180

Calcula la media, mediana y moda.

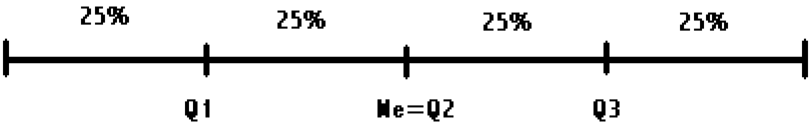
Ejercicio 5: La altura de los 30 alumnos de un aula viene dada por la siguiente tabla:

Altura (cm)	[140,150)	[150,160)	[160,170)	[170,180)	[180,190)
Nº alumnos	2	5	12	8	3

Calcula la altura media de dicha aula, así como la moda y la mediana.

b) **MEDIADAS DE POSICIÓN.-** Estas medidas dan los valores de algunas posiciones de interés en el conjunto de los datos.

b.1) CUARTILES: Si en lugar de partir la totalidad de los individuos en dos mitades (mediana), lo hacemos en cuatro partes iguales, obtenemos los cuarteles:



Q_1 = **Cuartil Inferior**: es el valor de la variable que deja el 25% de la población por debajo de él.

Q_2 = **Segundo Cuartil ó Mediana**: es el valor de la variable que deja el 50% de la población por debajo de él.

Q_3 = **Cuartil Superior**: es el valor de la variable que deja el 75% de la población por debajo de él.

Ejemplo: Calcula los 3 cuartiles de las edades de 12 pacientes que visitan a un dentista:
9 , 10, 27 , 12 , 14 , 21 , 18 , 10 , 20 , 13 , 12 , 18

<u>**Ejercicio 7:**</u> Hacer lo mismo con las siguientes edades: 2, 3, 3, 4, 6, 8, 12, 15, 19, 20, 25, 27, 35, 40, 48 (fíjate que ya están ordenadas).

<u>**Ejercicio 8:**</u> Hacer lo mismo con las siguientes edades: 15, 18, 12, 13, 22, 10, 15, 19, 16

EN EL CASO CONTINUO (o datos agrupados por intervalos)

1. Se busca el intervalo es el cual está el cuartil buscado : $[L_i\ ,\ \ L_{i-1})$
2. $Q_1 = L_i + \left(\dfrac{N \bullet \frac{25}{100} - N_{i-1}}{n_i} \right) \bullet a \quad ; \quad Q_2 = Me \quad ; \quad Q_3 = L_i + \left(\dfrac{N \bullet \frac{75}{100} - N_{i-1}}{n_i} \right) \bullet a$

Nota: Recuerda que $\frac{25}{100} = \frac{1}{4}$ y que $\frac{75}{100} = \frac{3}{4}$

<u>**Ejercicio 9:**</u> El peso de 50 recién nacidos en una semana en un hospital viene dado por la siguiente tabla:

Peso en Kg.	[2.5 , 3)	[3, 3.5)	[3.5 , 4)	[4, 4.5)
Nº de niños	6	23	12	9

Calcular los tres cuartiles

b.2) <u>**DECILES**</u>: Son los valores de la distribución que dividen la totalidad de los individuos en 10 partes iguales. Serían $D_1; D_2; D_3; ....; D_9$, donde D_K deja por debajo el $(K \bullet 10)\%$, es decir:

D_1 deja por debajo el 10% de los individuos.

D_2 deja por debajo el 20% de los individuos.

………..

D_9 deja por debajo el 90% de los individuos.

EN EL CASO CONTINUO vienen dados por la expresión: $D_K = L_i + \left(\dfrac{N \bullet \frac{K}{10} - N_{i-1}}{n_i} \right) \bullet a$

b.3) <u>PERCENTILES</u>: Son los valores de la distribución que dividen la totalidad de los individuos en 100 partes iguales.

EN EL CASO CONTINUO vienen dados por la expresión: $P_K = L_i + \left(\dfrac{N \bullet \frac{K}{100} - N_{i-1}}{n_i} \right) \bullet a$

<u>Nota</u>: Seguramente te habrás dado cuenta de que: $P_{25} = Q_1$; $P_{50} = D_5 = Q_2 = Me$; $P_{75} = Q_3$

<u>Ejercicio 10:</u> Las edades de lo 240 pacientes que pasaron por el servicio de urgencias del hospital insular en un día, vienen dadas por la siguiente tabla:

Edad	[0,10)	[10,20)	[20,30)	[30,40)	[40,50)	[50,60)	[60,70)	[70,80)	[80,90)	[90,100)
Nº	32	28	20	15	10	40	45	25	15	10

Calcula:

a. La edad media de los pacientes.
b. La moda y la mediana.
c. Q_2 y D_5
d. P_{45} , Q_3 , D_6 y P_{60}

c) **<u>MEDIDAS DE DISPERSIÓN.-</u>** Estas medidas, nos indican si los datos están muy agrupados en torno a un valor central o si por el contrario están muy dispersos respecto a este valor. Generalmente si estas medidas están próximas a 0, es por que los datos están agrupados, y si por lo contrario están lejos del 0, están dispersos, (aunque estar cerca y lejos de 0 siempre es muy relativo).

c.1) <u>RANGO o RECORRIDO</u>: Es la diferencia entre el mayor y el menor valor de una variable estadística.

Ejemplo: Supongamos que las notas de 8 alumnos son las siguientes 3, 3, 5, 6, 7, 7, 8, 9.

Rango = 9 – 3 = 6

c.2) <u>VARIANZA</u>: Se define como:

$$V = s^2 = \frac{1}{N} \bullet \sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2 = \sum f_i \bullet \left(x_i - \overline{x}\right)^2 = \sum f_i \bullet x_i^{\,2} - \overline{x}^2 = \frac{\sum x_i^{\,2} \bullet n_i}{N} - \overline{x}^2$$

Es decir: <u>Varianza = Media de los Cuadrados – Cuadrado de la Media</u>.

c.3) <u>DESVIACIÓN TÍPICA</u>: Es la raíz cuadrada positiva de la varianza. Se denota por **s**.

Es el parámetro de dispersión más utilizado, y además viene dado en las mismas unidades que los datos, cosa que no ocurría con la varianza.

c.4) <u>COEFICIENTE DE VARIACIÓN DE PEARSON</u>: Se define por $C_P = \dfrac{s}{\overline{x}}$

Medida útil para comparar distribuciones distintas, este coeficiente no depende de las unidades en que se midan los valores de la variable y suele expresarse en porcentaje. Este coeficiente no es muy útil si $\overline{x}$ está próximo a cero.

Ejemplo: Las notas de un examen realizado por 30 alumnos, son las siguientes:

5, 3, 4, 1, 2, 8, 9, 8, 7, 6, 6, 7, 9, 8, 7, 7, 1, 0, 1, 5, 9, 9, 8, 0, 8, 8, 8, 9, 5, 7.

Calcula el Rango, Varianza, Desviación típica y Coeficiente de Pearson.

Solución:

Notas (x_i)	n_i	$x_i \bullet n_i$	$x_i^2 \bullet n_i$
0	2	0	0
1	3	3	3
2	1	2	4
3	1	3	9
4	1	4	16
5	3	15	75
6	2	12	72
7	5	35	245
8	7	56	448
9	5	45	405
	N = 30	$\sum x_i \bullet n_i = 175$	$\sum x_i^2 \bullet n_i = 1277$

Rango = 9 – 0 = 9 $\qquad Media = \overline{x} = \dfrac{\sum x_i \bullet n_i}{N} = \dfrac{175}{30} = 5{,}83$

$$V = s^2 = \frac{\sum x_i^2 \bullet n_i}{N} - \overline{x}^2 = \frac{1277}{30} - (5{,}83)^2 = 42{,}57 - 33{,}99 = 8{,}58 \quad \Rightarrow \quad s = \sqrt{8{,}58} = 2{,}93$$

$$C_P = \frac{s}{\overline{x}} = \frac{2{,}93}{5{,}83} = 0{,}5$$

<u>**Nota**</u>: Si los datos están agrupados en intervalos utilizo las marcas de clase para calcular las medidas de dispersión.

<u>**Ejercicio 11:**</u> Durante el mes de Julio, en una determinada ciudad canaria, se registraron las siguientes temperaturas máximas: 32, 31, 28, 29, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33.

a. Forma la tabla de frecuencias
b. Representa gráficamente la distribución
c. Calcula la media, mediana y moda
d. Calcula el cuartil inferior y el superior
e. Calcula el recorrido, la varianza y la desviación típica

<u>**Ejercicio 12:**</u> El peso de 50 recién nacidos en una semana en un hospital viene dado por la siguiente tabla:

Peso en Kg.	[2.5 , 3)	[3, 3.5)	[3.5 , 4)	[4, 4.5)
Nº de niños	6	23	12	9

a. Representa gráficamente la distribución mediante un histograma y un polígono de frecuencias.
b. Calcula la media, la moda y la mediana
c. Calcula la varianza y la desviación típica
d. Calcula P_{40} ; D_7 y P_{70}

Ejercicio 13: Se aplicó un test sobre "satisfacción en el trabajo" a 88 empleados de una fábrica obteniéndose los siguientes resultados:

Puntuación	[38,44)	[44,50)	[50,56)	[56,62)	[62,68)	[68,74)	[74,80)
Nº trabajadores	7	8	15	25	18	9	6

a. Calcula la media, moda, mediana, rango, varianza, desviación típica y coeficiente de Pearson.
b. Calcula el tercer cuartil, P_{62} y D_3

Ejercicio 14: Los saldos deudores de 12 clientes de una empresa en miles de euros son:

120, 25, 131, 32, 117, 27, 142, 30, 130, 122, 28, 25

a. Representa gráficamente estos datos.
b. Calcula su media y desviación típica y explica porque es tan grande la desviación típica.

Ejercicio 15: En una población de 25 familias se observo la variable estadística X = "número de coches que tiene cada familia" obteniéndose los siguientes datos:

0, 1, 2, 3, 1, 0, 1, 1, 1, 4, 3, 2, 2, 1, 1, 2, 2, 1, 1, 1, 2 , 1 , 3, 2, 1.

a. Construye la tabla de frecuencias de esa distribución
b. Construye el diagrama de barras correspondiente.
c. Calcula la moda, mediana y desviación típica.
d. Calcula Q_1 , Q_3 y D_2

Ejercicio 16: Al estudiar la distribución de las edades de una población, se obtuvieron los siguientes datos:

Edad	[0,20)	[20,40)	[40,60)	[60,80)
Nº personas	15	¿?	15	16

Como puedes ver se ha extraviado el dato correspondiente al intervalo [20,40)

a. ¿Cuál sería ese dato si la edad media fuese de 35 años?
b. ¿Cuál sería la desviación típica si ese dato fuese 16? Interpreta el resultado.

TEMA 2: DISTRIBUCIONES BIDIMENSIONALES

1) INTRODUCCIÓN.-

Al hacer el estudio estadístico de un colectivo, puede ser que en cada observación se considere un solo carácter, obteniéndose una variable estadística unidimensional; o bien se estudien dos o más caracteres de un mismo individuo, obteniéndose así variables estadísticas bidimensionales o multidimensionales, respectivamente.

Es el caso del estudio de la talla y el peso de un recién nacido, la edad y la presión arterial de una persona, los ingresos y gastos de una familia, las calificaciones en varias asignaturas de un curso, etc.

En este tema sólo consideraremos el caso del estudio de dos caracteres, obteniendo para cada individuo dos valores que nos permitirán establecer una serie estadística doble, que recibirá el nombre de distribución bidimensional.

Tablas de Frecuencias.

Sean A y B los dos caracteres que estamos estudiando. El carácter A define una variable estadística X de valores $x_1; x_2;.....; x_n$ (n modalidades) y el carácter B define otra variable estadística Y de valores $y_1; y_2;.....; y_m$ (m modalidades). A cada individuo de la población se le asigna un par $(x_i; y_j)$. Todos estos pares constituyen una serie estadística doble, que nos permite construir una tabla de frecuencias absolutas de doble entrada que recibe el nombre de **tabla de frecuencias** o tabla de correlación.

Si n_{ij} es el número de veces que se presenta el par $(x_i; y_j)$, a n_{ij} se le denomina **frecuencia absoluta conjunta** de $(x_i; y_j)$.

Ejemplo: El Celta y el Deportivo han jugado entre ellos 38 partidos en primera división, cuyos resultados están reflejados en la siguiente tabla:

(Y) R.C. CELTA → (X) DEPORTIVO ↓	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$y_4 = 3$	$y_5 = 4$	Frec. absolutas marginales de X ↓
$x_1 = 0$	5	3	6	4	3	$A_1 = 21$
$x_2 = 1$	2	3	1	1	2	$A_2 = 9$
$x_3 = 2$	2	1	2	1	0	$A_3 = 6$
$x_4 = 3$	0	1	1	0	0	$A_4 = 2$
Frec. Ausol. marginales de Y →	$B_1 = 9$	$B_2 = 8$	$B_3 = 10$	$B_4 = 6$	$B_5 = 5$	**N = 38**

Las frecuencias absolutas marginales de X indican el número de veces que el Deportivo consiguió 0, 1, 2 ó 3 goles, y las frecuencias absolutas marginales de Y indican el número de veces que el Celta consiguió frente al Deportivo 0, 1, 2, 3 ó 4 goles respectivamente, verificándose que:

$A_i = \sum_j n_{ij}$ Frecuencia absoluta marginal de la variable X.

$B_j = \sum_i n_{ij}$ Frecuencia absoluta marginal de la variable Y.

$N = \sum_i \sum_j n_{ij} = \sum_i A_i = \sum_j B_j$ Número total de individuos de la población.

