Estadística unidimensional: tablas, gráficos y parámetros estadísticos.

Frecuencias y tablas.

Distribución de frecuencias

La **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

Tipos de frecuencias

Frecuencia absoluta

La **frecuencia absoluta** es el **número de veces** que aparece un determinado **valor** en un estudio estadístico.
Se representa por **f_i**.
La **suma de las frecuencias absolutas** es igual al número total de datos, que se representa por **N**.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

Para indicar resumidamente estas sumas se utiliza la letra griega **Σ** (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^{i=n} f_i = N$$

Frecuencia relativa

La **frecuencia relativa** es el **cociente** entre la f**recuencia absoluta** de un determinado valor y el **número total de datos**.
Se puede expresar en tantos por ciento y se representa por **n_i**.

$$n_i = \frac{f_i}{N}$$

La suma de las frecuencias relativas es igual a 1.

Frecuencia acumulada

La **frecuencia acumulada** es la **suma de las frecuencias absolutas** de todos los **valores inferiores o iguales** al **valor** considerado.
Se representa por **F_i**.

Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**. Se puede expresar en tantos por ciento.
Ejemplo:
Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:
32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.
En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

x_i	Recuento	f_i	F_i	n_i	N_i
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097
29	~~IIII~~ I	6	9	0.194	0.290
30	~~IIII~~ II	7	16	0.226	0.516
31	~~IIII~~ III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Este tipo de **tablas de frecuencias** se utiliza con **variables discretas**.

Distribución de frecuencias agrupadas

La **distribución de frecuencias agrupadas** o **tabla con datos agrupados** se emplea si las **variables** toman un **número grande de valores** o la **variable es continua**.
Se **agrupan** los **valores** en **intervalos** que tengan la **misma amplitud** denominados **clases**. A cada **clase** se le asigna su **frecuencia correspondiente**.

Límites de la clase

Cada **clase** está **delimitada** por el **límite inferior de la clase** y el **límite superior de la clase**.

Amplitud de la clase

La **amplitud de la clase** es la **diferencia** entre el **límite superior e inferior** de la **clase**.

Marca de clase

La **marca de clase** es el **punto medio** de cada **intervalo** y es el **valor** que representa a todo el **intervalo** para el **cálculo** de algunos **parámetros**.

Construcción de una tabla de datos agrupados

3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

1º Se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

2º Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos queramos establecer.

Es conveniente que el número de intervalos oscile entre 6 y 15.

En este caso, 48 - 3 = 45, incrementamos el número hasta 50 : 5 = 10 intervalos.

Se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece intervalo, se cuenta en el siguiente intervalo.

	c_i	f_i	F_i	n_i	N_i
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10, 15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.275
[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	42.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1
		40		**1**	

Representaciones gráficas.

Diagrama de barras

Un **diagrama de barras** se utiliza para de presentar **datos cualitativos** o **datos cuantitativos de tipo discreto (no agrupados en intervalos)**.

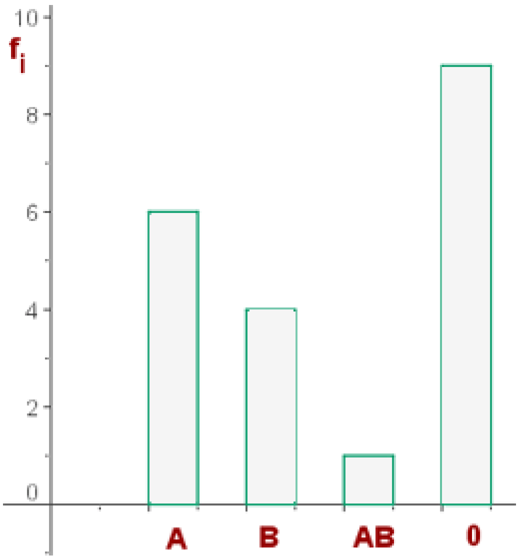
Se representan sobre unos ejes de coordenadas, en el **eje de abscisas** se colocan los **valores de la variable**, y sobre el **eje de ordenadas** las **frecuencias absolutas, frecuencias relativas o frecuencias acumuladas (absolutas o relativas)**.

Los **datos** se representan mediante **barras** de una **altura proporcional** a la **frecuencia**.

Ejemplo: Un estudio hecho al conjunto de los 20 alumnos de una clase para determinar su grupo sanguíneo ha dado el siguiente resultado:

Grupo sanguíneo	f_i
A	*6*
B	*4*
AB	*1*
0	*9*
	20

Solución:



Polígono de frecuencias

Un **polígono de frecuencias** se forma uniendo los **extremos** de las **barras** mediante **segmentos**.
También se puede realizar trazando los **puntos** que representan las **frecuencias** y uniéndolos mediante **segmentos**.
Ejemplo:
Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

Hora	Temperatura
6	7º
9	12°
12	14°
15	11°
18	12°
21	10°
24	8°

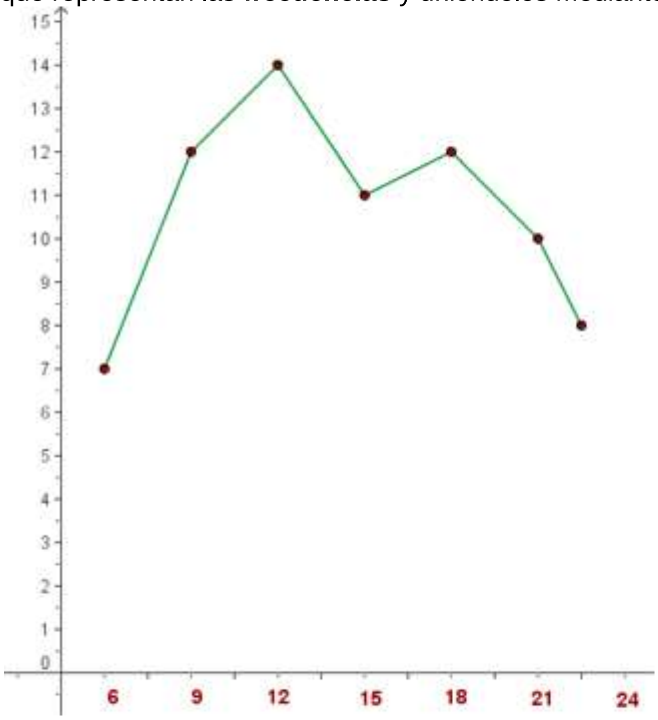


Diagrama de sectores

Un **diagrama de sectores** se puede utilizar para todo tipo de *variables*, pero se usa frecuentemente para las **variables cualitativas**.
Los **datos** se representan en un **círculo**, de modo que el **ángulo** de cada **sector** es **proporcional** a la **frecuencia absoluta** correspondiente.

$$\alpha = \frac{360°}{N} \cdot f_i$$

El diagrama circular se construye con la ayuda de un transportador de ángulos.
Ejemplo:
En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 9 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360°}{30} \cdot 12 = 144° \qquad \alpha_2 = \frac{360°}{30} \cdot 3 = 36°$$

$$\alpha_3 = \frac{360°}{30} \cdot 9 = 108° \qquad \alpha_4 = \frac{360°}{30} \cdot 6 = 72°$$

	Alumnos	Ángulo
Baloncesto	12	144°
Natación	3	36°
Fútbol	9	108°
Sin deporte	6	72°
Total	30	360°

Histograma

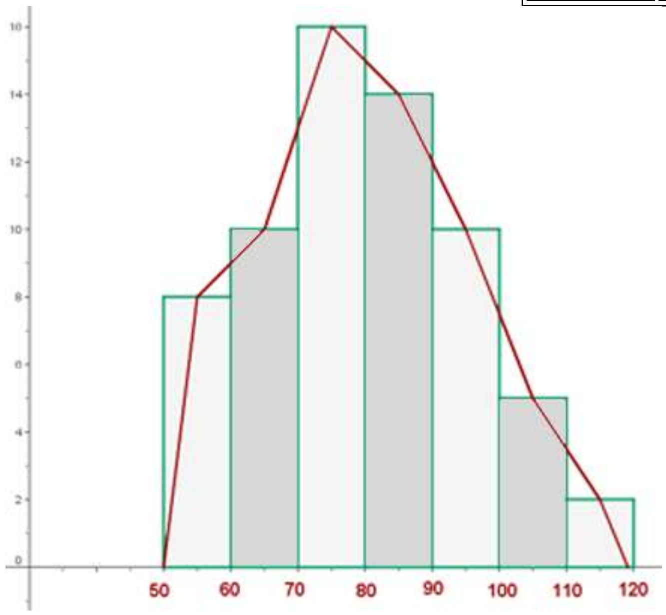
Un **histograma** es una **representación gráfica** de una **variable** en forma de **barras**.
Se utilizan para **variables continuas** o para **variables discretas**, con un gran número de datos, y que se han agrupado en **clases**.
En el **eje abscisas** se construyen unos **rectángulos** que tienen por **base la amplitud del intervalo**, y por **altura**, si los intervalos son de la misma amplitud la **frecuencia** de cada **intervalo**; si los intervalos son de distinta amplitud el cociente entre la **frecuencia** de cada intervalo entre la **amplitud** del mismo. (De forma que la **superficie** de cada **barra** sea **proporcional** a la **frecuencia** de los **valores** representados).

Polígono de frecuencia (de variables continuas o discretas agrupadas en intervalos)

Para construir el **polígono de frecuencia** se toma la **marca de clase** que coincide con el **punto medio** de cada **rectángulo**.