Si llamamos frecuencia relativa del par $(x_i ; y_j)$ a la razón $f_{ij} = \frac{n_{ij}}{N}$, podemos construir otra tabla en las frecuencias relativas de cada par:

(Y) R.C.CELTA → (X) DEPORTIVO ↓	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$y_4 = 3$	$y_5 = 4$	Frec. relativas marginales de X ↓
$x_1 = 0$	5/38	3/38	6/38	4/38	3/38	21/38
$x_2 = 1$	2/38	3/38	1/38	1/38	2/38	9/38
$x_3 = 2$	2/38	1/38	2/38	1/38	0	6/38
$x_4 = 3$	0	1/38	1/38	0	0	2/38
Frec. Relat. marginales de Y →	9/38	8/38	10/38	6/38	5/38	**1**

2) DISTRIBUCIONES MARGINALES Y CONDICIONADAS.-

La **Distribución Marginal** de la variable X, en una distribución bidimensional (X, Y) viene definida por los valores que toma dicha variable y las frecuencias de los mismos, independientemente de los valores de la otra variable Y.
Análogamente se define la distribución marginal de Y.

En nuestro ejemplo anterior, las dos distribuciones marginales vendrán dadas por las siguientes tablas:

X	Frec. absoluta marginal Ai	Frec. relativa marginal fi
0	21	21/38
1	9	9/38
2	6	6/38
3	2	2/38

X	Frec. absoluta marginal Bj	Frec. relativa marginal fj
0	9	9/38
1	8	8/38
2	10	10/38
3	6	6/38
4	5	5/38

La **Distribución Condicionada** de la variable Y, condicionada a que la variable X tome el valor x_i, es una variable con frecuencias:

$$f(Y = y_j / X = x_i) = \frac{n_{ij}}{\sum_k n_{ik}}$$

En nuestro ejemplo, la distribución condicionada de Y (goles que marca el Celta) condicionada a que X=0 (que el Deportivo no marque ningún gol) sería:

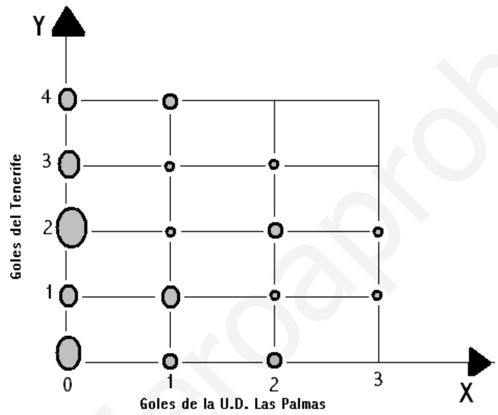
Y condicionada a que X=0	Y=0/X=0	Y=1/X=0	Y=2/X=0	Y=3/X=0	Y=4/X=0
Frecuencia relativa	5/21	3/21	6/21	4/21	3/21

Ejercicio 1: Construye la tabla de la distribución de X condicionada a que Y=1, también la de Y condicionada a que X=2.

3) REPRESENTACIONES GRÁFICAS.-

A partir de la tabla estadística correspondiente a una distribución bidimensional se puede obtener su representación gráfica. Entre los distintos tipos de representaciones nos referiremos a dos de ellos:

- **Diagrama de Dispersión o Nube de Puntos**: En un sistema de referencia cartesiano se asigna un carácter a cada uno de los ejes; sobre él se indicarán los puntos correspondientes a los pares $(x_i ; y_j)$, y en cada uno de ellos realizaremos una representación que muestre su correspondiente frecuencia.



- **Diagrama de Barras**: Consiste en fijar un sistema cartesiano tridimensional, representar en el plano horizontal los caracteres X e Y, y levantar en cada punto $(x_i ; y_j)$una barra paralela al eje vertical, con longitud proporcional a la frecuencia correspondiente a cada par.

Ejercicio 2: Se ha estudiado el número de helados que vende un quiosco, en función de la temperatura, a lo largo de 10 días siendo los resultados los siguientes:

Temperatura (X)	18	20	21	19	15	18	23	27	30	26
Número de helados(Y)	20	22	20	15	12	15	28	35	40	35

- Construye las tablas de las distribuciones marginales
- *f*(Y=20/X=18)
- *f*(X=15/Y=20)
- Representa la nube de puntos.

4) PARÁMETROS ESTADÍSTICOS.-

4.1) MEDIAS.- Si en una distribución bidimensional consideramos las distribuciones marginales de cada variable por separado, obtenemos dos distribuciones unidimensionales de las que podemos calcular la media de cada una de ellas:

- Media de la variable X.- Para la distribución marginal X:

X	x_1	x_2	…….	x_n
Frec. absoluta marginal Ai	A_1	A_2	…….	A_n
Frec. relativa marginal fi	f_1	f_2	…….	f_n

La media de la variable X será: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i \bullet A_i}{N} = \sum_{i=1}^{n} x_i \bullet f_i$

En nuestro ejemplo anterior, la media de los goles conseguidos por el Deportivo en los 38 partidos disputados frente al Celta, sería:

$$\bar{x} = \frac{0\bullet 21 + 1\bullet 9 + 2\bullet 6 + 3\bullet 2}{38} = 0\bullet\frac{21}{38} + 1\bullet\frac{9}{28} + 2\bullet\frac{6}{38} + 3\bullet\frac{2}{38} = \frac{27}{38} = 0{,}710526315$$

- Media de la variable Y.- De la misma forma calculamos la media de la variable Y.

Y	y_1	y_2	…….	y_m
Frec. absoluta marginal Bj	B_1	B_2	…….	B_m
Frec. relativa marginal fj	f_1	f_2	…….	f_m

La media de la variable Y será: $\bar{y} = \dfrac{\sum_{j=1}^{m} y_j \bullet B_j}{N} = \sum_{j=1}^{m} y_j \bullet f_j$

En nuestro ejemplo la media de goles conseguidos por el Celta sería:

$$\bar{y} = \frac{0\bullet 9 + 1\bullet 8 + 2\bullet 10 + 3\bullet 6 + 4\bullet 5}{38} = 0\bullet\frac{9}{38} + 1\bullet\frac{8}{38} + 2\bullet\frac{10}{38} + 3\bullet\frac{6}{38} + 4\bullet\frac{5}{38} = \frac{66}{38} = 1{,}736842105$$

Ejercicio 3: Calcula la temperatura media y la media del número de helados vendidos por el quiosco del ejercicio de la página anterior.

4.2) VARIANZAS.- Considerando las distribuciones marginales:

La **Varianza** de la variable **X:** $S_x^2 = \dfrac{\sum_{i=1}^{n} x_i^2 \bullet A_i}{N} - \left(\bar{x}\right)^2 = \sum_{i=1}^{n} x_i{}^2 \bullet f_i - \left(\bar{x}\right)^2$

La **Desviación Típica** es la raíz cuadrada positiva de la varianza: $S_x = \sqrt{S_x^2}$

La **Varianza** de la variable **Y:** $S_x^2 = \dfrac{\sum_{j=1}^{m} y_j{}^2 \bullet B_j}{N} - \left(\bar{y}\right)^2 = \sum_{j=1}^{m} y_j \bullet f_j - \left(\bar{y}\right)^2$

La **Desviación Típica** es la raíz cuadrada positiva de la varianza: $S_y = \sqrt{S_y^2}$

4.3) COVARIANZA.- Es un parámetro específico de una variable bidimensional. Se llama C**ovarianza** de un variable bidimensional, con valores $\left(x_i ; y_j\right)$, a la media aritmética de los productos de las variaciones de cada variable respecto de la media. Se representa por S_{xy} y su expresión es:

$$S_{xy} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i - \bar{x}\right)\bullet\left(y_j - \bar{y}\right)\bullet n_{ij}}{N}$$

El cálculo de la varianza usando la fórmula anterior suele ser muy complicado, por eso normalmente usaremos otra fórmula que se deduce del desarrollo de la anterior:

S_{xy} **= Media de los productos – Producto de las medias**

Es decir: $$S_{xy} = \frac{\sum_{i=1}^{n}\sum_{j=i}^{m} x_i \bullet y_j \bullet n_{ij}}{N} - \bar{x} \bullet \bar{y}$$

En nuestro ejemplo:
X =”Nº goles marcados por el Deportivo”, Y = ”Nº goles marcados por el Celta”

$$S_x^2 = \frac{0 \bullet 21 + 1 \bullet 9 + 4 \bullet 6 + 9 \bullet 2}{38} - \left(\frac{27}{38}\right)^2 = \frac{51}{38} - \frac{729}{1444} = \frac{1938}{1444} - \frac{729}{1444} = \frac{1209}{1444} = 0{,}837$$

$$S_y^2 = \frac{0 \bullet 9 + 1 \bullet 8 + 4 \bullet 10 + 9 \bullet 6 + 16 \bullet 5}{38} - \left(\frac{66}{38}\right)^2 = \frac{2560}{1444} = 1{,}773$$

$S_x = \sqrt{0{,}837} = 0{,}915$ y $S_y = \sqrt{1{,}773} = 1{,}3315$

$$S_{xy} = \frac{0 \bullet 25 + 1 \bullet 3 + 2 \bullet 2 + 3 \bullet 2 + 4 \bullet 4 + 6 \bullet 2 + 8 \bullet 0 + 9 \bullet 0 + 12 \bullet 0}{38} - \frac{27}{38} \bullet \frac{66}{38} = \frac{-224}{1444} = -0{,}155$$

Nota: Observa que la covarianza puede ser negativa, no así las varianzas. La covarianza indica el sentido de la variación conjunta de las dos variables. Si la covarianza es positiva indica que las dos variables varían en el mismo sentido (si aumenta una aumenta también la otra y si disminuye una disminuye la otra), y si la varianza es negativa entonces varían en sentido contrario.

Ejercicio 4: Calcula las varianzas, desviaciones típicas y covarianza en el ejercicio anterior que relaciona la temperatura y el número de helados vendidos por un quiosco.

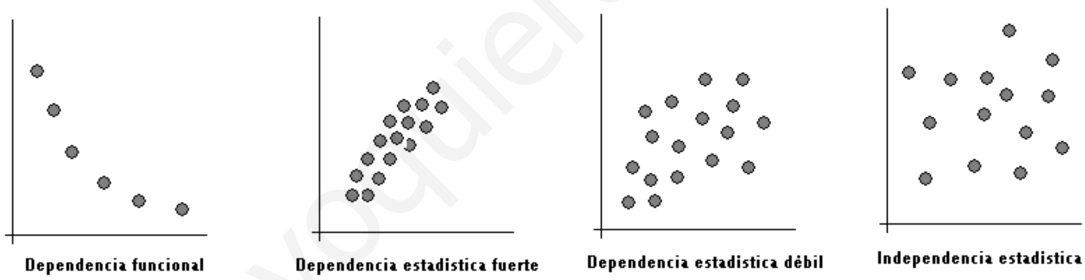
Ejercicio 5: Se ha preguntado a 8 alumnos que nos digan el número de horas de estudio y su nota alcanzado al final del curso, siendo los resultados los siguientes:

Horas estudio (X)	1	3	2	3	5	1	2	4
Nota Obtenida (Y)	4	6	7	9	8	3	5	6

- Dibuja la nube de puntos.
- Obtén la tabla de las distribuciones marginales.
- Calcula $f(X=3/Y=6)$; $f(X=2/Y=7)$; $f(X=4/Y=9)$; $f(Y=9/X=3)$
- Calcula las medias y las desviaciones típicas
- Calcula la Covarianza

5) DEPENDENCIA ESTADÍSTICA. REGRESIÓN LINEAL.-

La consideración simultánea de dos caracteres induce a pensar en la posibilidad de que exista una relación entre ellos. Es evidente que habrá casos en los que se pueda considerar que si existe esa relación, y habrá otros en los que la relación sea más bien escasa.



En el caso de dependencia estadística intentaremos ajustar la nube de puntos a una función matemática que se aproxime a ellos. En general puede ser cualquier función, pero nos limitaremos a ajustar rectas a esa nube de puntos. Serán las llamadas **rectas de regresión**.

Rectas de Regresión.- Denominaremos recta de regresión a la que mejor se ajuste a la nube de puntos.
La **recta de regresión de Y sobre X**, nos facilitará los valores estimados de Y conocidos los de X

$$y-\bar{y}=\frac{S_{xy}}{S_x^2}\bullet\left(x-\bar{x}\right)$$

La **recta de regresión de X sobre Y**, nos facilitará los valores estimados de X conocidos los de Y

$$x-\bar{x}=\frac{S_{xy}}{S_y^{\ 2}}\bullet\left(y-\bar{y}\right)$$

Nota: $\frac{S_{xy}}{S_x^2}$ y $\frac{S_{xy}}{S_y^2}$ se denominan coeficientes de regresión.

Nota: Tanto la recta de regresión de Y sobre X, como la recta de regresión de X sobre Y, pasan por el punto $(\bar{x};\bar{y})$

En nuestro ejemplo:

Recta de regresión de Y sobre X : $y-1{,}74=\dfrac{-0{,}155}{0{,}837}\bullet(x-0{,}71)$

Es decir: $y-1{,}74=-0{,}185\bullet(x-0{,}71) \quad \Rightarrow \quad y=-0{,}185x+1{,}87$

Como puedes observar la pendiente es negativa pues la covarianza es negativa, con lo cual al aumentar una variable disminuye la otra.

Mediante esta recta si sabemos los goles que marca el Deportivo, podemos estimar los goles que marcará el Celta.

Ejercicio 6: Si el Deportivo. marca 5 goles, ¿cuántos goles marcará el Celta?

Ejercicio 7: Calcula la recta de regresión de X sobre Y.
Si el Celta marca 5 goles, ¿cuántos marcará el Deportivo?
Dibuja los dos rectas de regresión.