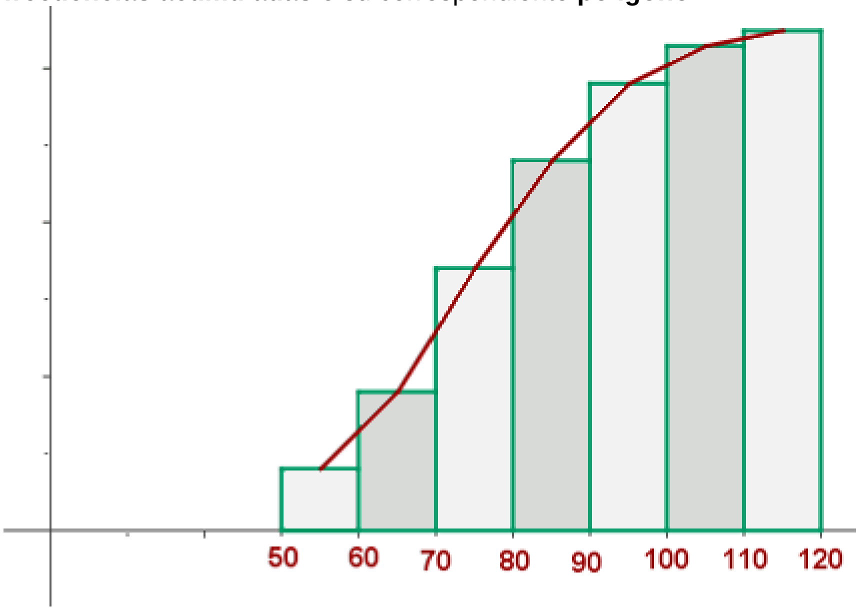
Ejemplo: El peso de 65 personas adultas viene dado por la siguiente tabla:

	c_i	f_i	F_i
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34
[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	105	5	63
[110, 120)	115	2	65
		65	



Histograma y polígono de frecuencias acumuladas

Si se representan las **frecuencias acumuladas** de una **tabla de datos agrupados** se obtiene el **histograma de frecuencias acumuladas** o su correspondiente **polígono**.



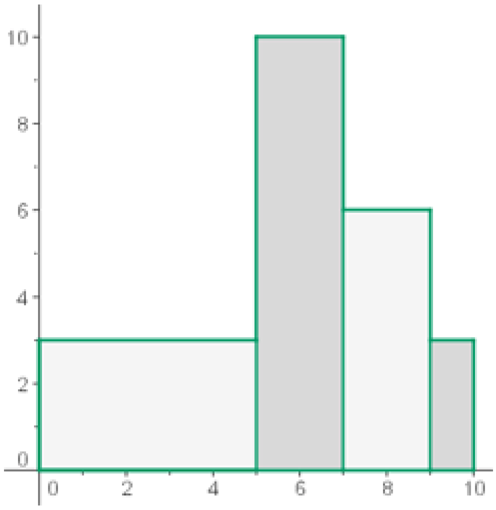
Histogramas con intervalos de amplitud diferente

Para **construir** un **histogramas** con **intervalo de amplitud diferente** tenemos que **calcular** las **alturas** de los **rectángulos** del **histograma**.

$$h_i = \frac{f_i}{a_i}$$

h_i es la altura del intervalo.
f_i es la frecuencia del intervalo.
a_i es la amplitud del intervalo.
Ejemplo:
En la siguiente tabla se muestra las calificaciones (suspenso, aprobado, notable y sobresaliente) obtenidas por un grupo de 50 alumnos.

	f_i	**h_i**
[0, 5)	15	3
[5, 7)	20	10
[7, 9)	12	6
[9, 10)	3	3
	50	



Medidas de centralización, dispersión y simetría.

Medidas de centralización

Nos indican en torno a qué valor (centro) se distribuyen los datos.
Las **medidas de centralización** son:

Media aritmética

La **media** es el valor **promedio** de la distribución.

Mediana

La **mediana** es la **puntación** de la escala que **separa la mitad superior** de la distribución y **la inferior**, es decir divide la serie de datos en **dos partes iguales**.

Moda

La **moda** es el **valor** que **más se repite** en una distribución.

Medidas de centralización

LA MODA

La moda, representada por *Mo,* es otro parámetro de posición que se calcula simplemente como el valor que más se repite en la muestra, es decir, el valor con una mayor frecuencia. En consecuencia, no siempre se sitúa hacia el centro de la distribución.

Puede haber más de una moda en el caso en que dos o más valores de la variable presenten la misma frecuencia. Por otro lado, la moda puede no existir cuando en un conjunto de datos, todos éstos son diferentes entre sí y no hay ningún dato que se repita más de una vez.

La **moda** es el **valor** que tiene **mayor frecuencia absoluta**.
Se representa por **M_o**.
Se puede hallar la **moda** para **variables cualitativas** y **cuantitativas**.
Hallar la **moda** de la distribución: 2, 3, 3, 4, 4, 4, 5, 5 $M_o = 4$

Si en un grupo hay **dos o varias puntuaciones** con la **misma frecuencia** y esa frecuencia es la máxima, la **distribución** es **bimodal** o **multimodal**, es decir, tiene **varias modas**.
1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9 M_o= 1, 5, 9

Cuando todas las **puntuaciones** de un grupo tienen la **misma frecuencia**, **no** hay **moda**.
2, 2, 3, 3, 6, 6, 9, 9

Si **dos puntuaciones adyacentes** tienen la **frecuencia máxima**, la **moda** es el **promedio** de las dos puntuaciones adyacentes.
0, 1, 3, 3, 5, 5, 7, 8Mo = 4

MEDIANA

Es el **valor** que ocupa el **lugar central** de todos los **datos** cuando éstos están **ordenados de menor a mayor**.
La **mediana** se representa por **M**$_e$.
La **mediana** se puede **hallar** sólo para **variables cuantitativas**.

Cálculo de la mediana

1. **Ordenamos** los **datos** de **menor a mayor**.
2. Si la serie tiene un **número impar de medidas** la **mediana** es la **puntuación central** de la misma.

2, 3, 4, 4, 5, 5, 5, 6, 6 Me = 5

3. Si la serie tiene un **número par** de puntuaciones la **mediana** es la **media** entre las dos **puntuaciones centrales**.

7, 8, 9, 10, 11, 12 Me = 9.5

***Por ejemplo**: Sea la variable aleatoria “números de televisores por hogar”. Se realiza una encuesta en 13 hogares, obteniéndose los siguientes resultados:*

3, 4, 2, 3, 2, 1, 1, 2, 1, 1, 2, 1 y 1

Halla la mediana de los mismos.

Solución: El primer paso es ordenar los datos de menor a mayor*: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4*

Como n es 13, impar, la Me (mediana) será igual a 2, de manera que queden 6 datos por debajo y 6 por encima de dicha posición.

Cálculo de la mediana para datos agrupados

La **mediana** se encuentra en el **intervalo** donde la **frecuencia acumulada** llega hasta la **mitad de la suma de las frecuencias absolutas**.

Es decir tenemos que buscar el intervalo en el que se encuentre $\frac{N}{2}$.

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

L$_i$ es el límite inferior de la clase donde se encuentra la mediana.

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas.

F$_{i-1}$ es la **frecuencia acumulada** anterior a la clase mediana.

a$_i$ es la amplitud de la clase.

La **mediana** es **independiente** de las **amplitudes** de los **intervalos**.

Ejemplo: **Calcula** la **mediana** de la distribución estadística que viene dada por la tabla:

	f_i	F_i
[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65
[69, 72)	27	92
[72, 75)	8	100
	100	

100/2 = 50
Clase de la mediana: [66, 69)

$$Me = 66 + \frac{50 - 23}{42} \cdot 3 = 67.93$$

MEDIA ARITMÉTICA

La **media aritmética** es el **valor** obtenido al **sumar** todos los **datos** y **dividir** el resultado entre el **número** total de **datos**.

$\bar{x}$ es el símbolo de la **media aritmética**.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

Ejemplo: Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Halla el peso medio.

$$\bar{x} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = 80 \text{ Kg}$$

Media aritmética para datos agrupados

Si los **datos** vienen **agrupados** en una tabla de frecuencias, la expresión de la **media** es:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{N}$$

Ejercicio de media aritmética

En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. **Calcula la puntuación media**.

	x_i	f_i	$x_i \cdot f_i$
[10, 20)	15	1	15
[20, 30)	25	8	200
[30,40)	35	10	350
[40, 50)	45	9	405
[50, 60	55	8	440
[60,70)	65	4	260
[70, 80)	75	2	150
		42	1 820

$$\bar{x} = \frac{1820}{42} = 43.33$$

Medidas de dispersión

Las **medidas de dispersión** nos informan sobre cuanto se alejan del centro los valores de la distribución.

Las **medidas de dispersión** son:

Rango o recorrido

El **rango** es la **diferencia** entre el **mayor** y el **menor** de los **datos** de una distribución estadística.

Por ejemplo: *Hallar el rango de los datos 2, 9, 8, 9, 15, 21, 5, 20.*

El Rango quedaría 21-2=19.

Desviación media

La **desviación media** es la **media aritmética** de los **valores absolutos** de las **desviaciones** respecto a la **media**.

Varianza

La **varianza** es la **media aritmética** del **cuadrado de las desviaciones** respecto a la **media**.

Desviación típica

La **desviación típica** es la **raíz cuadrada** de la **varianza**.

Medidas de dispersión

VARIANZA

La **varianza** es la **media aritmética del cuadrado de las desviaciones respecto a la media** de una distribución estadística.

La varianza se representa por σ^2.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}$$

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

Por ejemplo*: La varianza y desviación típica de los datos 2, 9, 8, 15, 21, 5, 20,* serían respectivamente 49,18 y 6,82.

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N} \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{N}$$

Para simplificar el **cálculo de la varianza** vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2 \quad \sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

Ejercicios de varianza

Ejercicio 1: **Calcula la varianza** de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8} = 15$$

Ejercicio 2: **Calcula la varianza** de la distribución de la tabla:

	x_i	f_i
[10, 20)	15	1
[20, 30)	25	8
[30,40)	35	10
[40, 50)	45	9
[50, 60	55	8
[60,70)	65	4
[70, 80)	75	2
		42

Solución:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

DESVIACIÓN TÍPICA

La **desviación típica** es la **raíz cuadrada de la varianza**.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

La **desviación típica** se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2 f_1+(x_2-\bar{x})^2 f_2+\ldots+(x_n-\bar{x})^2 f_n}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2+x_2^2+\ldots+x_n^2}{N}-\bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n}\frac{x_i^2}{N}-\bar{x}^2}$$

$$\sigma = \sqrt{\frac{x_1^2 f+x_2^2 f_2+\ldots+x_n^2 f_n}{N}-\bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n}\frac{x_i^2 f_i}{N}-\bar{x}^2}$$

Ejercicios de desviación típica

Ejercicio 1: Calcula la **desviación típica** de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma = \sqrt{\frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8}} = 3.87$$

Ejercicio 2: Calcula la **desviación típica** de la distribución de la tabla:

	x_i	f_i
[10, 20)	15	1
[20, 30)	25	8
[30,40)	35	10
[40, 50)	45	9
[50, 60	55	8
[60,70)	65	4
[70, 80)	75	2
		42

Solución:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42}-43.33^2} = 14.797$$

Parámetros de Forma

Las variables aleatorias continuas presentan frecuentemente una pauta de variabilidad que se caracteriza por el hecho de que los datos tienden a acumularse en torno a un valor central, que coincide con la media, decreciendo su frecuencia de forma aproximadamente simétrica a medida que se alejan por ambos lados de dicho valor. Los histogramas de estas variables continuas tienen forma de campana de Gauss, que es el modelo matemático de la distribución normal, siendo la distribución que con más frecuencia aparece en multitud fenómenos reales.

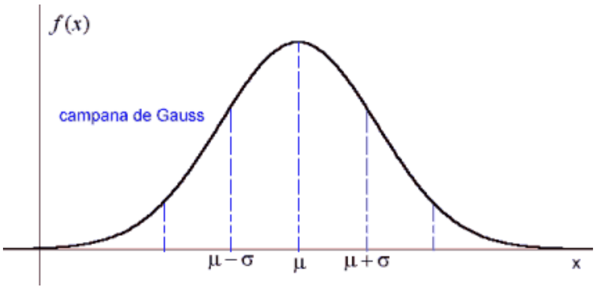


Imagen 1. La función de densidad de una distribución normal

Los parámetros de forma son indicativos de la forma típica que presenta la gráfica o histograma de los datos, es decir de cómo se distribuyen. Entre ellas destacan el coeficiente de asimetría y curtosis.

1. Coeficiente de Asimetría

Las medidas de asimetría permiten conocer si los datos están dispuestos de forma simétrica en torno a un valor central de posición, que generalmente es la media aritmética.

Para saber qué grado de asimetría presentan los datos es necesario el llamado Coeficiente de Asimetría (C.A), que se define como:

$$CA = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)\cdot S^3}$$

Si unos datos son simétricos, lo son respecto a su media y la suma de los cubos de las desviaciones de los datos respecto a su media será nula.

Por el contrario, tendremos una asimetría positiva (C.A > 0), cuando la media esté a la derecha de la mediana y gráficamente se obtiene un histograma en forma de L con una cola hacia la derecha, como se muestra en la figura 2. Así mismo, existe asimetría negativa (C.A <0) la media sea inferior a la mediana y el histograma resultante tiene una forma característica de J, con cola hacia la izquierda.

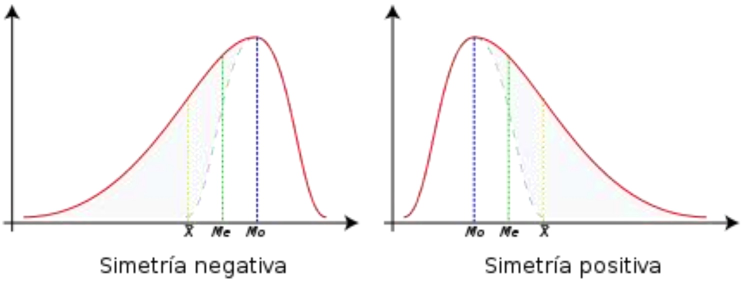


Imagen 2. Tipos de Asimetría

2. **Coeficiente de Curtosis o Apuntamiento (C.C)**

Con este parámetro se pretende medir cómo se reparten las frecuencias relativas de los datos entre el centro y los extremos, tomando como comparación la campana de Gauss. Miden si los valores se concentran más o menos frecuentemente en torno a la media respecto de lo que cabría esperar en una distribución normal

Se define como:

$$CC = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N-1) \cdot S^4}$$

Existen 3 grandes categorías de curtosis:

- ✓ Distribución platicúrtica (apuntamiento negativo) (CC<3): indica que en las colas o extremos hay más casos acumulados que en las colas de una distribución normal, es decir, datos alejados de la media que aparecen con una frecuencia excesiva, respecto de una distribución normal. Presentan un histograma simétrico pero más aplanada que una campana de gauss, como se muestra en la figura 3.
- ✓ Distribución leptocúrtica (apuntamiento positivo) (CC>3): se produce cuando datos alejados de la media aparecen con una frecuencia menor a lo que sería esperable. Presentan un histograma simétrico pero más apuntado que una campana de gauss, como se muestra en la figura 3.
- ✓ Distribución mesocúrtica (apuntamiento normal): coincide con la distribución normal.

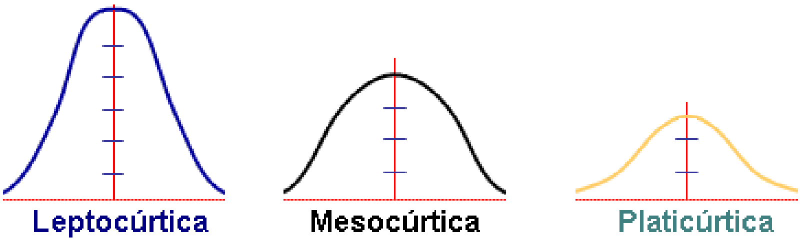


Imagen 3. Tipos de curtosis

***Por ejemplo**: el coeficiente de asimetría y de curstosis de los datos 2, 9, 8, 15, 21, 5, 20,* serían respectivamente 0,22 y -1,64, es decir prácticamente normal respecto del punto de vista de la asimetría y ligeramente planicúrtico.

Cuartiles y percentiles.

Medidas de posición

Las **medidas de posición** dividen un conjunto de datos en grupos con el mismo número de individuos.
Para calcular las **medidas de posición** es necesario que los **datos** estén ordenados de **menor a mayor**.
Las **medidas de posición** son:

Cuartiles

Los **cuartiles dividen** la serie de datos en **cuatro partes iguales**.

Deciles

Los **deciles** dividen la serie de datos en **diez partes iguales**.

Percentiles

Los **percentiles** dividen la serie de datos en **cien partes iguales**.

***Por ejemplo**: Los siguientes datos muestran el número de despedidos que se han producido en 15 empresas del sector del automóvil durante el año 2010:*

33	56	91	64	55	60	2	42	32	26	63	40	25	34	84

Hallar los cuartiles.

Lo primero que debemos hacer es ordenar los datos de menor a mayor:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	25	26	32	33	34	40	42	55	56	60	63	64	84	91

Se trata de un número impar de datos, luego la mediana es el valor central que ocupa la posición (N+1)/2 o en nuestro caso (15+1)/2=8, es decir, el dato 42.

Para el primer y el tercer cuartil, tenemos que N es impar (15) y que (N-1)/2=7 es impar. Por tanto, el primer cuartil C1 es la media de los primeros (N-1)/2 datos, como son 7 datos, será entonces el dato central de los primeros (N-1)/2 datos, o el dato 4. Luego C1=32

Para C3 seguimos el mismo procedimiento, con lo que C3=63.

Interpretación de los parámetros estadísticos.

Las medidas de posición resumen la distribución de datos. Permiten identificar el valor en torno al cual se agrupan mayoritariamente los datos, es decir, cuyo valor es representativo de todos ellos.
Los parámetros de dispersión nos informan sobre la heterogeneidad de los datos y miden en qué medida los datos se agrupan entorno a un valor central.
Los parámetros de forma son indicativos de la forma típica que presenta la gráfica o histograma de los datos, es decir de cómo se distribuyen. Las medidas de asimetría permiten conocer si los datos están dispuestos de forma simétrica en torno a un valor central de posición, que generalmente es la media aritmética.
El coeficiente de apuntamiento o curtosis mide si los valores se concentran más o menos frecuentemente en torno a la media respecto de lo que cabría esperar en una distribución normal.

Distribuciones estadísticas bidimensionales.

Distribuciones bidimensionales.

Relación funcional

Dos variables x e y están relacionadas funcionalmente cuando conocida la primera se puede saber con exactitud el valor de la segunda.

Ejemplos

Si se deja caer una piedra, existe una fórmula que nos permite calcular exactamente, la altura a la que se encuentra en función del tiempo transcurrido.
$h = \frac{1}{2} g t^2$.

Relación estadística

Dos variables x e y están relacionadas estadísticamente cuando conocida la primera se puede estimar aproximadamente el valor de la segunda.
Ejemplo:

Ingresos y gastos de una familia.
Producción y ventas de una fábrica.
Gastos en publicidad y beneficios de una empresa.

Variable estadística bidimensional

Una **variable bidimensional** es una variable en la que cada individuo está definido por un par de caracteres, **(X, Y)**.
Estos dos caracteres son a su vez **variables estadísticas** en las que sí existe relación entre ellas, una de las dos variables es la variable independiente y la otra variable dependiente.

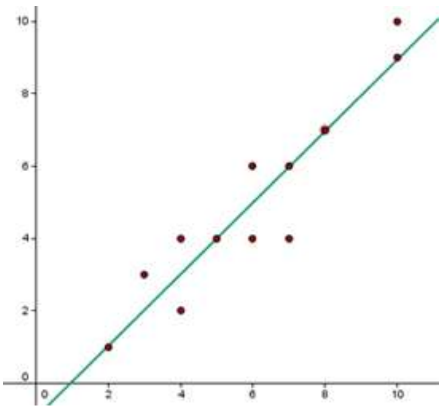
Distribuciones bidimensionales

Son aquellas en las que a cada individuo le corresponden los valores de dos variables, las representamos por el par (x_i, y_i).
Si representamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama **nube de puntos** o **diagrama de dispersión**.
Sobre la nube de puntos puede trazarse una recta que se ajuste a ellos lo mejor posible, llamada **recta de regresión**.

Ejemplo:

Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

Matemáticas	2	3	4	4	5	6	6	7	7	8	10	10
Física	1	3	2	4	4	4	6	4	6	7	9	10



Cálculo de parámetros.

Covarianza

La **covarianza** de una variable bidimensional es la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

La **covarianza** se representa por s_{xy} o σ_{xy}.

$$\sigma_{xy} = \frac{\sum f_i (x_i - \bar{x})(y_i - \bar{y})}{N} \qquad \sigma_{xy} = \frac{\sum f_i x_i y_i}{N} - \bar{x}\bar{y}$$

La **covarianza** indica el sentido de la correlación entre las variables
Si $\sigma_{xy} > 0$ la correlación es directa.
Si $\sigma_{xy} < 0$ la correlación es inversa.

La **covarianza** presenta como inconveniente, el hecho de que su valor depende de la escala elegida para los ejes. Es decir, la **covarianza** variará si expresamos la altura en metros o en centímetros. También variará si el dinero lo expresamos en euros o en dólares.