Ejercicio 8: Se ha preguntado a 8 alumnos que nos digan el número de horas de estudio y su nota alcanzado al final del curso, siendo los resultados los siguientes:

Horas estudio (X)	1	3	2	3	5	1	2	4
Nota Obtenida (Y)	4	6	7	9	8	3	5	6

- Calcula la recta de regresión de Y sobre X.
- ¿Qué nota se estima que puede sacar un alumno que estudia 2'5 horas?

6) CORRELACIÓN LINEAL.-

Se entiende por correlación la dependencia que existe entre las variables de una distribución. Una medida comúnmente aceptada para determinar esta dependencia es el llamado **coeficiente de correlación lineal.**

El **Coeficiente de Correlación Lineal** es: $\mathbf{r}=\dfrac{S_{xy}}{S_x\bullet S_y}$

El signo de **r** coincide con el signo de la covarianza S_{xy}, puesto que las desviaciones típicas S_x y S_y son positivas siempre.

Si **r > 0** la correlación es directa, es decir al aumentar una variable también aumenta la otra.
Si **r < 0** la correlación es inversa, es decir al aumentar una variable la otra disminuye.

Por como está definido, el coeficiente de correlación siempre toma valores en el intervalo comprendido entre [-1,1].
Si **r = 0** las dos rectas de regresión serían paralelas a los ejes y por lo tanto las variables serían <u>independientes o no correlacionadas</u>.

Si **r = 1 o r = -1** ambas rectas coinciden y la dependencia sería perfecta, es decir existe dependencia funcional.

Para los demás valores que pueda tomar r la dependencia estadística es tanto mas fuerte cuanto más cerca esté r de los valores 1 o -1. Si la dependencia es fuerte se dice que la correlación es significativa o fuerte. Si los valores de r están alejados de 1 o -1 la correlación es débil.

En nuestro ejemplo de los goles entre la U.D. Las Palmas y el Tenerife, el coeficiente de correlación es:

$$r = \frac{-0'155}{0'915 \bullet 1'3315} = -0'127$$

Como podemos ver la correlación está próxima a 0, es decir la correlación es muy débil, o casi inexistente por lo que las estimaciones que hagamos no serán muy de fiar. Al ser el coeficiente de correlación negativo indica que al aumentar una variable disminuye la otra, es decir cuantos más goles marque un equipo menos marcará el otro.

Ejercicio 9: Calcula e interpreta el coeficiente de correlación para el ejercicio que se encuentra al inicio de esta página.

Ejercicio 10: La distribución de las edades y la presión arterial de 10 personas es:

Edad (X)	30	28	35	42	51	42	63	32	70	67
Tensión (Y)	11’5	11’3	12’5	13’5	14’6	13	16’6	12	16’9	

- Representa la nube de puntos. ¿Se puede proceder a un ajuste lineal?
- Calcula el coeficiente de correlación e interpreta el resultado.
- ¿Qué tensión se espera que tenga una persona de 60 años?
- ¿Cuál sería la edad estimada para una persona que tenga 14 de tensión?
- Dibuja las rectas de regresión.

Ejercicio 11: Se midió el contenido en oxigeno en mg/l del agua de un lago a distintas profundidades, obteniéndose los siguientes datos:

Profundidad (en metros)	15	25	35	45	55	65
Mg/l de oxigeno	6’5	5’9	4’7	3’8	2’7	1’4

- Calcula la recta de regresión del contenido de oxigeno (Y) respecto de la profundidad (X). Dibuja la nube de puntos y la recta de regresión.
- Estudia e interpreta el coeficiente de correlación entre ambas variables.
- ¿Qué contenido en oxigeno se puede predecir para una profundidad de 70 metros?

Ejercicio 12: Una fábrica de una cierta marca de refrescos estudió al azar 10 semanas del año, observando la temperatura media correspondiente a cada semana y la cantidad de refrescos pedidos durante las mismas. La información obtenida fue la siguiente:

Temperatura media	10	28	12	31	30	19	24	5	9	15
Cantidad de refrescos	21	65	19	72	75	39	67	11	12	24

- ¿Puede la fábrica planificar su producción en función de la temperatura?, ¿de qué forma?

Ejercicio 13: La siguiente tabla da el número de calzado y los pesos de 55 estudiantes. Con estos datos estudia la dependencia lineal o correlación entre las dos variables.

Peso / Nº calzado	55	60	65	70	75	80	85
39	1						
40		3	3	4			
41		3	4	6			1
42			8	8	7	2	
43			2		1		
44							2

Ejercicio 14: El índice de mortalidad (Y) de siete grupos de fumadores que consumían diariamente **x** cigarrillos, aparece en la tabla siguiente:

Nº de cigarrillos	X	3	5	6	15	20	40	45
Índice de mortalidad	Y	0’2	0’3	0’3	0’5	0’7	1’4	1’5

- Estudia la correlación entre X e Y.
- ¿Qué mortalidad se puede predecir para un consumidor de 60 cigarrillos diarios?

Ejercicio 15: Las notas obtenidas por 5 alumnos en Matemáticas y Música son:

Matemáticas	6	4	8	5	3’5
Música	6’5	4’5	7	5	4

- Determina las rectas de regresión y calcula la nota esperada en Música para un alumno que tiene un 7’5 en Matemáticas.

Ejercicio 16: Una empresa dispone de los datos de la siguiente tabla:

Millones en publicidad	1	2	3	4	5	6	7	8
Ventas	15	16	14	18	21	19	19	21

- Estima las ventas esperadas al invertir 10 millones en publicidad. Explica la fiabilidad de la estimación realizada. (Los datos de las ventas son también en millones)

Ejercicio 17: Una distribución bidimensional de la edad y del peso de unos niños está dada por la siguiente tabla:

EDAD	2	3	5	7	9
PESO	14	20	30	42	46

- Calcula el coeficiente de correlación entre las dos variables. Interpreta el resultado.
- Calcula la recta de regresión del peso sobre la edad
- ¿Cuál sería el peso estimado de un niño de 10 años?

TEMA 3: CÁLCULO DE PROBABILIDADES

1) ÁLGEBRA DE SUCESOS.

1.1) Experimentos Aleatorios y Determinísticos.- Un fenómeno **Determinístico** es aquel en el cual conociendo las causas, se pueden prever los efectos, por ejemplo en una reacción química, en un problema de física, etc.

Un fenómeno **Aleatorio** es aquel en el que no se puede establecer la relación causa-efecto, por ejemplo en la lotería, en la tirada de un dado, etc. Un fenómeno aleatorio está caracterizado por dos premisas:

a. No se sabe cual será el resultado.
b. Sin embargo, conocemos el conjunto de los posibles resultados (este conjunto se denomina espacio muestral y se denota por la letra E ó Ω)

En la práctica no se puede estudiar todas las posibles situaciones de un fenómeno (requiere mucho tiempo y dinero), lo que hace la estadística es considerar una **muestra** o sacar unas conclusiones que podrán ser aplicables al resto de la **población**.
Por ejemplo, si una fábrica de neumáticos quiere estudiar la resistencia de los mismos, es evidente que no puede probar la resistencia de todos los neumáticos, entonces se escoge un cierto número de ellos y se estudian esos.

Se denomina **Población** a un conjunto de elementos en los que se estudia una determinada característica.

Se denomina **Muestra** a un subconjunto representativo de la población que se aísla para su análisis estadístico. Es muy importante que la muestra sea representativa para que los resultados obtenidos se puedan extender con ciertas garantías al resto de la población.

1.2) Espacio Muestral y Sucesos.- Llamamos **Espacio Muestral** de un experimento aleatorio al conjunto de todos los posibles resultados del experimento. Se designa por **E** o Ω.

A cada uno de los elementos del espacio muestral se llamará **Suceso Elemental**.

Por ejemplo, si consideramos el experimento consistente en lanzar dos dados y anotar la suma de los números que aparecen en las caras superiores, el espacio muestral sería:

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

Un suceso elemental sería por ejemplo {7} = “que la suma de los dados sea 7”

Distinguimos Distintos Tipos de Sucesos:

- Se llama **Suceso** de un experimento aleatorio a cada uno de los subconjuntos del espacio muestral E.

Por ejemplo, si consideramos el experimento aleatorio consistente en tirar un dado, el espacio muestral sería E = {1,2,3,4,5,6}. Algunos de los sucesos de este experimento aleatorio podrían ser : A = “salir par” = {2,4,6} ; B = “salir impar” = {1,3,5} ; C = “salir múltiplo de 3 “ = {3,6} ; D = “salir mayor que 2” = {3,4,5,6} ; E = “salir menor o igual que 3” = {1,2} ; etc.

- Suceso Imposible o Suceso Nulo es aquel que no se verifica nunca. Se designa por ϕ . En nuestro ejemplo anterior, un suceso nulo sería ϕ = {7,10,12}

- Suceso Seguro es aquel que siempre se verifica, coincide con E.

- Suceso Contrario o Suceso Complementario de un suceso A es el suceso que se verifica cuando no se verifica A, y viceversa. Se representa por $\overline{A}$ o por A^*.

En nuestro ejemplo anterior $\overline{A}$ = {1,3,5} ; $\overline{B}$ = {2,4,6} ; $\overline{C}$ = {1,2,4,5} ; etc.
Se verifica que: $A \cup \overline{A} = E$ y $A \cap \overline{A} = \phi$

1.3) Operaciones con Sucesos.

- Unión de Sucesos. Dados dos sucesos A y B de un mismo experimento aleatorio, llamamos Suceso Unión de A y B al suceso que se realiza cuando se realiza A o B. Se representa por $A \cup B$.

En nuestro ejemplo anterior: $B \cup C$ = {1,3,5,6} = "salir impar **o** múltiplo de 3"

- Intersección de Sucesos. Sucesos Incompatibles. Dados dos sucesos A y B de un experimento aleatorio, llamamos Suceso Intersección de A y B al suceso que se realiza cuando se realizan simultáneamente los sucesos A y B. Se representa por $A \cap B$.

En nuestro ejemplo anterior: $B \cap C$ = {3} = "salir impar **y** múltiplo de 3"="salir 3"

Dados dos sucesos *A* y *B* de un experimento aleatorio, diremos que *A* y *B* son Incompatibles si $A \cap B = \phi$ (es decir no se pueden verificar a la vez). En caso contrario, es decir $A \cap B \neq \phi$ diremos que *A* y *B* son Compatibles.

En nuestro ejemplo A y B son Incompatibles y B y C son Compatibles.

- Diferencia de Sucesos. Dados dos sucesos A y B de un experimento aleatorio, llamamos Suceso Diferencia, y se escribe A-B, al suceso formado por los sucesos elementales de A que no son de B.

En nuestro ejemplo A-C={2,4}

2) EXPERIMENTOS COMPUESTOS. ESPACIOS COMPUESTOS.

Consideremos el experimento aleatorio que consiste en el lanzamiento de un dado y una moneda, en realidad se trata de dos experimentos simples. Los experimentos formados por varios experimentos simples se llaman Experimentos Compuestos y su espacio muestral será un Espacio Compuesto.

En este caso E = { (1,C),(1,X),(2,C),(2,X),(3,C),(3,X),(4,C),(4,X),(5,C),(5,X),(6,C),(6,X) } que se obtiene a partir de $E_1 = \{1,2,3,4,5,6\}$ y $E_2 = \{C, X\}$ donde $E = E_1 x E_2$

Ejercicios Resueltos:

1. Con el juego del dominó y considerando la suma de puntos de cada ficha, obtén:
   a. Espacio muestral.
   b. A = "Salir número primo"
   c. B = "Salir número impar"

   Solución:
   En el dominó, las fichas van de (0,0) a (6,6); por tanto su suma va desde 0 hasta 12, por tanto:
   a. E = {0,1,2,3,4,5,6,7,8,9,10,11,12}
   b. A = {2,3,5,7,11}
   c. B = {1,3,5,7,9,11}

2. Sean A, B y C tres sucesos del espacio muestra E. Utilizando estos sucesos expresa:
   a. Los tres sucesos suceden simultáneamente.
   b. Ocurren A o B, pero no C.
   c. Ocurre alguno de los tres sucesos.
   d. Ninguno de los tres sucede

   Solución:
   a. $A\cap B\cap C$
   b. $(A\cup B)\cap\overline{C}$
   c. $A\cup B\cup C$
   d. $\left(\overline{A\cup B\cup C}\right)=\overline{A}\cap\overline{B}\cap\overline{C}$

3. Sea una urna con 9 bolas numeradas del 1 al 9. Sacamos una bola, miramos el número y la devolvemos. Sean los sucesos:
   A = "Salir número primo"
   B = "Salir número impar"
   C = "Salir múltiplo de 3"

   Calcula los sucesos:
   a. $A\cap B$
   b. $B\cap C$
   c. $(A\cup B)\cap C$
   d. $A\cap\overline{B}$
   e. $B-C$
   f. $\overline{A\cup B}$

   Solución:
   E = {1,2,3,4,5,6,7,8,9}
   A = {2,3,5,7} B = {1,3,5,7,9} C = {3,6,9}

   a. $A\cap B$ ={3,5,7}
   b. $B\cap C$ ={3,9}
   c. $(A\cup B)\cap C$ = {1,2,3,5,7,9} ∩ {3,6,9} = {3,9}
   d. $A\cap\overline{B}$ = {2,3,5,7} ∩ {2,4,6,8} = {2}
   e. B – C = { 1,5,7}
   f. $\overline{A\cup B}=\overline{\{1,2,3,5,7,9\}}=\{4,6,8\}$

3) PROBABILIDAD.-

Ley de los Grandes Números.- En la siguiente tabla se muestra los resultados de lanzar un dado un número determinado de veces:

Nº de Lanzamientos	20	40	60	80	100	120	140	160	180	200
Nº de Caras	11	20	31	43	53	62	72	83	92	101
Frec. Absoluta (ni)	11	20	31	43	53	62	72	83	92	101
Frec. Relativa (fi)	0.55	0.5	0.517	0.537	0.53	0.516	0.514	0.519	0.511	0.505

Si repitiésemos el experimento un número de veces muy grande, veríamos que las frecuencias relativas del suceso "cara" tienden a estabilizarse en torno al valor 0.5, es decir que la frecuencia relativa del suceso "cara" tomará valores aproximados por exceso o defecto a 0.5, de tal forma que las oscilaciones alrededor de ese valor serán cada vez más pequeñas.

Llegamos así a la llamada "**ley de los Grandes Números**", que enunciada de forma sencilla dice que "**La frecuencia relativa de un suceso tiende a estabilizarse en torno a un número, a medida que el número de pruebas crece indefinidamente**"
A este número, al que la frecuencia relativa de un suceso se acerca cuanto mayor es el número de pruebas realizadas, lo llamaremos **Probabilidad** del suceso.
Esta definición presenta un inconveniente de tipo práctico, por eso veremos otra definición de probabilidad, aunque esta sirve para dar una idea intuitiva del concepto de probabilidad.

Definición Axiomática de la Probabilidad.- Llamaremos **Probabilidad** a una ley que asocia a cada suceso **A**, de un experimento aleatorio, un número real que llamaremos probabilidad de **A** y representaremos por **p(A),** que cumple los siguientes axiomas:

1. $p(A) \geq 0$, es decir la probabilidad de un suceso siempre es positiva o nula.
2. La probabilidad del suceso seguro es igual a la unidad: $p(E) = 1$. Por tanto la probabilidad de un suceso no puede ser mayor que uno.
3. La probabilidad de la unión de dos sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos, es decir **si A y B son Incompatibles** $p(A \cup B) = p(A) + p(B)$**.**

Propiedades de la probabilidad (se deducen de los axiomas)

a) Probabilidad del suceso contrario: $p(\overline{A}) = 1 - p(A)$ (por tanto $p(A) \leq 1$)

b) Probabilidad del suceso imposible: $P(\phi) = 0$

Ejemplo: Calcula la probabilidad de que al tirar 3 monedas se obtenga por lo menos una cara.

Solución:

E = {(c,c,c);(c,c,x),(c,x,c),(x,c,c),(c,x,x),(x,c,x),(x,x,c)} donde los sucesos elementales son equiprobables, y la probabilidad de cada suceso será 1/8.
Por norma, cuando aparezca "por lo menos" es conveniente recurrir al suceso contrario.
Si A = {Obtener por lo menos una cara} → $\overline{A}$ = {Obtener 3 cruces} y como $p(\overline{A}) = 1/8$ entonces $p(A) = 1 - p(\overline{A}) \Rightarrow p(A) = 7/8$.

c) Probabilidad de la unión de sucesos incompatibles.- El tercer axioma de la probabilidad dice que: $p(A \cup B) = P(A) + P(B)$ si A y B son incompatibles ($A \bigcap B = \phi$)

Esta propiedad se puede extender al caso de más de dos sucesos incompatibles y por tanto:

$p(A \bigcup B \bigcup C) = p(A) + p(B) + p(C)$ con A, B y C incompatibles dos a dos.

Y en general:

$$p(A_1 \bigcup A_2 \bigcup ... \bigcup A_n) = p(A_1) + p(A_2) + ... + p(A_n) \text{ con } A_i \bigcap A_j = \phi \,;\, i \neq j \,;\, 1 \leq i, j \leq n$$

Probabilidad de la unión de sucesos compatibles. Para dos sucesos A y B de un mismo experimento aleatorio se verifica:

$$p(A \bigcup B) = p(A) + p(B) - p(A \bigcap B)$$

En caso de tres sucesos A, B y C tendríamos:

$$p(A \bigcup B \bigcup C) = p(A) + p(B) + p(C) - p(A \bigcap B) - p(A \bigcap C) - p(B \bigcap C) + p(A \bigcap B \bigcap C)$$

Ejemplo: De una baraja española se extrae una carta. Consideremos los sucesos A="salir oro", B="salir rey" y C="salir el as de espadas". Calcula la probabilidad de:

1) A U B 2) A U C

Solución:

1. A y B son compatibles, pues $A \bigcap B$="salir rey de oros" por tanto:

$$p(A \bigcup B) = p(A) + p(B) - p(A \bigcap B) = \frac{10}{40} + \frac{4}{40} - \frac{1}{40} = \frac{13}{40}$$

2. A y C son incompatibles pues $A \bigcap B = \phi$ por tanto:

$$p(A \bigcup B) = p(A) + p(B) = \frac{10}{40} + \frac{1}{40} = \frac{11}{40}$$

Ejemplos:

1. Se consideran dos sucesos A y B, asociados a un experimento aleatorio con P(A) = 0'7, P(B) = 0'6. ¿Pueden ser A y B incompatibles?

   Solución: A y B son incompatibles si A ∩ B = Ø.

   P (A U B) = P(A) + P(B) – P(A∩ B) = 0'7 + 0'6 – P(A∩ B) y como la probabilidad de cualquier suceso no puede ser mayor que 1 → P(A∩ B) > 0 → A∩ B ≠ Ø por tanto no son incompatibles.

2. Un estudiante hace dos pruebas en un mismo día. La probabilidad de que pase la primera es 0'6, la de que pase la segunda es 0'8 y la probabilidad de que pase ambas es 0'5.
   a. ¿Son A y B compatibles o incompatibles?
   b. Calcula la probabilidad de que pase al menos una prueba?
   c. Calcula la probabilidad de que no pase ninguna prueba.

   Solución: Sean: A = "pasar la primera prueba" y B = "pasar la segunda prueba"
   a. Evidentemente son compatibles pues P(A ∩ B) ≠ 0 → A∩ B ≠ Ø
   b. P(A U B) = P(A) + P(B) – P(A ∩ B) = 0'6 + 0'8 – 0'5 = 0'9
   c. $P(\overline{A \bigcup B}) = 1 - P(A \bigcup B)$ = 1 – 0'9 = 0'1

4) REGLA DE LAPLACE.

Consideremos un experimento aleatorio en el que todos los posibles sucesos elementales son igualmente probables, es decir **equiprobables**, entonces para un suceso A de este experimento se cumple:

$$p(A) = \frac{N^o \;\; de \;\; casos \;\; favorables}{N^o \;\; de \;\; casos \;\; posibles}$$

Ejemplo: En una urna tengo 10 bolas, 3 son de color rojo, 4 verdes y 3 azules. Calcula:

a. Si saco una bola:
   i. Probabilidad de que sea roja.
   ii. Probabilidad de que sea verde.
b. Si saco dos bolas a la vez (es decir sin reemplazamiento)
   i. Probabilidad de que las dos sean azules.

Solución:

a.
   i. P(R) = 3/10
   ii. P(V) = 4/10
b.
   i. P(AA) = $\frac{\binom{3}{2}}{\binom{10}{2}} = \frac{3}{45} = \frac{1}{15}$

5) PROBABILIDAD CONDICIONADA.

Teniendo en cuenta la relación entre frecuencia relativa y probabilidad podemos definir el concepto de probabilidad condicionada.
Llamamos **Probabilidad Condicionada del suceso B respecto del suceso A**, que denotamos por P(B/A), al siguiente cociente:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) \neq 0$$

Representa la probabilidad del suceso B supuesto que ya se dio A.

Análogamente, tendríamos la probabilidad de A condicionada a B:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) \neq 0$$

De las dos relaciones anteriores se obtiene:

$$P(A \cap B) = P(A) \bullet P(B/A) \qquad\qquad P(A \cap B) = P(B) \bullet P(A/B)$$

Ejemplo: En el experimento consistente en sacar una carta de una baraja de 40 cartas consideremos los sucesos: S = "sacar una sota" y E = "sacar una espada". Calcula P(S/E) y P(E/S)

Solución:

$$P(S/E) = \frac{P(S\bigcap E)}{P(E)} = \frac{1/40}{10/40} = \frac{1}{10} \qquad P(E/S) = \frac{P(S\bigcap E)}{P(S)} = \frac{1/40}{4/40} = \frac{1}{4}$$

Ejemplo: De una urna que contiene 9 bolas rojas y 5 negras, se extraen sucesivamente dos bolas. Calcula la probabilidad de los siguientes sucesos:

a. A = "las dos bolas sean negras"
b. B = " las dos bolas sean rojas"
c. C = "la primera sea roja y la segunda negra"
d. D = "una sea roja y otra negra"

Solución:

a. Sea N1 = "sacar bola negra en la primera extracción"
Sea N2 = "sacar bola negra en la segunda extracción"

$$P(A) = P(N1\cap N2) = P(N1) \cdot P(N2/N1) = \frac{5}{14} \bullet \frac{4}{13} = \frac{10}{91} \text{ ó también } P(A) = \frac{\binom{5}{2}}{\binom{14}{2}} = \frac{10}{91}$$

b. Sea V1 = " sacar bola roja en la primera extracción"
Sea V2 =" sacar bola roja en la segunda extracción"

$$P(B) = P(V1\cap V2) = P(V1) \cdot P(V2/V1) = \frac{9}{14} \bullet \frac{8}{13} = \frac{36}{91} \text{ ó también } P(B) = \frac{\binom{9}{2}}{\binom{14}{2}} = \frac{36}{91}$$

c. $P(C) = P(V1 \cap N2) = P(V1) \cdot P(N2/V1) = \frac{9}{14} \bullet \frac{5}{13} = \frac{45}{182}$

d.

$$P(D) = P((V1\cap N2) \cup (N1\cap V2)) = P(V1\cap N2) + P(N1\cap V2) = \frac{9}{14} \bullet \frac{5}{13} + \frac{5}{14} \bullet \frac{9}{13} = \frac{90}{182} = \frac{45}{91}$$

Ó también $P(D) = \frac{\binom{9}{1} \bullet \binom{5}{1}}{\binom{14}{2}} = \frac{45}{91}$

6) SUCESOS DEPENDIENTES E INDEPENDIENTES.

Dados dos sucesos A y B con probabilidades no nulas, diremos que B es INDEPENDIENTE de A si P(B/A) = P(B) (es decir A no afecta a la probabilidad de B)

Se cumple:

B es independiente de A $\Leftrightarrow$ P(B/A) = P(B) $\Leftrightarrow \dfrac{P(A\bigcap B)}{P(A)} = P(B) \Leftrightarrow$

P(A∩B) = P(A) · P(B) $\Leftrightarrow \dfrac{P(A\bigcap B)}{P(B)} = P(A) \Leftrightarrow$ P(A/B) = P(A) $\Leftrightarrow$ A es independiente de A

Diremos pues que dos sucesos **A y B son independientes si P(B/A) = P(B) 0 si P(A/B) = P(A),** es decir entre A y B no observamos ninguna relación de causa-efecto. En caso contrario diremos que son **dependientes.**

Ejemplo: Dada una urna con 5 bolas negras y 3 bolas blancas, se consideran los sucesos N=”sacar bola negra” y B=”sacar bola blanca”. Son N y B compatibles en los siguientes casos:

a. Se extraen las bolas sin reemplazamiento.
b. Se extraen las bolas con reemplazamiento.

Solución:
a. P(N/B) = 5/7 pero P(N) = 5/8 es decir dependientes.
b. P(N/B) = 5/8 y P(N) = 5/8 es decir son independientes.

7) PROBABILIDAD DE LA INTERSECCIÓN DE SUCESOS.

Distinguiremos dos casos:

a. Cuando los sucesos son dependientes.
b. Cuando los sucesos son independientes.

Probabilidad de la Intersección de Sucesos Dependientes:
A partir de la definición de probabilidad condicionada obtenemos:

P(A∩B) = P(A) · P(B/A) si A y B son dependientes

Análogamente, para el caso de tres sucesos obtenemos:

P(A∩B∩C) = P(A) · P(B/A) · P(C/(A∩B)) si A, B y C son dependientes

Probabilidad de la Intersección de Sucesos Independientes.
Si A y B son independientes, entonces:

P(A∩B) = P(A) · P(B) si A y B son independientes.(pues en este caso P(B)=P(B/A)

En general para el caso de n sucesos independientes:

P(A1∩A2∩…….∩An) = P(A1) · P(A2) · ……· P(An)

Ejemplo: Calcula la probabilidad de obtener cuatro caras en cuatro lanzamientos de una moneda.

Solución: Los sucesos son independientes, puesto que lo que sale en un lanzamiento no influye el los demás.

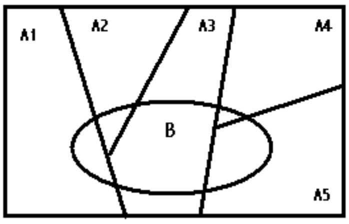
P(C∩C∩C∩C) = P(C) · P(C) · P(C) · P(C) = ½ · ½ · ½ · ½ = 1/16

8) TEOREMA DE PROBABILIDAD TOTAL Y BAYES.

Si A1, A2,……,An forman una partición del espacio muestral E, es decir:

- A1 U A2 U …… U An = E
- Ai ∩Aj = Ø si $i \neq j$ (i,j = 1,2,….,n)

Y sea B un suceso cualquiera del espacio muestral E, se verifica que :



$$P(B) = \sum P(B \cap Ai) = \sum P(Ai) \bullet P(B/Ai)$$

Fórmula de Bayes.