Ejemplos: Halla la **covarianza** de la distribución de las notas de 12 alumnos de una clase en Matemáticas y Física, sabiendo que son las siguientes:

Matemáticas	2	3	4	4	5	6	6	7	7	8	10	10
Física	1	3	2	4	4	4	6	4	6	7	9	10

Solución:

x_i	2	3	4	4	5	6	6	7	7	8	10	10	72
y_i	1	3	2	4	4	4	6	4	6	7	9	10	60
$x_i \cdot y_i$	2	9	8	16	20	24	36	28	42	56	90	100	431

Después de tabular los datos hallamos las medias aritméticas:

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

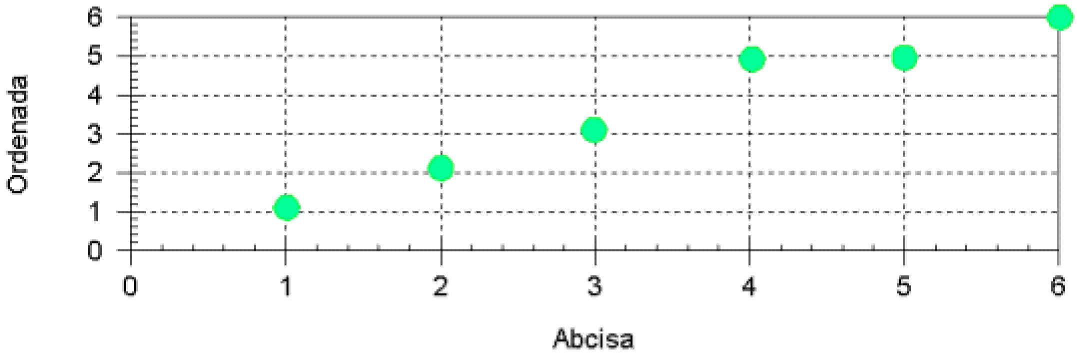
Nube de puntos.

Una **nube de puntos** o diagrama de dispersión consiste en representar cada par de valores de las variables en un sistema de coordenadas cartesianas en el que los ejes X e Y representan las variables de la distribución bidimensional.

Por ejemplo supongamos los siguientes datos:

x	1	2	3	4	5	6
y	1	2	3	5	5	6

La representación gráfica correspondiente sería:



De esta forma es sencillo verificar el tipo de relación que se establece entre las dos variables.

Correlación.

La **correlación** trata de establecer la relación o dependencia que existe entre las dos variables que intervienen en una **distribución bidimensional**.

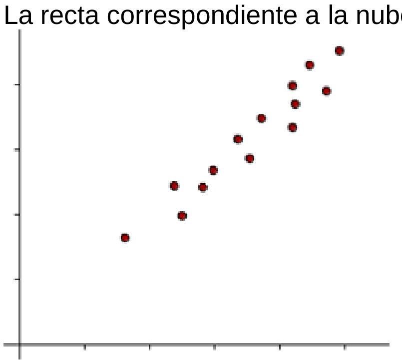
Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra. En caso de que suceda, diremos que las variables están correlacionadas o que hay **correlación** entre ellas.

Tipos de correlación

1º **Correlación directa**

La correlación directa se da cuando al aumentar una de las variables la otra aumenta.

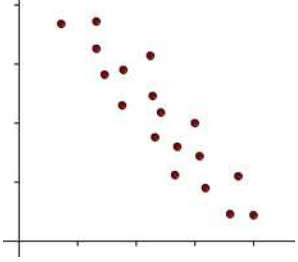
La recta correspondiente a la nube de puntos de la distribución es una recta creciente.



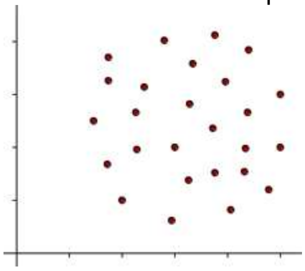
2º **Correlación inversa**

La correlación inversa se da cuando al aumentar una de las variables la otra disminuye.

La recta correspondiente a la nube de puntos de la distribución es una recta decreciente.

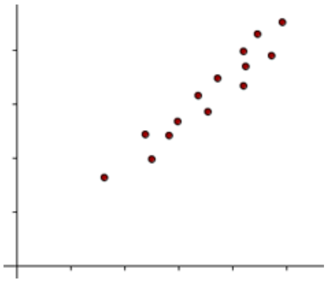


3º **Correlación nula**
La correlación nula se da cuando no hay dependencia de ningún tipo entre las variables.
En este caso se dice que las variables están incorreladas y la nube de puntos tiene una forma redondeada.

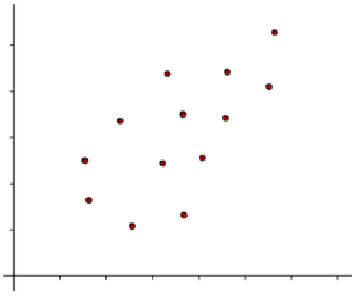


Grado de correlación
El **grado de correlación** indica la proximidad que hay entre los puntos de la nube de puntos. Se pueden dar tres tipos:

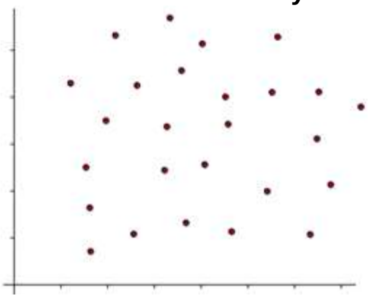
1. **Correlación fuerte**
La correlación será fuerte cuanto más cerca estén los puntos de la recta.



2. **Correlación débil**
La correlación será débil cuanto más separados estén los puntos de la recta.



3. **Correlación muy débil o prácticamente nula**



Coeficiente de correlación lineal

El **coeficiente de correlación lineal** es el cociente entre la **covarianza** y el producto de las **desviaciones típicas** de ambas variables.

El **coeficiente de correlación lineal** se expresa mediante la letra **r**.

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

Propiedades

1. El **coeficiente de correlación** no varía al hacerlo la escala de medición.

Es decir, si expresamos la altura en metros o en centímetros el coeficiente de correlación no varía.

2. El signo del **coeficiente de correlación** es el mismo que el de la **covarianza**.

Si la covarianza es positiva, la correlación es directa.

Si la covarianza es negativa, la correlación es inversa.

Si la covarianza es nula, no existe correlación.

3. El **coeficiente de correlación lineal** es un número real comprendido entre −1 y 1.

$$-1 \leq r \leq 1$$

4. Si el **coeficiente de correlación lineal** toma valores cercanos a −1 la correlación es **fuerte e inversa**, y será tanto más fuerte cuanto más se aproxime r a −1.
5. Si el **coeficiente de correlación lineal** toma valores cercanos a 1 la correlación es **fuerte y directa**, y será tanto más fuerte cuanto más se aproxime r a 1.
6. Si el **coeficiente de correlación lineal** toma valores cercanos a 0, la correlación es **débil**.
7. Si r = 1 ó −1, los puntos de la nube están sobre la recta creciente o decreciente. Entre ambas variables hay **dependencia funcional**.

Ejemplo: Halla el **coeficiente de correlación** de las notas obtenidas por 12 alumnos de una clase en Matemáticas y Física e interpretarlo:

Matemáticas	2	3	4	4	5	6	6	7	7	8	10	10
Física	1	3	2	4	4	4	6	4	6	7	9	10

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
2	1	2	4	1
3	3	9	9	9
4	2	8	16	4
4	4	16	16	16
5	4	20	25	16
6	4	24	36	16
6	6	36	36	36
7	4	28	49	16
7	6	42	49	36
8	7	56	64	49
10	9	90	100	81
10	10	100	100	100
72	60	431	504	380

1º Hallamos las medias aritméticas.

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

2º Calculamos la covarianza.

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

3º Calculamos las desviaciones típicas.

$$\sigma_x = \sqrt{\frac{504}{12} - 6^2} = 2.45 \qquad \sigma_y = \sqrt{\frac{380}{12} - 25} = 2.58$$

4º Aplicamos la fórmula del **coeficiente de correlación lineal**.

$$r = \frac{5.92}{2.45 \cdot 2.58} = 0.94$$

Al ser el **coeficiente de correlación** positivo, la correlación es directa.

Como **coeficiente de correlación** está muy próximo a 1 la correlación es muy fuerte.

Rectas de regresión. Estimación.

Recta de regresión

La **recta de regresión** es la que mejor se ajusta a la **nube de puntos**.

La **recta de regresión** pasa por el punto $(\bar{x}, \bar{y})$ llamado **centro de gravedad**.

Recta de regresión de Y sobre X

La recta de regresión de Y sobre X se utiliza para estimar los valores de la Y a partir de los de la X.
La **pendiente** de la recta es el cociente entre la covarianza y la varianza de la variable X.

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

Recta de regresión de X sobre Y

La **recta de regresión** de X sobre Y se utiliza para estimar los valores de la X a partir de los de la Y.
La **pendiente** de la recta es el cociente entre la covarianza y la varianza de la variable Y.

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

Si la correlación es nula, r = 0, las rectas de regresión son perpendiculares entre sí, y sus eucaciones son:

$y = \bar{y}$ $x = \bar{x}$

Ejemplo: Halla las **rectas de regresión** y represéntalas correspondientes a las notas de 12 alumnos de una clase en Matemáticas y Física:

Matemáticas	2	3	4	4	5	6	6	7	7	8	10	10
Física	1	3	2	4	4	4	6	4	6	7	9	10

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
2	1	2	4	1
3	3	9	9	9
4	2	8	16	4
4	4	16	16	16
5	4	20	25	16
6	4	24	36	16
6	6	36	36	36
7	4	28	49	16
7	6	42	49	36
8	7	56	64	49
10	9	90	100	81
10	10	100	100	100
72	60	431	504	380

1º Hallamos las medias ariméticas.

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

2º Calculamos la covarianza.

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

3º Calculamos las varianzas.