Si A1, A2,……,An forman una partición del espacio muestral E, es decir:

- A1 U A2 U …… U An = E
- Ai ∩Aj = Ø si $i \neq j$ (i,j = 1,2,….,n)

Y sea B un suceso cualquiera del espacio muestral E, se verifica que:

$$P(Ai/B) = \frac{P(Ai \cap B)}{P(B)} = \frac{P(B/Ai) \bullet P(Ai)}{\sum P(B/Ai) \bullet P(Ai)}$$

Ejemplo: En una empresa el 30% de los trabajadores son mujeres. El 20% de las mujeres fuman y el 35% de los hombres también fuman.. ¿cuál es la probabilidad de que un trabajador fume?. Si un trabajador fuma, ¿cuál es la probabilidad de que sea mujer?

Solución:

Sean los sucesos: H = "ser hombre" P(H) = 0'7
M = "ser mujer" P(M) = 0'3
F = "ser fumador" P(F/M) = 0'2 ; P(F/H) = 0'35

Los sucesos H y M son una partición del espacio muestral.

a. P(F)= P(M) · P(F/M) + P(H) · P(F/H) = 0'3 · 0'2 + 0'7 · 0'35 = 0'305

b. $$P(M/F) = \frac{P(M) \bullet P(F/M)}{P(M) \bullet P(F/M) + P(H) \bullet P(F/H)} = \frac{0'3 \bullet 0'2}{0'3 \bullet 0'2 + 0'7 \bullet 0'35} = \frac{0'06}{0'305} = 0'1967$$

Ejemplo: Disponemos de una urna A con 4 bolas rojas y 3 bolas negras, y otra B con 5 rojas y 8 negras. Se elige una urna al azar y se extrae una bola que resulta ser negra. ¿Cuál es la probabilidad de que sea de la urna B?

Solución:

$$P(B/N) = \frac{P(B) \bullet P(N/B)}{P(B) \bullet P(N/B) + P(A) \bullet P(N/A)} = \frac{\frac{1}{2} \bullet \frac{8}{13}}{\frac{1}{2} \bullet \frac{8}{13} + \frac{1}{2} \bullet \frac{3}{7}} = \frac{\frac{8}{13}}{\frac{8}{13} + \frac{3}{7}} = \frac{56}{95} = 0'5895$$

Ejemplo: En un edificio se usan dos ascensores, el primero lo usan el 45% de los inquilinos y el resto usan el segundo. El porcentaje de fallos del primero es del 5% mientras que del segundo es del 8%. Si un cierto día un inquilino queda atrapado en el ascensor, calcula la probabilidad de que fuese en el primer ascensor:

Solución:
Sean los sucesos: A = "coger el primer ascensor" P(A) = 0'45
B = "coger el segundo ascensor" P(B) = 0'65
F = "fallar el ascensor" P(F/A)= 0'05 y P(F/B)= 0'08
Entonces:

$$P(A/F)=\frac{P(A)\bullet P(F/A)}{P(A)\bullet P(F/A)+P(B)\bullet P(F/B)}=\frac{0'45\bullet 0'05}{0'45\bullet 0'05+0'55\bullet 0'08}=\frac{0'0225}{0'0665}=\frac{225}{665}=\frac{45}{133}=0'3383$$

Nota: Hacer estos tres últimos ejemplos usando diagramas en árbol.

EJERCICIOS PROPUESTOS:

1. Considera el experimento aleatorio del lanzamiento de un dado y una moneda. Obtén:
   a. Espacio muestral
   b. Suceso par y cara
   c. Suceso impar y cruz

2. Una urna contiene 3 bolas blancas y 2 negras. Obtén los posibles resultados y sus probabilidades al extraer dos bolas.
   a. Sin reemplazamiento.
   b. Con reemplazamiento.

3. De una baraja de 40 cartas extraemos una carta. Sean los sucesos:
   A ="sacar copas" B ="sacar as" C ="sacar as de oros"
   Determina los sucesos siguientes:
   a. A∩B
   b. AUC
   c. B∩C
   d. $A\bigcap\overline{B}$
   e. $\left(A\bigcup B\right)\bigcap\overline{C}$
   f. B – C

4. Considera el espacio muestral E = {a,b,c,d} en el que los 4 sucesos elementales tienen la misma probabilidad. Sean S1 = {a,b} y S2 = {a,c}
   a. ¿Son S1 y S2 sucesos incompatibles?
   b. Calcula la probabilidad del suceso S1 U S2 y la probabilidad del suceso contrario de S1.

5. Dados los sucesos A y B con probabilidades: P(A) = 1/3 ; P(B) = ¼ y P(A∩B)=1/12. Obtén:
   a. $P(\overline{A}\bigcup\overline{B})$
   b. $P(\overline{A\bigcup B})$

6. En un banco hay dos sistemas de seguridad, A y B. El sistema A funciona 90 de cada 100 veces, y el B, 80 de cada 100 veces, y los dos a la vez, 75 de cada 100 veces. ¿Cuál es la probabilidad de que no funcione ninguno de los dos sistemas?

7. La probabilidad de que un estudiante apruebe una oposición es 0'5, la de otro es 0'4 y la de que aprueben ambos es 0'1. Calcula:
   a. La probabilidad de que al menos uno apruebe.
   b. La probabilidad de que ninguno apruebe.
   c. La probabilidad de que sólo apruebe uno.

8. Si la intersección de dos sucesos independientes es 0'2, y la de su unión es 0'7, ¿cuál es la probabilidad de cada uno de los sucesos?

9. De una baraja se extraen simultáneamente tres cartas al azar. Encuentra la probabilidad de que:
   a. Las tres sean bastos.
   b. Alguna de las cartas sea un oro.

10. Se ha seguido la pista a 100000 coches durante un año. Éstos son de tres marcas distintas A, B y C. Unos han tenido accidentes (Ac) y otros no (No Ac). Se reparten según la siguiente tabla:

	A	**B**	**C**
Ac.	650	200	150
No Ac.	49350	19800	29850

    Calcula:
    a. ¿Cuál de las tres marcas es más segura?
    b. ¿Cuáles de los sucesos A, B y C son independientes de Ac y de No Ac?

11. Por una investigación realizada entre los alumnos de una clase, se sabe que el 50% aprueban matemáticas; el 60% aprueban lengua; el 70% economía; el 30% matemáticas y lengua; el 40% matemáticas y economía, y el 50% lengua y economía. ¿Son independientes los sucesos aprobar cada asignatura?

12. Sean A y B dos sucesos, tales que P(A) = 0'40, P(B/A) = 0'25 y P(B) = **b**. Halla:
    a. $P(A \cap B)$
    b. P(A U B) si **b** = 0'5
    c. El menor valor posible de **b**.
    d. El mayor valor posible de **b**.

13. La ruleta de un casino consta de 40 casilla, numeradas del 1 al 40. Los números acabados en 1, 2, 3, 4 05 son rojos y el resto negros. Puesta en marcha la ruleta, se consideran los sucesos siguientes: A ="salir un número de la primera decena"; B ="salir número par" y C="salir un número rojo". Averigua:
    a. P(C – A)
    b. Probabilidad de que el número sea de la primera decena sabiendo que es rojo.
    c. ¿Son independientes A y B? ¿y A y C?

14. La probabilidad de que un globo sonda sea recuperado es 1/9. Si tres globos son lanzados al espacio, ¿cuál es la probabilidad de recuperarlos en cada caso?
    a. Sólo uno
    b. Los tres
    c. Al menos uno

15. De una baraja española de 40 cartas se extrae una al azar, ¿cuál es la probabilidad de que sea bastos o menor que 5?

16. Una urna A contiene 6 bolas blancas y 4 negras, una segunda urna B contiene 5 bolas blancas y 2 negras. Se selecciona una urna al azar y de ella se extraen 2 bolas sin reemplazamiento. Calcula la probabilidad de que sean: a) Blancas. b) Del mismo color. c) de distinto color.

17. En una ciudad el 55% de los habitantes consumen pan integral, el 30% consume pan de multicereales y el 20% consume ambos.
   a. Sabiendo que un habitante consume pan integral, ¿cuál es la probabilidad de que coma pan de multicereales?
   b. Sabiendo que un habitante consume pan de multicereales, ¿cuál es la probabilidad de que no consuma pan integral?
   c. ¿Cuál es la probabilidad de que una persona de esa ciudad no consuma ninguno de los dos tipos de pan?

18. El contenedor A tiene un 10% de piezas defectuosas y el contenedor B tiene un 5% de piezas defectuosas. Si ambos contenedores tienen el mismo número de piezas, se elige un contenedor al azar, y dentro de él se escoge una pieza también al azar.
   a. Halla la probabilidad de que la pieza sea defectuosa.
   b. Si la pieza obtenida es defectuosa, calcula la probabilidad de que la pieza provenga del contenedor A. Calcula también en este caso la probabilidad de la pieza provenga del contenedor B.

19. En un centro de enseñanza todos los alumnos aprueban alguna asignatura. Se conoce que el 30% de los alumnos aprueban la asignatura A, el 40% la asignatura B y el 5% aprueban ambas. Calcula las siguientes probabilidades de un alumno:
   a. Apruebe cualquier otra asignatura.
   b. Apruebe la A y no la B.
   c. Si aprueba la asignatura B, no apruebe la A.

20. Una urna A contiene dos bolas blancas y una negra, y otra urna B contiene dos bolas negras y una blanca. Se extraen dos bolas de la urna A y sin mirar el color, se introducen en la B. A continuación, se extrae una bola de la urna B. ¿Cuál es la probabilidad de que esa bola sea negra?

21. En una universidad existen tres facultades A, B y C. En A hay matriculados 150 chicas y 50 chicos, en B 300 chicas y 200 chicos y en C 150 chicas y 150 chicos.
   a. Calcula la probabilidad de que un estudiante elegido al azar sea chico.
   b. Si un estudiante elegido al azar resulta chico, ¿cuál es su facultad más probable?

22. Se dispone de dos urnas A y B, de idéntico aspecto. La urna A contiene 4 bolas rojas y 2 amarillas, mientras que B contiene 5 bolas rojas y 3 amarillas. Un individuo se dirige a una de las urnas y extrae, sin reemplazamiento, dos bolas. Halla la probabilidad de que:
   a. Ambas sean rojas
   b. Las dos bolas sean del mismo color

23. Se estima que sólo el 20% de los que compran acciones en Bolsa tienen conocimientos bursátiles. De ellos, el 80% obtiene beneficios. De los que compran acciones sin conocimientos bursátiles sólo el 10% obtiene beneficios. Se desea saber:
   a. El tanto por ciento que obtiene beneficios en la compra de acciones.
   b. Si se elige al azar una persona que ha comprado acciones en Bolsa y resulta que ha obtenido beneficios, ¿cuál es la probabilidad de que tenga conocimientos bursátiles?

24. En un baile de disfraces se reúnen 10 matrimonios. Si se eligen al azar dos personas, calcula la probabilidad de que:
   a. Las dos personas sean esposos. (Solución: 1/19)
   b. Una sea hombre y otra mujer. (Solución: 10/19)

25. En un concurso de televisión, el concursante elige al azar una permutación de 5 letras diferentes. Al compararla con una fijada previamente, gana 1000 euros por cada letra colocada en el mismo lugar que en la permutación fijada inicialmente. Si acierta las 5, el premio es de 100.000 euros.
   a. ¿Qué probabilidad tiene el concursante de ganar 100.000 euros?
   b. ¿Y de ganar 4000 euros?

26. De una cesta en la que hay 4 higos podridos, un niño jugando mete un higo en otra cesta en la que hay 6 higos podridos y 18 sanos. Si luego sacamos un higo de esta cesta y vemos que no está podrido, calcula la probabilidad de que el higo que metió el niño en esta cesta estuviese sano.

27. La probabilidad de que al llamar a la centralita de la Facultad de Matemáticas el teléfono esté comunicando es de 0’3. La probabilidad de que la telefonista nos diga que la extensión del departamento de Estadística comunica es 0’2. calcula la probabilidad de que logremos comunicar con el Departamento de Estadística para que nos resuelvan este problema.

28. La probabilidad de que una bomba lanzada por un avión haga blanco en un objetivo es de 1/3. Calcula.
   a. Probabilidad de alcanzar el objetivo si se tiran 3 bombas.(Sol: 19/27)
   b. Probabilidad de que las tres bombas alcancen el objetivo. (Sol: 1/27)

29. Un producto es fabricado en tres fases independientes: A, B y C. El proceso de fabricación es tal que la probabilidad de un defecto en A es 0’03, de un defecto en B es 0’04 y de un defecto en C es 0’08. ¿Cuál es la probabilidad de que un producto no sea defectuoso?
   (Solución: 0’856704)

30. La probabilidad de que un hombre y una mujer vivan 50 años o más es de 0’6 y 0’7 respectivamente. Se pide:
   a. Probabilidad de que vivan los dos después de 50 años.
   b. Probabilidad de que viva sólo la mujer.
   c. Probabilidad de que viva por lo menos uno de los dos.
   d. Probabilidad de que no viva ninguno de los dos.

31. En un centro hay 1000 alumnos repartidos así:

	Chicos	Chicas
Usan Gafas	187	113
No Usan Gafas	413	287

   Se elige al azar uno de ellos. ¿Cuál es la probabilidad de que sea:
   a) Chico; b) Chica; c) Use gafas; d) No use gafas; e) Sea una chica con gafas.
   f) Se elige alguno al azar y me dicen que es una chica. ¿Cuál es la probabilidad de que use gafas?

32. Una clase se compone de veinte alumnos, de los cuales doce son chicas. La mitad de los alumnos y la tercera parte de las alumnas aprueban matemáticas. Calcular la probabilidad de que al elegir una persona al azar, resulte ser:
   a) Alumna que aprueba matemáticas; b) Alumno que suspenda matemáticas.
   c) Que apruebe matemáticas; d) Que sea chico, sabiendo que a suspendido.

33. Sean A y B dos sucesos tales que P(AUB) =3/4; P($\overline{B}$) =2/3; y P(A∩B)=¼ . Hallar P(A) y P(B).

34. Determinar si son compatibles o incompatibles los sucesos A y B en los siguientes casos:
 a) P(A) = ¼ ; P(B) = ½ ; P(AUB) = 2/3 b) P(A) = 0 ; P(B) = ½

35. En un cajón de un armario, Juan guarda desordenadamente 3 pares de calcetines blancos y 4 pares de calcetines rojos; otro cajón contiene 4 corbatas blancas, 3 rojas y 2 azules. Para vestirse saca al azar del primer cajón un par de calcetines, y del segundo, una corbata. Hallar la probabilidad de que los calcetines y la corbata sean del mismo color.

36. Un producto está formado de dos partes A y B. El proceso de fabricación es tal que la probabilidad de un defecto en A es 0,06 y la probabilidad de un defecto en B es 0,07. ¿Cuál es la probabilidad de que el producto no sea defectuoso?. Si se sabe que el producto es defectuoso, ¿cuál es la probabilidad de que el defecto se haya producido en A?

37. En cierto país, donde la enfermedad X es endémica, se sabe que un 12% de la población padece dicha enfermedad. Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiables, ya que da positiva en el 90% de los casos de personas realmente enfermas y también da positiva en el 5% de personas sanas. ¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

38. En un banquete el 40% de los comensales han tomado pescado, el 50% carne, un 25% ambos platos y el resto de los son vegetarianos. Se elige un comensal al azar. Hallar probabilidad de:
 a) Como carne o pescado b) No coma pescado c) Sea vegetariano d)Coma sólo carne.

39. En una clase de treinta alumnos, cinco han leído “El Quijote”, 12 han leído “El lazarillo de Tormes” y 3 han leído ambos libros. Se elige un alumno al azar. Calcula:
 a) Probabilidad de que haya leído alguno de los dos libros. b) Que no haya leído ninguno.
 c) Que no haya leído ninguno d) Que haya leído sólo “El Quijote”
 e) ¿Quién Escribió “El Quijote”?, ¿y “El lazarillo de Tormes”?

40. Determina si son dependientes o independientes los sucesos A y B en los siguientes casos:
 a) P(A)=1/5 ; P(B)=1/2 ; P(A∩B)=1/10 b) P(A)=0,2 ; P(B)=0,3 ; P(A∩B)=0,6

41. Una Urna A tiene 6 bolas blancas y 4 negras, una segunda Urna B contiene 5 bolas blancas y 2 negras. Se extrae una bola de la urna A, y luego otra de la urna B. ¿Cuál es la probabilidad de que sean de distinto color?

42. Tengo una Urna A con 6 bolas blancas y 4 negras, y una segunda Urna B con 5 bolas blancas y 2 negras. Tiro un dado y si sale 1 ó 2 saco una bola de la urna A, y en caso contrario la saco de la urna B. Calcula: a) Probabilidad de que la bola sea blanca. b) si la bola ha sido blanca, ¿cuál es la probabilidad de que sea de la urna A?.

43. En una determinada ciudad se sabe la mitad de se población tiene conocimientos de informática, la tercera parte tiene conocimientos de contabilidad y una quinta parte tiene conocimientos de informática y contabilidad. Se elige una persona al azar. Calcular la probabilidad de que:
 a) Tenga conocimientos de contabilidad; **b**) Tenga conocimientos de informática o contabilidad.
 c) No tenga ningún conocimiento de ninguna; **d**) Sólo tenga conocimientos de contabilidad

TEMA 4: DISTRIBUCIONES:

Variable Aleatoria:

Una Variable Aleatoria asociada a un experimento será una aplicación del espacio muestral E (o Ω) en el conjunto de los números reales R.

$$\xi : E \longrightarrow R$$
$$A \longrightarrow X_A \in R$$

Ejemplo: Consideremos un experimento consistente en lanzar una moneda al aire tres veces. El espacio muestral sería $E = \{(c,c,c),(c,c,x),(c,x,c),(x,c,c),(c,x,x),(x,c,x),(x,x,c),(x,x,x)\}$
Cada uno de esos 8 posibles resultados se denomina suceso elemental, y la unión de varios sucesos elementales forma un suceso.

Vamos estudiar ahora la variable aleatoria ξ = "Nº de caras". Los posibles valores que toma esta variable serán: 0, 1, 2, 3.

$(c,c,c) \xrightarrow{\xi} 3$ $(c,c,x) \xrightarrow{\xi} 2$ $(c,x,c) \xrightarrow{\xi} 2$ $(x,c,c) \xrightarrow{\xi} 2$
$(c,x,x) \xrightarrow{\xi} 1$ $(x,c,x) \xrightarrow{\xi} 1$ $(x,x,c) \xrightarrow{\xi} 1$ $(x,x,x) \xrightarrow{\xi} 0$

Es frecuente confundir la variable aleatoria con el conjunto de valores asignados, que es lo que se llama Recorrido de la variable.

Ejercicio: Calcula en el ejemplo anterior $P(\xi = 3)$, $P(\xi = 2)$, $P(\xi = 1)$ y $P(\xi = 0)$

Ejercicio: Considera el experimento consistente en lanzar una moneda al aire dos veces, y sea ***X*** = "Nº de cruces". Calcula: $P(X = 2)$, $P(X = 1)$ y $P(X = 0)$

Ejercicio: Considera el experimento consistente en lanzar dos dados, y sea la variable aleatoria X = "Suma de los dos dados":

- ¿Qué valores puede tomar esta variable?
- ¿Cuál es el espacio muestral de dicho experimento?
- Calcula el suceso B = "Que sumen menos de cinco"
- Calcula el suceso C = "Que sumen un número par"
- Calcula el suceso $B \cap C$. ¿Cómo llamarías a este suceso?
- Calcula el suceso $B \cup C$. ¿Cómo llamarías a este suceso?
- Calcula: $P(X = 3)$, $P(X = 7)$, $P(X \leq 5)$, $P(C)$, $P(B \cap C)$ y $P(B \cup C)$

Distinguiremos dos tipos de variables aleatoria, según sea el recorrido de la variable.

Variable Aleatoria Discreta.

Una Variable Aleatoria $\xi : E \longrightarrow R$ es discreta cuando toma un número finito (o numerable) de valores.

La Variable Aleatoria del ejemplo anterior sería discreta, puesto que toma los valores 0, 1, 2 y 3.

Función Masa de Probabilidad.

Se llama Función Masa de Probabilidad asociada a una variable aleatoria discreta "ξ" a una aplicación que asigna a cada valor x_i su correspondiente probabilidad p_i.

$$\{x_1, x_2, ..., x_n\} \longrightarrow [0,1]$$
$$x_i \longrightarrow p_i = p(\xi = x_i)$$

En nuestro ejemplo $p(0) = \frac{1}{8}$, $p(1) = \frac{2}{8}$, $p(2) = \frac{3}{8}$, $p(3) = \frac{1}{8}$

La Función Masa de Probabilidad se suele expresar mediante una tabla del tipo:

ξ	x_1	x_2	$\cdots$	x_n
$p_i = p(\xi = x_i)$	p_1	p_2	$\cdots$	p_n

En nuestro ejemplo:

Nº de caras	0	1	2	3
Probabilidad	1/8	3/8	3/8	1/8

Cuya representación gráfica sería:

Entre las propiedades de la función masa de probabilidad cabe destacar:

a. $p_i \geq 0 \quad \forall i$
b. $p_1 + p_2 + ... + p_n = 1$

Ejercicio: Calcula la función masa de probabilidad del experimento consistente en tirar dos dados y sumarlos.

Ejercicio: Dada la siguiente función masa de probabilidad de la variable aleatoria X.

x_i	1	2	3	4	5
p_i	***a***	0,2	***b***	***c***	0,15

Calcula ***a, b y c*** sabiendo que $p(x \leq 2) = 0{,}35$ y $p(x > 3) = 0{,}55$

Función de Distribución asociada a una Variable Aleatoria Discreta.

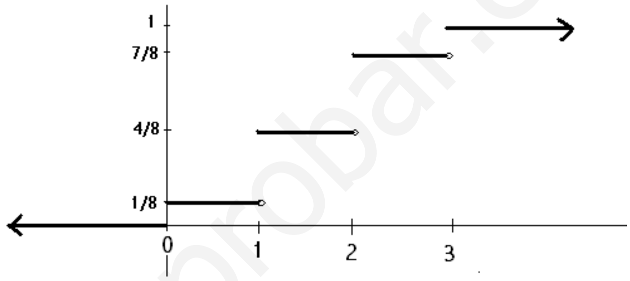
La nueva función que vamos a definir es una generalización del concepto de distribución de frecuencias relativas acumuladas en el caso discreto.

La Función de Distribución de una V.A.D. es una función real de variable real definida de la siguiente manera:

$$F : R \longrightarrow [0,1]$$
$$x \longrightarrow F(x) = p(\xi \le x) = \sum_{xi \le x} p(\xi = x_i)$$

En nuestro ejemplo de la tirada de tres monedas la función de distribución sería:

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ \frac{1}{8} & si \quad 0 \le x < 1 \\ \frac{4}{8} & si \quad 1 \le x < 2 \\ \frac{7}{8} & si \quad 2 \le x < 3 \\ 1 & si \quad x \ge 3 \end{cases}$$



Conocida la función masa de probabilidad queda determinada la función de distribución y viceversa, mediante la relación : $p(\xi = x_k) = F(x_k) - F(x_{k-1})$

Propiedades de la función de distribución:

1. Como F(x) es un probabilidad: $0 \le F(x) \le 1 \quad \forall x \in R$
2. $\lim_{x \to -\infty}$ F(x) = 0 y $\lim_{x \to \infty}$ F(x) = 1
3. En una variable aleatoria discreta (v.a.d.) el valor de F(x) es constante entre cada dos valores consecutivos de la variable, por eso resulta ser una función escalonada.
4. Es una función no decreciente: $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2) \quad \forall x_1, x_2 \in R$
5. $p(a < \xi \le b) = F(b) - F(a)$

Ejercicio: Indica y representa la función de distribución en el experimento de la tirada de dos dados.

Ejercicio: Indica y representa la función de distribución para el número de caras que se obtienen en el tirada de dos monedas.

Ejercicio: Dada la siguiente función masa de probabilidad:

x_i	3	5	7	9
p_i	0,2	a	0,3	0,25

Indica y representa su función de distribución.

Características de una Variable Aleatoria Discreta:

Esperanza Matemática o Media:

La esperanza matemática de una variable aleatoria es una generalización del concepto de media aritmética.

Si ξ es una Variable Aleatoria Discreta, se define la esperanza matemática de ξ como:

$E(\xi) = \sum_i x_i \bullet p(\xi = x_i)$ (También se puede denotar como μ)

Ejercicio: Calcula la esperanza del número de puntos obtenidos en la tirada de un dado.

Ejercicio: Calcula la esperanza del número de caras en la tirada de tres monedas.

Nota: Recuerda que la media de una muestra de datos venía dada por:

$$\bar{x} = \sum_i x_i \bullet f_i = \sum_i \frac{x_i \bullet n_i}{N}$$

Varianza y Desviación Típica:

La Varianza de una variable aleatoria será una generalización del concepto de varianza para una muestra de datos. La Desviación típica es la raíz cuadrada de la Varianza.

Si ξ es una variable aleatoria con esperanza μ (o $E(\xi)$), se llama Varianza de ξ a :

$$\sigma^2 = \text{var}(\xi) = \sum_i (x_i - \mu) \bullet p_i = \left(\sum_i x_i^2 \bullet p_i \right) - \mu^2 \quad \text{(es decir } \sigma^2 = E(\xi^2) - (E(\xi))^2 \text{)}$$

La Desviación típica es la raíz cuadrada de la Varianza y se denota por σ.

Ejercicio: Calcula la Varianza y la Desviación Típica de la variable aleatoria discreta definida como el "número de puntos obtenidos en la tirada de un dado"

Nota: Recuerda que la desviación típica y la varianza en una muestra de datos venia dada por:

$$s^2 = \sum_i (x_i - \bar{x})^2 \bullet f_i = \left(\sum_i x_i^2 \bullet f_i \right) - \bar{x}^2 \quad \text{donde} \quad f_i = \frac{n_i}{N} \text{ frec. relativa.}$$

$$s = \sqrt{s^2}$$

Ejercicio: La variable aleatoria discreta X viene dada por la función de probabilidad siguiente:

X	2	3	5	6	8
p	0,2	0,1	0,4	0,2	0,1

- Calcula la media y la desviación típica de la variable X
- Indica y representa su función de distribución.
- Calcula $p(X \le 5)$; $p(X > 6)$; $p(X < 2)$; $p(2 \le X < 6)$; $p(X < 10)$

Distribución Binomial (Es un caso de distribución de una v.a.d)

Supongamos un experimento aleatorio con las siguientes características:

1. En cada prueba del experimento sólo son posibles dos resultados, el suceso *A* y el suceso contrario A^*.
2. El resultado obtenido en cada prueba es independiente de los anteriores.
3. La probabilidad del suceso A es constante y por lo tanto no varía de una prueba a otra. Sea **p = p(*A*)** y **q = p(** A^* **) (**evidentemente p + q = 1)

Un experimento que tenga esas características se dice que sigue el modelo de la Distribución Binomial.