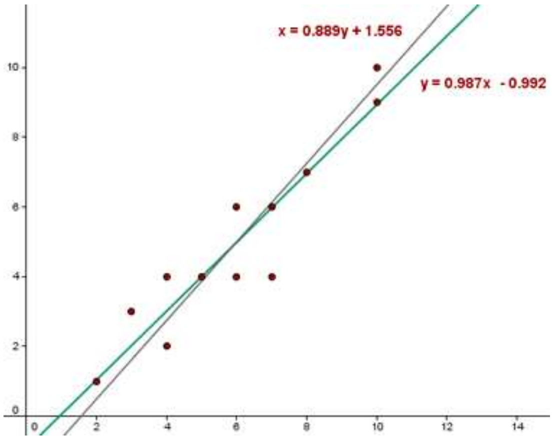
$$\sigma_x^2 = \frac{504}{12} - 6^2 = 6 \qquad \sigma_y^2 = \frac{380}{12} - 25 = 6.66$$

4º Recta de regresión de Y sobre X.

$$y - 5 = \frac{5.92}{6}(x - 6) \qquad y = 0.987x - 0.922$$

5º Recta de regresión de X sobre Y.

$$x - 6 = \frac{5.92}{6.66}(y - 5) \qquad x = 0.889y + 1.556$$



Ejercicio 1: Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.
1 Halla la ecuación de la recta de regresión de la edad sobre el peso.
2 ¿Cuál sería el peso aproximado de un niño de seis años?

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
2	14	4	196	28
3	20	9	400	60
5	32	25	1 024	160
7	42	49	1 764	294
8	44	64	1 936	352
25	152	151	5 320	894

$$\bar{x} = \frac{25}{5} = 5 \qquad \bar{y} = \frac{152}{5} = 30.4$$

$$\sigma_x^2 = \frac{151}{5} - 5^2 = 5.2 \qquad \sigma_y^2 = \frac{5320}{5} - 30.4^2 = 139.84$$

$$\sigma_{xy} = \frac{894}{5} - 5 \cdot 30.\,4 = 26.8$$

$$x - 5 = 0.192\,(y - 30) \qquad x = 0.192y - 0.76$$

$$y - 30.4 = 5.15\,(x - 5) \qquad y = 5.15x + 4.65$$

$$y = 5.15 \cdot 6 + 4.65 = 35.55 \text{ Kg}$$

Ejercicio 2: Un centro comercial sabe en función de la distancia, en kilómetros, a la que se sitúe de un núcleo de población, acuden los clientes, en cientos, que figuran en la tabla:

Nº de clientes (X)	8	7	6	4	2	1
Distancia (Y)	15	19	25	23	34	40

1. Calcula el coeficiente de **correlación lineal**.
2. Si el centro comercial se sitúa a 2 km, ¿cuántos clientes puede esperar?
3. Si desea recibir a 500 clientes, ¿a qué distancia del núcleo de población debe situarse?

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
8	15	120	64	225
7	19	133	49	361
6	25	150	36	625
4	23	92	16	529
2	34	68	4	1 156
1	40	40	1	1 600
28	156	603	170	4 496

$\bar{x} = \frac{28}{6} = 4.67$ $\qquad$ $\bar{y} = \frac{156}{6} = 26$

$\sigma_x^2 = \frac{170}{6} - 4.67^2 = 6.53$ $\qquad$ $\sigma_y^2 = \frac{4496}{6} - 26^2 = 73.33$

$\sigma_x = \sqrt{6.53} = 2.55$ $\qquad$ $\sigma_y = \sqrt{73.33} = 8.56$

$\sigma_{xy} = \frac{603}{6} - 4.677 \cdot 26 = -20.92$

$r = \frac{-20.92}{2.55 \cdot 8.56} = -0.96$

Correlación negativa muy fuerte.

$x - 4.67 = \frac{-20.92}{73.33}(y - 26)$ $\qquad$ $x = -0.29y + 12.09$

$x = -0.29 \cdot 2 + 12.09 = 1151$ clientes

$y - 26 = \frac{-20.92}{6.53}(x - 4.67)$ $\qquad$ $y = -3.2x + 40.96$

$y = -3.2 \cdot 5 + 40.96 = 24.96$ km

Ejercicio 3: Las notas obtenidas por cinco alumnos en Matemáticas y Química son:

Matemáticas	6	4	8	5	3. 5
Química	6. 5	4. 5	7	5	4

Determina las **rectas de regresión** y calcula la nota esperada en Química para un alumno que tiene 7.5 en Matemáticas.

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
6	6. 5	36	42. 25	39
4	4. 5	16	20. 25	18
8	7	64	49	56
5	5	25	25	25
3. 5	4	12. 25	16	14
26. 5	27	153. 25	152. 5	152

$\bar{x} = \frac{26.5}{5} 5.3$ $\qquad$ $\bar{y} = \frac{27}{5} = 5.4$

$\sigma_x^2 = \frac{153.25}{5} - 5.3^2 \; 2.56$ $\qquad$ $\sigma_y^2 = \frac{152.5}{5} - 5.4^2 = 1.3$

$\sigma_{xy} = \frac{152}{5} - 5.3 \cdot 5.4 = 1.78$

$y - 5.4 = 0.7\,(x - 5.3)$ $\qquad$ $y = 0.7x + 1.69$

$x - 5.3 = 1.33\,(y - 5.4)$ $\qquad$ $x = 1.73y - 1.838$

$y = 0.7 \cdot 7.5 + 1.69 = 6.94$

Ejercicio 4: Un conjunto de datos bidimensionales (X, Y) tiene **coeficiente de correlación** r = -0.9, siendo las medias de las distribuciones marginales $\bar{x} = 1$, $\bar{y} = 2$. Se sabe que una de las cuatro ecuaciones siguientes corresponde a la **recta de regresión** de Y sobre X:

y = -x + 2 $\qquad$ 3x - y = 1 $\qquad$ 2x + y = 4 $\qquad$ y = x + 1

Selecciona razonadamente esta **recta**.

Solución: Como el **coeficiente de correlación lineal es negativo**, la **pendiente** de la **recta** también será **negativa**, por tanto descartamos la 2ª y 4ª.

Un punto de la recta ha de ser ($\bar{x}$, $\bar{y}$), es decir, (1, 2).

2 ≠ - 1 + 2 $\qquad$ 2 . 1 + 2 = 4

La recta pedida es: $\qquad$ 2x + y = 4.

Ejercicio 5: Las estaturas y pesos de 10 jugadores de baloncesto de un equipo son:

Estatura (X)	186	189	190	192	193	193	198	201	203	205
Pesos (Y)	85	85	86	90	87	91	93	103	100	101

Calcula:

1 La **recta de regresión** de Y sobre X.

2 El **coeficiente de correlación**.

3 El peso estimado de un jugador que mide 208 cm.

Solución:

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
186	85	34 596	7 225	15 810
189	85	35 721	7 225	16 065
190	86	36 100	7 396	16 340
192	90	36 864	8 100	17 280
193	87	37 249	7 569	16 791
193	91	37 249	8 281	17563
198	93	39 204	8 649	18 414
201	103	40 401	10 609	20 703
203	100	41 209	10 000	20 300
205	101	42 025	10 201	20 705
1 950	921	380 618	85 255	179 971

$$\bar{x} = \frac{1950}{10} = 195 \qquad \bar{y} = \frac{921}{10} = 92.1$$

$$\sigma_x^2 = \frac{380618}{10} - 195^2 = 36.8 \qquad \sigma_y^2 = \frac{85255}{10} - 92.1^2 = 43.09$$

$$\sigma_x = \sqrt{36.8} = 6.07 \qquad \sigma_y = \sqrt{43.09} = 6.56$$

$$y - 92.1 = 1.02\,(x - 195) \qquad y = 1.02x - 106.8$$

$$r = \frac{37.61}{6.07 \cdot 6.56} = 0.94$$

Correlación positiva muy fuerte.

$$y = 1.02 \cdot 208 - 106.8 = 105.36 \text{ kg}$$

Ejercicio 6: A partir de los siguientes datos referentes a horas trabajadas en un taller (X), y a unidades producidas (Y), determinar la **recta de regresión** de Y sobre X, el **coeficiente de correlación lineal** e interpretarlo.

Horas (X)	80	79	83	84	78	60	82	85	79	84	80	62
Producción (Y)	300	302	315	330	300	250	300	340	315	330	310	240

Solución:

$$\bar{x} = \frac{936}{12} = 78 \qquad \bar{y} = \frac{3632}{12} = 302.67$$

$$\sigma_x^2 = \frac{73\,760}{12} - 78^2 = 62.67 \qquad \sigma_y^2 = \frac{1\,109\,254}{12} - 302.67^2 = 828.7$$

$$\sigma_x = \sqrt{62.67} = 7.92 \qquad \sigma_y = \sqrt{828.7} = 28.8$$

$$y - 302.47 = 3.47\,(x - 78) \qquad y = 3.47x + 32.01$$

$$r = \frac{217.41}{7.92 \cdot 28.8} = 0.95$$

Correlación positiva muy fuerte

Ejercicio 7: Se ha solicitado a un grupo de 50 individuos información sobre el número de horas que dedican diariamente a dormir y ver la televisión. La clasificación de las respuestas ha permitido elaborar la siente tabla:

Nº de horas dormidas (X)	6	7	8	9	10
Nº de horas de televisión (Y)	4	3	3	2	1
Frecuencias absolutas (f_i)	3	16	20	10	1

1 Calcula el **coeficiente de correlación**.

2 Determina la ecuación de la **recta de regresión** de Y sobre X.

3 Si una persona duerme ocho horas y media, ¿cuánto cabe esperar que vea la televisión?

Solución:

x_i	y_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$	$y_i \cdot f_i$	$y_i^2 \cdot f_i$	$x_i \cdot y_i \cdot f_i$
6	4	3	18	108	12	48	72
7	3	16	112	784	48	144	336
8	3	20	160	1280	60	180	480
9	2	10	90	810	20	40	180
10	1	1	10	100	1	1	10
		50	390	3082	141	413	1078

$$\bar{x} = \frac{390}{50} = 7.8 \qquad \bar{y} = \frac{141}{50} = 2.82$$

$$\sigma_x^2 = \frac{3082}{50} - 7.8^2 = 0.8 \qquad \sigma_y^2 = \frac{413}{50} - 2.82^2 = 0.3076$$

$$\sigma_x = \sqrt{0.8} = 0.89 \qquad \sigma_y = \sqrt{0.3076} = 0.55$$

$$\sigma_{xy} = \frac{1078}{50} - 7.8 \cdot 2.82 = -0.436$$

$$r = \frac{-0.436}{0.89 \cdot 0.55} = -0.88$$

Es una **correlación negativa y fuerte**.

$$y - 2.82 = \frac{-0436}{0.8}(x - 7.8) \qquad y = -0.545x + 7.071$$

$$y = -0.545 \cdot 8.5 + 7.071 = 2.44 \text{ horas}$$

Ejercicio 8: La tabla siguiente nos da las notas del test de aptitud (X) dadas a seis dependientes a prueba y ventas del primer mes de prueba (Y) en cientos de euros.

X	25	42	33	54	29	36
Y	42	72	50	90	45	48

1 Halla el **coeficiente de correlación** e interpretar el resultado obtenido.
2 Calcula la **recta de regresión** de Y sobre X. Predecir las ventas de un vendedor que obtenga 47 en el test.

Solución:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
25	42	625	1 764	1 050
42	72	1 764	5 184	3 024
33	50	1 089	2 500	1 650
54	90	2 916	8 100	4 860
29	45	841	2 025	1 305
36	48	1 296	2 304	1 728
209	347	8 531	21 877	13 617

$$\bar{x} = \frac{219}{6} = 36.5 \qquad \bar{y} = \frac{347}{6} = 57.83$$

$$\sigma_x^2 = \frac{8531}{6} - 36.5^2 = 89.58 \qquad \sigma_y^2 = \frac{21877}{6} - 57.83^2 = 301.86$$

$$\sigma_x = \sqrt{89.58} = 9.46 \qquad \sigma_y = \sqrt{301.86} = 17.37$$

$$\sigma_{xy} = \frac{13617}{6} - 36.5 \cdot 57.83 = 158.71$$

$$r = \frac{158.71}{9.46 \cdot 17.37} = 0.97$$

$$y - 57.83 = 1.77(x - 36.5) \qquad y = 1.77x - 6.78$$

$$y = 1.77 \cdot 47 - 6.78 = 76.41$$

Introducción a la probabilidad.

Sucesos. Operaciones con sucesos.

Suceso: Es cada uno de los resultados posibles de una experiencia aleatoria.

Ejemplos:

Al lanzar una moneda salga cara.

Al lanzar una moneda se obtenga 4.

Espacio muestral: Es el conjunto de todos los posibles resultados de una experiencia aleatoria, lo representaremos por E (o bien por la letra griega Ω).

Ejemplos: Espacio muestral de una moneda: E = {C, X}.

Espacio muestral de un dado: E = {1, 2, 3, 4, 5, 6}.

Suceso aleatorio es cualquier subconjunto del espacio muestral.

Ejemplo Al tirar un dado un suceso sería que saliera par, otro, obtener múltiplo de 3, y otro, sacar 5.

Un ejemplo completo

Una bolsa contiene bolas blancas y negras. Se extraen sucesivamente tres bolas. Calcular:

1. El espacio muestral.

   E = {(b,b,b); (b,b,n); (b,n,b); (n,b,b); (b,n,n); (n,b,n); (n,n ,b); (n, n,n)}

2. El suceso A = {extraer tres bolas del mismo color}.

   A = {(b,b,b); (n, n,n)}

3. El suceso B = {extraer al menos una bola blanca}.

   B= {(b,b,b); (b,b,n); (b,n,b); (n,b,b); (b,n,n); (n,b,n); (n,n ,b)}

4. El suceso C = {extraer una sola bola negra}.

   C = {(b,b,n); (b,n,b); (n,b,b)}

Sucesos contrarios

El suceso $\overline{A}$ **= E - A** se llama **suceso contrario** o complementario de A.

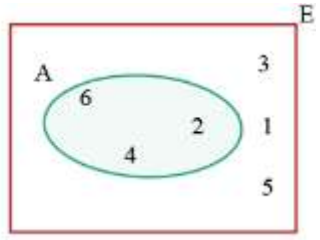
Es decir, se verifica siempre y cuando no se verifique A.

Ejemplo:

Consideramos el experimento que consiste en lanzar un dado, si A = "sacar par". Calcula $\overline{A}$.

A = {2, 4, 6}

$\overline{A}$ = {1, 3, 5}



Unión de sucesos

La **unión de sucesos, A $\cup$ B**, es el suceso formado por todos los elementos de A y de B.

Es decir, el suceso A $\cup$ B se verifica cuando ocurre uno de los dos, A o B, o ambos.

A $\cup$ B se lee como "**A o B**".

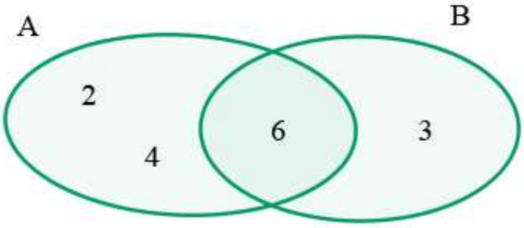
Ejemplo:

Consideramos el experimento que consiste en lanzar un dado, si A = "sacar par" y B = "sacar múltiplo de 3".
Calcular $A \bigcup B$.

$A = \{2, 4, 6\}$

$B = \{3, 6\}$

$A \bigcup B = \{2, 3, 4, 6\}$



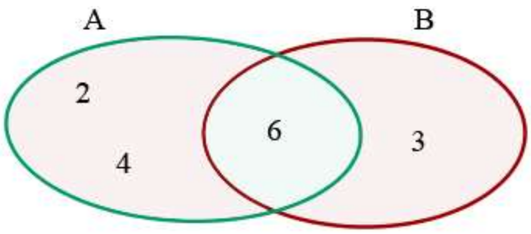
Intersección de sucesos

La **intersección de sucesos, $A \bigcap B$**, es el suceso formado por todos los elementos que son, a la vez, de A y B.

Es decir, el suceso $A \bigcap B$ se verifica cuando ocurren simultáneamente A y B.

$A \bigcap B$ se lee como "**A y B**".

Ejemplo: Consideramos el experimento que consiste en lanzar un dado, si A = "sacar par" y B = "sacar múltiplo de 3". Calcular $A \bigcap B$.



$A = \{2, 4, 6\}$

$B = \{3, 6\}$

$A \bigcap B = \{6\}$

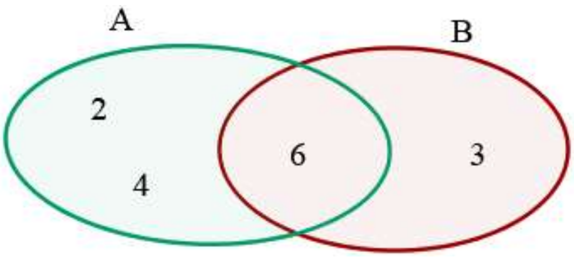
Diferencia de sucesos

La **diferencia de sucesos, A − B**, es el suceso formado por todos los elementos de A que no son de B.

Es decir, la **diferencia de los sucesos** A y B se verifica cuando lo hace A y no B.

A − B se lee como "**A menos B**".

Ejemplo: Consideramos el experimento que consiste en lanzar un dado, si A = "sacar par" y B = "sacar múltiplo de 3". Calcular A − B.



$A = \{2, 4, 6\}$

$B = \{3, 6\}$

$A - B = \{2, 4\}$

Propiedad de la diferencia de sucesos

$$A - B = A \cap \overline{B}$$

Números combinatorios.

Factorial de un número natural

Es el producto de los “n” factores consecutivos desde “n” hasta 1. El **factorial de un número** se denota por **n!**.

$$n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \ldots \cdot 3 \cdot 2 \cdot 1$$

$$0! = 1$$

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

Números combinatorios

El número C_m^n se llama también **número combinatorio**. Se representa por $\binom{m}{n}$ y se lee "**m sobre n**".

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

Ejemplo:

$$\binom{7}{4} = \frac{7!}{4! \cdot 3!} = \frac{7 \cdot \cancel{6} \cdot 5 \cdot \cancel{4!}}{\cancel{4!} \cdot \cancel{3 \cdot 2}} = 35$$

Propiedades de los números combinatorios

1. $\binom{m}{0} = \binom{m}{m} = 1$
2. $\binom{m}{n} = \binom{m}{m-n}$

Los números de este tipo se llaman **complementarios**.

3. $\binom{m}{n-1} + \binom{m}{n} = \binom{m+1}{n}$

$$\binom{11}{4} + \binom{11}{5} = \binom{12}{5}$$

Ejemplo:

Hallar el número de combinaciones de 75 elementos de orden 72.

$$\binom{75}{72} = \binom{75}{75-72} = \binom{75}{3} = \frac{75 \cdot 74 \cdot 73}{3 \cdot 2 \cdot 1} = 67525$$

Probabilidad.

Definición de probabilidad

La probabilidad de un suceso es un número, comprendido entre 0 y 1, que indica las posibilidades que tiene de verificarse cuando se realiza un experimento aleatorio.

Experimentos deterministas: Son los experimentos de los que podemos predecir el resultado antes de que se realicen.

Ejemplo: Si dejamos caer una piedra desde una ventana sabemos, sin lugar a dudas, que la piedra bajará. Si la arrojamos hacia arriba, sabemos que subirá durante un determinado intervalo de tiempo; pero después bajará.

Experimentos aleatorios: Son aquellos en los que no se puede predecir el resultado, ya que éste depende del **azar**.

Ejemplos: Si lanzamos una moneda no sabemos de antemano si saldrá cara o cruz.

Si lanzamos un dado tampoco podemos determinar el resultado que vamos a obtener.

Teoría de probabilidades

La **teoría de probabilidades** se ocupa de **asignar** un cierto **número** a cada **posible resultado** que pueda ocurrir en un **experimento aleatorio**, con el fin de cuantificar dichos resultados y saber si un suceso es más probable que otro.

Axiomas de la probabilidad

1. La probabilidad es positiva y menor o igual que 1.

$0 \leq p(A) \leq 1$

2. La probabilidad del suceso seguro es 1.

$p(E) = 1$

3. Si A y B son incompatibles, es decir $A \cap B = \varnothing$ entonces:

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

Propiedades de la probabilidad

1. La suma de las probabilidades de un suceso y su contrario vale 1, por tanto la probabilidad del suceso contrario es:

$$p(\overline{A}) = 1 - p(A)$$

2. Probabilidad del suceso imposible es cero.

$$p(\varnothing) = 0$$

3. La probabilidad de la unión de dos sucesos es la suma de sus probabilidades restándole la probabilidad de su intersección.