La variable que expresa el número de veces que ocurre *A* la llamaremos Variable Aleatoria Binomial y la representaremos por **B(n,p)** siendo **n** el número de veces que se hace el experimento y **p = p(A).**

Ejemplo: Una compañía de tabacos determinó que el porcentaje de fumadores en una ciudad es del 30%. Se toma un muestra de 10 personas. Comprueba si la variable X = ”número de personas de la muestra que fuman” sigue una Distribución Binomial.

1. En cada prueba sólo son posibles dos resultados: A = “fumador”; A^* = ”no fumador”
2. El resultado obtenido en la pregunta “¿ Fuma o No Fuma?” , en cada individuo es independiente de los otros.
3. La probabilidad del suceso *A*, **p = p(A) = 0,3** es constante en todos los casos y no varía de una persona a otra.

Por tanto $X \in B(10;0,3)$, ya que **n= 10** , es el número de preguntas (pruebas) que hago

Función Masa de probabilidad de la Distribución Binomial:

Realizamos **n** pruebas del experimento y queremos saber la probabilidad de que el suceso *A* que tiene una probabilidad **p** en cada prueba, salga exactamente **r** veces.

1. Para que se dé **r** veces el suceso *A*, se tiene que dar **n-r** veces el suceso A^*.
   Uno de los caso sería $\underbrace{A\cap A\cap\ldots\cap A}_{r \text{ veces}}\cap\underbrace{A^*\cap A^*\cap\ldots\cap A^*}_{n-r \text{ veces}}$
   La probabilidad de este suceso será $p\cdot p\cdot\ldots\cdot p\cdot q\cdot q\cdot\ldots\cdot q = p^r\cdot q^{n-r}$
2. Las distintas maneras en las que pueden salir **r** veces *A* y **n-r** veces A^* será el número $\binom{n}{r} = \dfrac{n!}{r!\cdot(n-r)!}$
3. Si representamos por *X* la variable aleatoria **B(n,p)** concluimos:

$$\text{P(obtener } \mathbf{r} \text{ veces el suceso } A) = p(X = r) = \binom{n}{r}\cdot p^r\cdot q^{n-r}$$

Ejemplo: Continuando con el ejemplo anterior de los fumadores y no fumadores, calcular la probabilidad de que de las 10 personas de la muestra fumen 4.

Sea X = "Número de personas que fuman"
p = 0,3 , q = 0,7 , n = 10 , r = 4 y $X \in B(10;0{,}3)$

$$p(X=4)=\binom{10}{4}\cdot(0{,}3)^4\cdot(0{,}7)^6 = 0{,}2$$

Ejercicio: Suponiendo que la probabilidad de ser varón es de 0,51; calcula la probabilidad de que una pareja con 6 hijos tenga:

a. Dos varones.
b. Una niña.
c. Por lo menos un varón.
d. Por lo menos una niña.
e. Tres niños y tres niñas.
f. Como mucho dos niñas.

Ejercicio: Una prueba de inteligencia está compuesta por 10 preguntas, cada una con 4 respuestas de las que sólo una es correcta. Un alumno tiene prisa y decide contestar al azar. Calcula:

a. Probabilidad de acertar exactamente 4 preguntas.
b. Probabilidad de no acertar ninguna.
c. Probabilidad de acertar todas.
d. Probabilidad de acertar por lo menos 8.
e. Probabilidad de acertar como mucho 3.
f. Probabilidad de fallar 3.

Esperanza, Varianza y Desviación Típica de la Distribución Binomial:

Sea $X \in B(n,p)$ se verifica que:

$$\mu = E(X) = n\bullet p \qquad \sigma^2 = \mathrm{var}(X) = n\bullet p\bullet q \qquad \sigma = \sqrt{n\bullet p\bullet q}$$

Ejercicio: La probabilidad de éxito de una determinada vacuna antigripal es 0,72. Calcula la probabilidad de que una vez administrada a 15 pacientes:

a. Ninguno sufra la gripe.
b. Todos tengan la gripe.
c. Dos tengan la gripe.
d. De los 15 pacientes, ¿Cuántos se esperan que no cojan la gripe?

Ejercicio: Se sabe que la probabilidad de que un alumno apruebe la P.A.U. en la convocatoria de Junio es de un 80%. Si un Instituto presenta 110 alumnos a dicha prueba, ¿Cuántos alumnos cabe esperar que aprueben?. Calcula también la desviación típica.

Ejercicio: En una frutería, el 25% de las manzanas están podridas. Si escogemos 4 manzanas al azar, calcula:

a. La función masa de probabilidad de la V.A.D. X = "Número de manzanas podridas". Indica también su función de distribución.
b. Probabilidad de que tres manzanas estén sanas.
c. ¿Cuántas manzanas se esperan que estén podridas de las 4 que he escogido"
d. Probabilidad de que ninguna esté podrida.
e. Probabilidad de que al menos dos estén sanas.

Variable Aleatoria Continua.

Una variable $\xi : E \longrightarrow R$ es continua cuando puede tomar, todos los valores posibles dentro de un cierto intervalo de la recta real.

Por ejemplo, en el experimento que consiste en elegir un alumno al azar de todo el instituto, el espacio muestral E está formado por todos los alumnos del instituto. La variable aleatoria X que asocia a cada alumno su estatura, es continua, ya que puede tomar infinitos valores, por ejemplo, del intervalo [1,50 – 1,95].

Como X puede tomar infinitos valores en un intervalo, no tiene sentido hablar de la probabilidad en un punto, puesto que esta es nula, es decir p(X = a) = 0.

Sin embargo, sí interesa conocer las probabilidades correspondientes a intervalos.

Nota: La función análoga a la función masa de probabilidad del caso discreto, se denomina función de densidad en el caso continua, y es una función continua a partir de la cual, al igual que en el caso discreto, obtenemos la función de distribución, pero en el caso continuo no vamos a trabajar con estas funciones pues es necesario ver antes el tema de derivadas e integrales.

Distribución Normal (Es un caso de una Variable Aleatoria Continua)

Existe una cantidad considerable de fenómenos naturales, tales como la estatura o el peso de una persona, la frecuencia cardiaca, la altura alcanzada por un árbol de una determinada especie, etc., que una vez estudiados presentan ciertas características comunes, tales como la aproximación de los valores a la media, la simetría de estos valores con respecto a la media, pocos valores alejados de la media, etc.

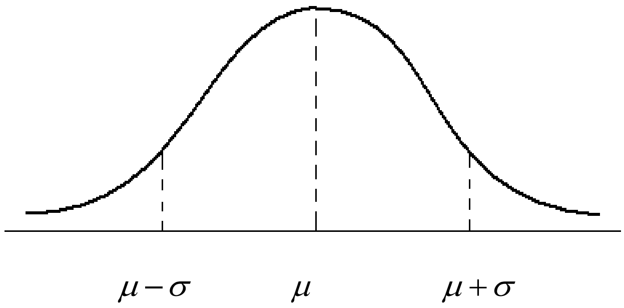
Este comportamiento es considerado "normal" y una vez representada la correspondiente función de densidad, esta adquiere una forma típicamente acampanada, y se conoce como la Campana de Gauss.

Se dice que una variable aleatoria ξ tiene una Distribución Normal de media μ y desviación típica σ si su función de densidad es:

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

Se dice que $\xi \in N(\mu, \sigma)$ y evidentemente $E(\xi) = \mu$ y $\mathrm{var}(\xi) = \sigma^2$

La gráfica de ***f(x)*** es de la forma:



Los valores de la función de distribución de la normal no son fáciles de calcular usando integrales, pero tampoco nos será necesario pues existen unas tablas de la función de distribución de la $N(0,1)$ y serán suficientes, pues cualquier otra distribución normal $N(\mu,\sigma)$ se puede pasar a una $N(0,1)$, mediante un sencillo cambio que se conoce como **tipificación**.

Manejo de las tablas de la distribución Normal $N(0,1)$

La tabla de la *N(0,1)* nos da directamente el valor de $F(a) = p(X \leq a) \quad a \geq 0$
Para calcular las demás probabilidades debemos ponerlas en función de probabilidades como la anterior que ya conocemos. Veamos los diferentes casos:

1. $p(X \geq a) = 1 - p(X \leq a) = 1 - F(a)$
2. $p(X \leq -a) = p(X \geq a) = 1 - p(X \leq a) = 1 - F(a)$
3. $p(X \geq -a) = F(a)$
4. $p(a \leq X \leq b) = p(X \leq b) - p(X \leq a) = F(b) - F(a)$
5. $p(-a \leq X \leq b) = p(X \leq b) - p(X \leq -a) = F(b) - (1 - F(a))$
6. $p(-a \leq X \leq -b) = p(X \leq -b) - p(X \leq -a) = (1 - F(b)) - (1 - F(a)) = F(a) - F(b)$

Ejercicio: Sea $X \in N(0,1)$, Calcula:

a. $p(X < 1{,}57)$
b. $p(0{,}5 < X \leq 1{,}3)$
c. $p(X > -2{,}15)$
d. $p(-1{,}2 \leq X < -0{,}5)$
e. $p(X \leq 2)$
f. $p(-1{,}46 < X < 0{,}75)$
g. $p(X \leq -1{,}11)$
h. $p(X > 0)$

Tipificación de una Variable $N(\mu,\sigma)$

Como ya dijimos, la única distribución normal de la que disponemos una tabla es la *N(0,1)*, que tiene media $\mu = 0$ y desviación típica $\sigma = 1$.

Si $X \in N(\mu,\sigma)$ entonces $Z = \dfrac{X-\mu}{\sigma}$ es una variable que cumple $Z \in N(0,1)$

Ejemplo: Sea $X \in N(12,2)$, calcula $p(X < 15)$ y $p(10 < X \leq 17)$

Solución:

$$p(X < 15) = p(Z < \frac{15-12}{2}) = p(Z < 1{,}5) = 0{,}9332$$

$$p(10 < X \leq 17) = p(\frac{10-12}{2} < Z \leq \frac{17-12}{2}) = p(-1 < Z \leq 2{,}5) = F(2{,}5) - (1 - F(1)) =$$
$$= 0{,}9938 - (1 - 0{,}8413) = 0{,}8351$$

Ejercicio: Sea $X \in N(30,4)$, calcula:

a. $p(X < 35)$
b. $p(X \geq 20)$
c. $p(22 < X < 34)$
d. $p(X < 30)$
e. $p(X \leq 24)$
f. $p(32 < X \leq 36)$

Ejercicio: El número de aprobados de una asignatura se distribuyen normalmente con media 15 y desviación típica 6. Calcula la probabilidad de que haya más de 20 aprobados.

Ejercicio: Se aplica a 300 alumnos de la E.S.O. un test de agresividad y se observa que los resultados se distribuyen normalmente con media 30 y desviación típica 12. Se pide:

a. ¿Qué proporción de alumnos tendrá una puntuación entre 20 y 35?
b. ¿Cuántos alumnos tendrán una puntuación superior a 42?

Ejercicio: Una persona viaja diariamente de su casa a la oficina y se sabe que el tiempo que tarda en dicho viaje se distribuye según una Normal con $\mu = 35{,}5$ y $\sigma = 3{,}11$ minutos. Si sale de su casa todos los días a las 8:20 horas y debe estar en la oficina a las 9:00 horas, ¿cuántos días al año se espera que llegue tarde si hace 240 viajes anuales?

Ejercicio: La cantidad de café depositada en cada bolsa por una máquina envasadora automática sigue una distribución normal de media $\mu = 1040$ gramos y desviación típica $\sigma = 50$ gramos.

a. Calcula el tanto por ciento de paquetes que contienen más de 1 Kg.
b. Calcula la probabilidad de que un paquete tenga un peso comprendido entre 950 gramos y 1050 gramos.
c. Calcula α, sabiendo que el 97,5% de los paquetes contienen menos de α gramos.

Ejercicio: En una ciudad, la temperatura máxima durante el mes de Junio está distribuida normalmente con media 26º y una varianza de 16 . Calcula el número de días que se espera que tengan una temperatura máxima comprendida entre 22º y 28º.

Aproximación de la Distribución Binomial a la Distribución Normal

La Importancia de la distribución normal viene dada por la gran cantidad de experimentos (variables aleatorias) que siguen una distribución normal, pero también por la gran cantidad de variables aleatorias que se pueden aproximar a un normal, entre ellas la distribución binomial *B(n,p)* que se puede aproximar por una $N(n\cdot p; \sqrt{n\cdot p\cdot q})$.

En la distribución Binomial *B(n,p)* si **n** es muy grande, entonces los cálculos son muy largos y laboriosos, por tanto:

Dada una variable $X \in B(n, p)$, sabemos que $\mu = E(x) = n\cdot p$ y $\sigma = \sqrt{n\cdot p\cdot q}$, entonces la variable $Y \in N(n\cdot p; \sqrt{n\cdot p\cdot q})$ es una buena aproximación de *X*.

Esta aproximación será mejor cuanto mayor sea **n** y para valores de **p** y **q** no muy próximos a cero. Se considera que si **n·p** > 5 y **n·q** >5 la aproximación es bastante buena.

Importante: Al realizar una aproximación de una binomial a una normal, debemos tener en cuanta que la variable de partida es una variable aleatoria discreta, por lo tanto $p(2 < X < 3)$ no es lo mismo que $p(2 \leq X \leq 3)$, cosa que si ocurría en una variable continua. De la misma forma $p(X = 4)$ tampoco es cero, cosa que ocurría en una variable continua. Por tanto es necesario usar un **factor de corrección**, y así:

Sea $X \in B(n, p)$ e $Y \in N(n{\cdot}p; \sqrt{n{\cdot}p{\cdot}q})$ una aproximación de *X*, entonces:

- $p(X = a) = p(a - 0{,}5 \leq Y \leq a + 0{,}5)$
- $p(a \leq X \leq b) = p(a - 0{,}5 \leq Y \leq b + 0{,}5)$
- $p(a \leq X < b) = p(a - 0{,}5 \leq Y \leq b - 0{,}5)$
- $p(a < X \leq b) = p(a + 0{,}5 \leq Y \leq b + 0{,}5)$
- $p(a < X < b) = p(a + 0{,}5 \leq Y \leq b - 0{,}5)$
- los demás casos se corrigen de la misma forma.