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

4. Si un suceso está incluido en otro, su probabilidad es menor o igual a la de éste.

$$Si \;\; A \subset B, \text{entonces} \qquad p(A) \leq p(B)$$

5. Si $A_1, A_2, \ldots, A_k$ son incompatibles dos a dos entonces:

$$p(A_1 \cup A_2 \cup \cdots \cup A_k) = p(A_1) + p(A_2) + \cdots + p(A_k)$$

6 Si el espacio muestral E es finito y un suceso es $S = \{x_1, x_2, \ldots, x_n\}$ entonces:

$$p(S) = p(x_1) + p(x_2) + \cdots + p(x_n)$$

Ejemplo: La probabilidad de sacar par, al tirar un dado, es: P(par) = P(2) + P(4) + P(6)

Regla de Laplace

Si realizamos un experimento aleatorio en el que hay n sucesos elementales, todos igualmente probables, **equiprobables**, entonces si A es un suceso, la **probabilidad** de que ocurra el suceso A es:

$$P(A) = \frac{\text{número de casos favorables a A}}{\text{número de casos posibles}}$$

Ejemplos:

1 Halla la probabilidad de que al lanzar dos monedas al aire salgan dos caras.

Casos posibles: {cc, cx, xc, xx}.

Casos favorables: 1.

$$P(2\,caras) = \frac{1}{4}$$

2 En una baraja de 40 cartas, halla la P(as) y P(copas).

Casos posibles: 40.

Casos favorables de ases: 4.

$$P(as) = \frac{4}{40} = \frac{1}{10}$$

Casos favorables de copas: 10.

$$P(copas) = \frac{10}{40} = \frac{1}{4}$$

3 Calcula la probabilidad de que al lanzar un dado al aire, salga:

1 Un número par.

Casos posibles: {1, 2, 3, 4, 5, 6}.

Casos favorables: {2, 4, 6}.

$$P(par) = \frac{3}{6} = \frac{1}{2}$$

2 Un múltiplo de tres.

Casos favorables: {3, 6}.

$$P(\dot{3}) = \frac{2}{6} = \frac{1}{3}$$

3 Mayor que 4.

Casos favorables: {5, 6}.

$$P(> 4) = \frac{2}{6} = \frac{1}{3}$$

Probabilidad de la unión de sucesos compatibles

$A \cap B \neq \emptyset$ **p(A ∪ B) = p(A) + p(B) − p(A ∩ B)**

p(A ∪ B ∪ C) = p(A) + p(B) + p(C) − p(A ∩ B) − p(A ∩ C) − p(B ∩ C) + p(A ∩ B ∩ C)

Ejemplo: Calcula la probabilidad de obtener un múltiplo de 2 ó un 6 al lanzar un dado.

$$P(\dot{2} \cup 6) = \frac{3}{6} + \frac{1}{6} - \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Probabilidad de la diferencia de sucesos

$$p(A - B) = p(A \cap \bar{B}) = p(A) - p(A \cap B)$$

Probabilidad condicionada.

Sean A y B dos sucesos de un mismo espacio muestral E.

Se llama **probabilidad** del suceso B **condicionado** a A y se representa por **P(B/A)** a la **probabilidad del suceso B una vez ha ocurrido el A**.

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

Ejemplo: Calcula la probabilidad de obtener un 6 al tirar un dado sabiendo que ha salido par.

$$P(6/par) = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

Sucesos independientes

Dos sucesos A y B son independientes si **p(A/B) = p(A)**

Sucesos dependientes

Dos sucesos A y B son dependientes si **p(A/B) ≠ p(A)**

Probabilidad de la intersección de sucesos independientes

$p(A \cap B) = p(A) \cdot p(B)$

Ejemplo: Se tiene una baraja de 40 cartas, una vez que se ha sacado una se vuelve a meter (extracción con reemplazamiento). ¿Cuál es la probabilidad de extraer dos ases?

$$p(A \cap B) = p(A_1) \cdot p(A_2) = \frac{4}{40} \cdot \frac{4}{40} = \frac{1}{100}$$

Probabilidad de la intersección (en general)

$p(A \cap B) = p(A) \cdot p(B/A)$

Ejemplo: Se tiene una baraja de 40 cartas, se extraen dos cartas. ¿Cuál es la probabilidad de extraer dos ases?

$$p(A_1 \cap A_2) = p(A_1) \cdot p(A_2 / A_1) = \frac{4}{40} \cdot \frac{3}{39} = \frac{1}{130}$$

Tabla de contingencia de doble entrada

	A	$\bar{A}$	
B	$A \cap B$	$\bar{A} \cap B$	
$\bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	

Fórmulas de probabilidad

Regla de Laplace

$$P(A) = \frac{\text{número de casos favorables a A}}{\text{número de casos posibles}}$$

Probabilidad del suceso contrario

$$p(\bar{A}) = 1 - p(A)$$

$$p(A) = p(A \cap B) + p(A \cap \bar{B})$$

Probabilidad de la unión

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

Probabilidad de la intersección **$p(A \cap B) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B)$**

Probabilidad condicionada

$$P(B/A) = \frac{p(A \cap B)}{p(A)} \qquad P(A/B) = \frac{p(A \cap B)}{p(\bar{B})}$$

Probabilidad de la diferencia de sucesos $p(A-B)=p\left(A\cap\overline{B}\right)=p\left(A\right)-p\left(A\cap B\right)$

Probabilidad total $\boxed{P(B)=P(A_1)\cdot P(B/A_1)+P(A_2)\cdot P(B/A_2)+\cdots+P(A_n)\cdot P(B/A_n)=\sum_{i=1}^{n}P(A_i)\cdot P(B/A_i)}$,

Teorema de Bayes

$$\boxed{P\left(A_j/B\right)=\frac{P\left(B\cap A_j\right)}{P(B)}=\frac{P\left(A_j\right)\cdot P\left(B/A_j\right)}{\sum_{i=1}^{n}P\left(A_i\right)\cdot P\left(B/A_i\right)}}$$

Leyes de De Morgan

$$\boxed{(A\cup B)^c=\left(\overline{A\cup B}\right)=A^c\cap B^c=\overline{A}\cap\overline{B}}$$

$$\boxed{(A\cap B)^c=\left(\overline{A\cap B}\right)=A^c\cup B^c=\overline{A}\cup\overline{B}}$$

Sucesos independientes *Dos sucesos A y B son independientes si* ***p(A/B) = p(A)*** *o también si* ***p(B/A)=p(B)*** *o también si* ***p(A∩B)=p(A)·p(B)***

Sucesos incompatibles A y B son incompatibles si **no tienen elementos en común**, es decir **p(A ∩ B) = 0**

Distribuciones de probabilidad. Variable discreta.

Función de probabilidad.

Función de probabilidad de una distribución discreta de probabilidad

Se llama función de probabilidad de una variable aleatoria discreta X a la aplicación que asocia a cada valor de x_i de la variable su probabilidad p_i, que es un número comprendido entre 0 y 1. Además la suma de las probabilidades de todos los sucesos elementales es siempre 1

$$0 \le p_i \le 1$$

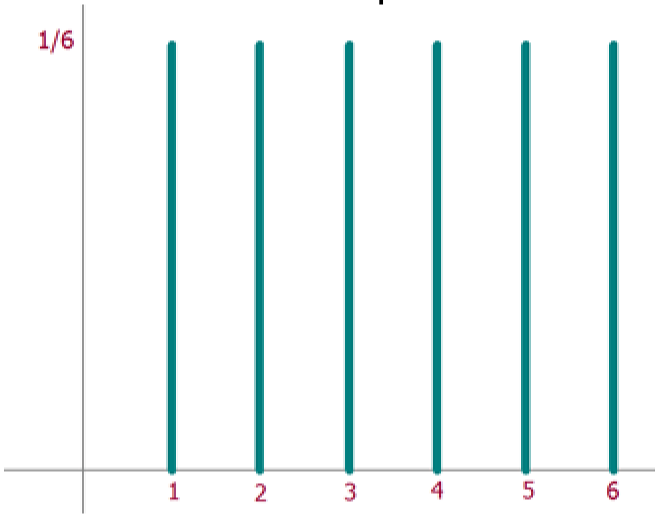
$$p_1 + p_2 + p_3 + \cdots + p_n = \Sigma\, p_i = 1$$

Ejemplo: Calcula la **distribución de probabilidad** de las puntuaciones obtenidas al lanzar un dado.

x	1	2	3	4	5	6	
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	**$\Sigma p_i = 1$**

Representación

La representación de una **distribución discreta de probabilidad es un diagrama de barras.**



Ejercicio 1: Se lanza un par de dados. Se define la variable aleatoria X como la suma de las puntuaciones obtenidas. Halla la función de probabilidad, la esperanza matemática y la varianza

Solución:

x	p_i	$x \cdot p_i$	$x^2 \cdot p_i$
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4 /36	20/3 6	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4 /36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
		7	**54.83**

$$\mu = 7$$

$$\sigma = \sqrt{54.83 - 7^2} = 2.415$$

Función de distribución.

Sea X una variable aleatoria discreta cuyos valores suponemos ordenados de menor a mayor. Llamaremos **función de distribución de la variable X**, y escribiremos F(x) a la función **F(x) = p(X ≤ x)**

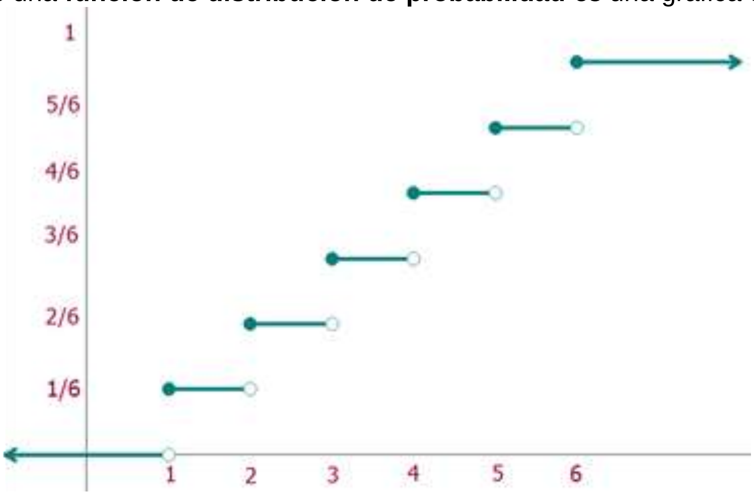
La función de distribución asocia a cada valor de la variable aleatoria la probabilidad acumulada hasta ese valor.