Ejemplo: Se lanza 15 veces un dado. Calcula la probabilidad de salir entre 7 y 10 veces (ambos inclusive) número par.

Sea X=”número de veces que sale par” , $X \in B(15;0{,}5)$

***********Vamos a resolverlo directamente, usando la binomial.

$$p(7 \leq X \leq 10) = p(X = 7) + p(X = 8) + p(X = 9) + p(X = 10) = \binom{15}{7}\cdot\left(\frac{1}{2}\right)^7\cdot\left(\frac{1}{2}\right)^8 +$$

$$\binom{15}{8}\cdot\left(\frac{1}{2}\right)^8\cdot\left(\frac{1}{2}\right)^7 + \binom{15}{9}\cdot\left(\frac{1}{2}\right)^9\cdot\left(\frac{1}{2}\right)^6 + \binom{15}{9}\cdot\left(\frac{1}{2}\right)^{10}\cdot\left(\frac{1}{2}\right)^5 = \left[\binom{15}{7} + \binom{15}{8} + \binom{15}{9} + \binom{15}{10}\right]\cdot\frac{1}{2^{15}} =$$

$$= \frac{6435 + 6435 + 5005 + 3003}{32768} == \frac{20878}{32768} = 0{,}6371459$$

***********Usemos ahora la aproximación a una normal.

$$X \in B(15; \frac{1}{2}) \quad \Rightarrow \quad Y \in N(15\cdot\frac{1}{2}; \sqrt{15\cdot\frac{1}{2}\cdot\frac{1}{2}}) \quad \Rightarrow \quad Y \in N(7{,}5;1{,}9364)$$

Entonces:

$$p(7 \leq X \leq 10) = p(6{,}5 \leq Y \leq 10{,}5) = p(\frac{6{,}5 - 7{,}5}{1{,}9364} \leq Z \leq \frac{10{,}5 - 7{,}5}{1{,}9364}) = p(-0{,}51 \leq Z \leq 1{,}55) =$$

$$F(1{,}55) - F(-0{,}51) = F(1{,}55) - (1 - F(0{,}51)) = 0{,}9394 - 1 + 0{,}6950 = 0{,}6344$$

Como podemos ver el valor obtenido usando la aproximación a una normal, es “bastante” aproximado al valor real obtenido directamente con la binomial.

Ejercicio: Se sabe que el 25% de los alumnos de un instituto, van a clases particulares de matemáticas.

a. Si tomamos 4 alumnos al azar, calcula la probabilidad de que 3 de ellos vayan a clases particulares
b. Si tomamos una muestra de 100 alumnos, calcula la probabilidad de que vayan a clases particulares entre 20 y 35 alumnos

Ejercicio: En un centro comercial se sabe que el 35% de los clientes pagan con tarjeta.

a. Si en una caja han pagado 120 clientes, ¿cuál es el número esperado de clientes que no han pagado.
b. Si en una caja han pagado 200 clientes, ¿cuál es la probabilidad de que hayan pagado con tarjeta entre 60 y 85 clientes?.
c. Si en una caja han pagado 400 clientes, ¿cuál es la probabilidad de que al menos 260 no lo hayan hecho con trajeta?

(P.A.U – Junio 2004)

Ejercicio: Una de las pruebas de acceso a la universidad para personas mayores de 25 años consiste en un test con 100 preguntas, cada una de las cuales con dos posibles respuestas, siendo sólo una de ellas correcta. Para superar esta pruebe debe obtenerse, al menos, 60 respuestas correctas. Si una persona contesta al azar:

a. ¿Cuál será el número esperado de respuestas correctas?
b. ¿Qué probabilidad tendrá de superar la prueba?

(P.A.U. – Septiembre 2003)

Ejercicio: Se sabe, tras varios sondeos, que en una determinada población únicamente el 15% es favorable a los tratamientos de psicoterapia. Elegida, al azar, una muestra de 50 personas, se desea saber:

a. La probabilidad de que haya entre 10 y 20 personas favorables a dichos tratamientos.
b. La probabilidad de que haya más de 5 personas favorables a los tratamientos.

(P.A.U. – Septiembre 1997)

Ejercicio: Un estudio indica que la proporción de individuos que enfermarán después de suministrarle una determinada vacuna es del 5%. Se toma una muestra de 400 individuos vacunados. Determinar:

a. El número esperado de individuos que no enfermarán.
b. La probabilidad de que el número de individuos que enferman sea, como mínimo, igual a 24.
c. Determinar la probabilidad de que el número de individuos que no enferman sea, como mínimo 372.

(P.A.U. – Septiembre 2001)

EJERCICIOS PROPUESTOS:

1. Una variable aleatoria discreta tiene la siguiente función de probabilidad:

x_i	0	1	2	3	4
p_i	0,1	a	b	c	0,2

Sabiendo que $p(X \leq 2) = 0{,}7$ y que $p(X \geq 2) = 0{,}75$:

   a. Encuentra su esperanza matemática y su desviación típica.
   b. Obtén la función de distribución

2. La probabilidad de que un estudiante de 1º curso obtenga el título de licenciado en Geografía e Historia es de 0,3. Encuentra la probabilidad de que de un grupo de 7 estudiantes matriculados en 1º curso:
   a. Ninguno de los 7 acabe la carrera
   b. Todos obtengan el título
   c. Por lo menos dos terminen la carrera.
   d. Encuentra también la media y la desviación típica del número de alumnos que terminan la carrera.

3. En un estudio hecho por TVE, se pudo saber que sólo el 15% de los españoles son partidarios de que se retransmitan combates de boxeo. Elegida una muestra aleatoria formada por 10 personas, se pide:
   a. ¿Cuál es la probabilidad de que exactamente la mitad sea favorable?
   b. ¿Cuál es la probabilidad de que por lo menos la mitad sea favorable?
   c. La probabilidad de que sólo uno sea favorable
   d. La probabilidad de que al menos 9 sean favorables

4. La distribución de probabilidad de una variable aleatoria discreta viene dada por:

X	1	2	3	4	5
p(x)	0,15	0,25	0,2	m	0,15

   a. Halla "m" para que se trate de una función de probabilidad.
   b. Calcula y representa gráficamente su función de distribución.
   c. Halla $p(X \leq 4)$ y $p(2 \leq X < 4)$

P.A.U – Junio 1998

5. Después de aplicar un test de aptitud numérica a un grupo de alumnos de E.S.O., se detectó que el 65% tiene una aptitud numérica que podemos considerar inaceptable. ¿Cuál es la probabilidad de que de un grupo de 5, por lo menos dos tengan una capacidad numérica aceptable?

6. En una zona geográfica determinada, el 60% de los votos fueron para el partido A. Si se consideran a 5 votantes al azar de dicha zona, se pide:
   a. Probabilidad de que votaran exactamente tres a dicho partido.
   b. Probabilidad de que ninguno lo votase.
   c. Probabilidad de que no lo votasen todos.

7. La probabilidad de que un alumno supere el examen de selectividad es de 0,8.
   a. Calcula la probabilidad de que de un grupo de 8 alumnos, por lo menos aprueben 2.
   b. De un instituto se presentan exactamente 100 alumnos. Calcula.

i. Probabilidad de que por lo menos aprueben 80.
ii. Probabilidad de que aprueben más de 90.
iii. Probabilidad de que no aprueben ni siquiera la mitad

8. Los pesos de los pollos de una granja se distribuyen normalmente con una media de 1 kg. 800 gr. y una desviación típica de 300 gr. Si se rechazan los que pesen menos de 1 kg. 500 gr.:
    a. ¿Qué tanto por ciento habrá que rechazar?
    b. ¿Cuál es la probabilidad de que elegido uno al azar, pese más de 2 kg.?.

9. La probabilidad de que la causa de un accidente automovilístico sea el exceso de alcohol es de 0,6. Bajo estas hipótesis se pide:
    a. Si en un fin de semana se producen 10 accidentes, ¿cuál es la probabilidad de que por lo menos 3 de ellos sean debidos al alcohol?
    b. ¿Cuántos de esos 10 accidentes puedes estimar que se evitarían si todos los conductores se mantuviesen sin ingerir alcohol?

10. Una finca produce naranjas con un diámetro medio de 58 mm. y desviación típica 7 mm. Según su diámetro se catalogan como **Extra** las de 62 mm. o más; **Cat. I** las comprendidas entre 55 y 62 mm. y **Cat. II** las de menos de 55 mm. ¿Qué porcentaje produce de cada tipo?

11. Calcula la probabilidad de que al lanzar 100 veces una moneda, el número de caras obtenido esté comprendido entre 45 y 55 (ambos inclusive)

12. Para aprobar unas oposiciones se necesita obtener un mínimo de 100 puntos en una prueba. Por experiencias anteriores se sabe que la distribución de los puntos obtenidos por los opositores sigue una normal de media 110 puntos y desviación típica 15.
    a. ¿Cuál es la probabilidad de que un opositor apruebe?
    b. Si sabemos que hay 1000 opositores y sólo 300 plazas, ¿cuántos puntos de deberán exigir para ajustar el número de plazas al número de opositores?

13. En un gran estadio deportivo se quieren instalar focos para iluminar el terreno de juego. El suministrador asegura que el tiempo de vida de los focos es, aproximadamente, normal con media de 40 horas y desviación típica de 4 horas.
    a. Escogiendo un foco al azar, ¿cuál es la probabilidad de que dure por lo menos 30 horas?
    b. Si se comprueba que sólo 1400 focos duran más de 30 horas, ¿cuántos focos habrá aproximadamente en el estadio?

14. Consideramos tres distribuciones binomiales: B(10; 0,1) , B(200; 0,1) y B(200; 0,5). Explica cuál de ellas se aproxima mejor usando una distribución normal.

15. Los ingresos diarios de una empresa tienen una distribución normal con media 35560 euros y desviación típica 2530 euros. Justifica si es razonable o no esperar obtener un día ventas por un importe superior a 55000 euros. Calcula cuántos días en un año se espera obtener ventas superiores a 40620 euros.

16. El nivel medio de colesterol en sangre en la población adulta entre 50 y 60 años, es de 190 mg. por cada 100 ml. de sangre. La desviación típica es de 25 mg. por 100 ml. Si las medidas se distribuyen aproximadamente de forma normal, calcula:
    a. El porcentaje de la población que tiene niveles superiores a 250 mg., que es cuando considera un nivel preocupante.
    b. La probabilidad de que un individuo tenga un nivel de colesterol comprendido entre 150 mg. y 200 mg.
    c. El porcentaje de los que tienen menos de 130 mg.
    d. Si se sabe que el 60% de esa población tiene más de α mg., calcula α.

17. El peso medio de las almejas recogidas en la ría de Noia sigue una distribución normal de media 20 gr. y desviación típica 4 gr. Si a una fábrica conservera sólo le sirven aquellas que pesen entre 18 y 26 gr. ¿Qué proporción de almejas cumplen esta característica?.
    Por otra parte, un conocido restaurante de Noia sólo compra almejas que pesen por lo menos 30 gr., ¿de cuántos quilos podrá disponer un día en el que se han recogido 500 kg.?

18. La probabilidad de que un reloj salga defectuosa de fábrica es del 5%. En un lote de 6 relojes , ¿cuál es la probabilidad de que alguno sea defectuoso?, ¿y en un lote de 200 relojes?. ¿Cuál sería el número de relojes defectuosos esperados en un lote de 1000 relojes?

19. Al elegir 100 personas de una población resultó que su talla media era de 170 cm. y su desviación típica 10 cm. Si dichos datos se distribuyen normalmente, ¿cuántas de estas personas miden entre 190 y 215 cm?, ¿y a lo sumo 160 cms?

    (P.A.U. – Septiembre 1997)

20. Sabemos que el tiempo de espera en la cola de una sucursal de un banco se distribuye normalmente con media 15 minutos y desviación típica de 5 minutos. Si tomamos a 40 clientes que hoy han sido atendidos, se pide:
    a. ¿Cuál es la probabilidad de que el tiempo medio que han tenido que esperar sea menor que 17 minutos?
    b. ¿Cuál es la probabilidad de que se encuentre entre 12 y 16 minutos?
    c. Entre que valores se halla la media de tiempos de espera con una seguridad del 40%. (pertenece al próximo tema: Muestreo)

    (P.A.U. – Junio 1997)

21. La probabilidad de que un alumno matriculado en 2º Curso de Bachillerato abandone los estudios es de 0,2. Si en un centro hay 100 alumnos de ese nivel, se pide:
    a. ¿De que distribución se trata?. Razona la respuesta.
    b. ¿Qué condiciones debe cumplir para que se pueda aproximar a una continua?
    c. Halla la probabilidad de que abandonen menos de 30 alumnos.
    d. Halla la probabilidad de que abandonen entre 10 y 20 alumnos.

    (P.A.U. – Septiembre 2000)

22. El peso de las piñas tropicales cultivadas en una determinada finca siguen una variable normal de media 1,4 Kg. y desviación típica 0,6 kg. Si en la presente cosecha se han recogido un total de 325000 kg. de piña, determinar:
    a. La cantidad de piñas tropicales que pesan más de 1,6 kg.
    b. La probabilidad de que una piña pese entre 1,3 y 1,5 kg.
    c. La cantidad de piñas cuyo peso difiere medio kilo de la media. (PAU – Sept. 2001)