Ejemplo: Calcula la **función de distribución de probabilidad** de las puntuaciones obtenidas al lanzar un dado.

x	p_i
$x < 1$	0
$1 \leq x < 2$	$\frac{1}{6}$
$2 \leq x < 3$	$\frac{2}{6}$
$3 \leq x < 4$	$\frac{3}{6}$
$4 \leq x < 5$	$\frac{4}{6}$
$5 \leq x < 6$	$\frac{5}{6}$
$6 \leq x$	1

Representación

La representación de una **función de distribución de probabilidad** es una gráfica escalonada.



Ejercicio 2: Sea X una variable aleatoria discreta cuya función de probabilidad es:

x	$\mathbf{p_i}$
0	0,1
1	0,2
2	0,1
3	0,4
4	0,1
5	0,1

1 Calcula y representa gráficamente la función de distribución

2 Calcula las siguientes probabilidades:

a) p (X < 4.5)

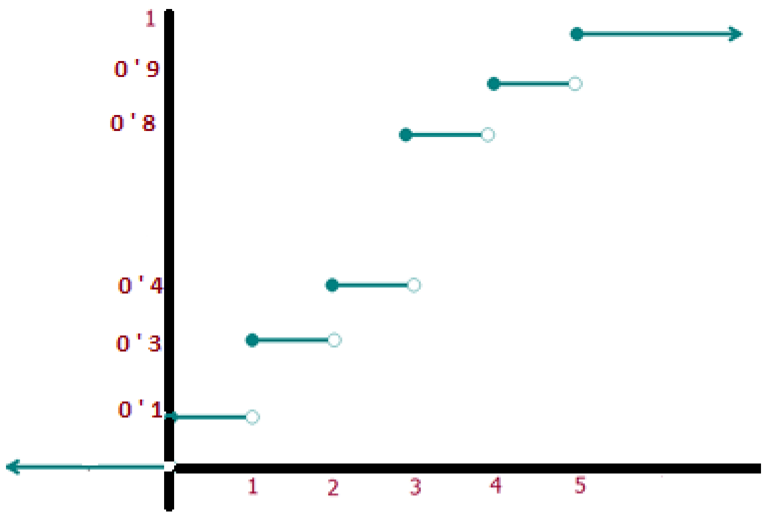
b) p (X ≥ 3)

c) p (3 ≤ X < 4.5)

Soluciones:

1 Calcula y representa gráficamente la función de distribución

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.1 & 0 \le x < 1 \\ 0.3 & 1 \le x < 2 \\ 0.4 & 2 \le x < 3 \\ 0.8 & 3 \le x < 4 \\ 0.9 & 4 \le x < 5 \\ 1 & 5 \le x \end{cases}$$



2 Calcula las siguientes probabilidades:

a) p (X < 4.5) p (X < 4.5) = F (4.5) = 0.9

b) p (X ≥ 3) p (X ≥ 3) = 1 - p(X < 3) = 1 - 0.4 = 0.6

c) p (3 ≤ X < 4.5) p (3 ≤ X < 4.5) = p (X < 4.5) - p(X < 3) = 0.9 - 0.4 = 0.5

Distribución binomial.

Una **distribución binomial o de Bernoulli** tiene las siguientes características:

1. En cada prueba del experimento sólo son posibles **dos resultados**: **éxito** y **fracaso**.
2. La **probabilidad de éxito es constante**, es decir, que no varía de una prueba a otra. Se representa por **p**.
3. La **probabilidad de fracaso** también es **constante**, Se representa por **q**,

$$q = 1 - p$$

4. El **resultado** obtenido en cada prueba es **independiente** de los resultados obtenidos anteriormente.
5. La **variable aleatoria binomial**, **X**, expresa el **número de éxitos obtenidos** en las **n** pruebas. Por tanto, los valores que puede tomar **X** son: **0, 1, 2, 3, 4, ..., n**.

La **distribución bimomial** se expresa por **B(n, p)**

Cálculo de probabilidades en una distribución binomial.

Función de probabilidad de la distribución binomial

La **función de probabilidad de la distribución binomial**, también denominada **función de la distribución de Bernoulli**, es:

$$p(X = k) = \binom{n}{k} p^k \cdot q^{n-k}$$

n es el número de pruebas. **k** es el número de éxitos.
p es la probabilidad de éxito. **q** es la probabilidad de fracaso.

El número combinatorio $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Ejemplo: La última novela de un autor ha tenido un gran éxito, hasta el punto de que el 80% de los lectores ya la han leído. Un grupo de 4 amigos son aficionados a la lectura. ¿Cuál es la probabilidad de que el grupo hayan leído la novela 2 personas?

Solución: n = 4 p = 0.8 q = 0.2 B(4, 0.2)

$$p(X = 2) = \binom{4}{2} 0.8^2 \cdot 0.2^2 = \frac{4 \cdot 3}{2} \cdot 0.64 \cdot 0.04 = 0.1536$$

Parámetros de la distribución binomial

Media $\mu = n \cdot p$

Varianza $\sigma^2 = n \cdot p \cdot q$

Desviación típica $\sigma = \sqrt{n \cdot p \cdot q}$

Ejemplo:

La probabilidad de que un artículo producido por una fábrica sea defectuoso es 0.02. Se envió un cargamento de 10.000 artículos a unos almacenes. Halla el número esperado de artículos defectuosos, la varianza y la desviación típica.

Solución:

$$\mu = 10.000 \cdot 0.02 = 200$$

$$\sigma^2 = 10.000 \cdot 0.02 \cdot 0.98 = 196$$

$$\sigma = \sqrt{196} = 14$$

Ejercicio 1: Se lanza una moneda cuatro veces. Calcular la probabilidad de que salgan más caras que cruces

Solución: Es una distribución binomial B(4, 0.5) p = 0.5 q = 0.5

$$p(X \geq 3) = p(X = 3) + p(X = 4) =$$

$$= \binom{4}{3} 0.5^3 \cdot 0.5 + \binom{4}{4} 0.5^4 = 0.3125$$

Ejercicio 2: Un agente de seguros vende pólizas a cinco personas de la misma edad y que disfrutan de buena salud. Según las tablas actuales, la probabilidad de que una persona en estas condiciones viva 30 años o más es 2/3. Hállese la probabilidad de que, transcurridos 30 años, vivan:

1. Las cinco personas
2. Al menos tres personas
3. Exactamente dos personas

Soluciones: 1 Las cinco personas

Es una distribución binomial B(5, 2/3) p = 2/3 q = 1/3

$$p(X=5)=\binom{5}{5}\left(\frac{2}{3}\right)^5=0.132$$

2 Al menos tres personas

$$p(X\geq 3)=p(X=3)+p(X=4)+p(X=5)=$$

$$=\binom{5}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2+\binom{5}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)+\binom{5}{5}\left(\frac{2}{3}\right)^5=0.791$$

3 Exactamente dos personas

$$p(X=2)=\binom{5}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3=0.164$$

Ejercicio 3: Si de seis a siete de la tarde se admite que un número de teléfono de cada cinco está comunicando, ¿cuál es la probabilidad de que, cuando se marquen 10 números de teléfono elegidos al azar, sólo comuniquen dos?

Solución: Es una distribución binomial B(10, 1/5) p = 1/5 q = 4/5

$$p(X=2)=\binom{10}{2}\left(\frac{1}{5}\right)^2\cdot\left(\frac{4}{5}\right)^8=0.3020$$

Ejercicio 4: La probabilidad de que un hombre acierte en el blanco es 1/4. Si dispara 10 veces ¿cuál es la probabilidad de que acierte exactamente en tres ocasiones? ¿Cuál es la probabilidad de que acierte por lo menos en una ocasión?

Soluciones: Es una distribución binomial B(10, 1/4) p = 1/4 q = 3/4

$$p(X=3)=\binom{10}{3}\left(\frac{1}{4}\right)^3\cdot\left(\frac{3}{4}\right)^7=0.25$$

$$p(\text{al menos uno})=1-\binom{10}{0}\left(\frac{1}{4}\right)^0\cdot\left(\frac{3}{4}\right)^{10}=0.9437$$

Ejercicio 5: En una urna hay 30 bolas, 10 rojas y el resto blancas. Se elige una bola al azar y se anota si es roja; el proceso se repite, devolviendo la bola, 10 veces. Calcula la media y la desviación típica

Soluciones: Es una distribución binomial B(10, 1/3) p = 1/3 q = 2/3

$$\mu=10\cdot\frac{1}{3}=3.33$$

$$\sigma=\sqrt{10\cdot\frac{1}{3}\cdot\frac{2}{3}}=1.49$$

DISTRIBUCIÓN DE POISSON

La distribución de Poisson se utiliza para modelar datos discretos como aproximación a la Binomial dada la dificultad que existía de encontrar tablas Binomiales adecuadas cuando n es grande y p pequeña. La distribución de probabilidad de Poisson proporciona buenas aproximaciones cuando np <= 5.

Se aproxima a la binomial cuando p es igual o menor a 0.1, y el tamaño de muestra es grande (n > 16) por tanto np > 1.6.

Una Variable aleatoria X tiene distribución Poisson si toma probabilidades con.

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0,1,\ldots$$

Con media y varianza:

$$\mu = np$$

$$\sigma = \sqrt{\mu} = \sqrt{np}$$

Ejemplo 1. Suponga que una compañía de seguros asegura las vidas de 5000 hombres de 42 años de edad. Si los estudios actuariales muestran que la probabilidad de que un hombre muera en cierto año es 0.001, entonces la probabilidad de que la empresa pague exactamente 4 indeminizaciones y= 4 en un cierto año es:

$$P(y=4) = p(4) = \frac{5000!}{4!*4996!}(0.001)^{4}(0.999)^{4996}$$

Aproximando con la distribución de Poisson, se toma la tasa media de sucesos λ= np = (5000)*(0.001)= 5, teniendo:

$$P(y=4) = \frac{\lambda^{4}e^{-\mu}}{4!} = \frac{5^{4}e^{-5}}{4!} = 0.1745$$

Ejemplo 2. Una planta tiene 20 máquinas, si la probabilidad de que falla una en cierto día es 0.05. Encuentre la probabilidad de que durante un día determinado fallen dos máquinas.

np = 20 *0.05 = 1.0

$$P(y=2) = \frac{1^{2}e^{-1}}{2!} = 0.184$$

Si se calcula con la distribución Binomial se tiene:

$$P(y=2) = p(2) = \frac{20!}{2!*18!}(0.05)^{2}(0.95)^{18} = 0.188$$

La aproximación es mejor conforme se aproxima a np = 5.

La distribución de Poisson además de ser útil como aproximación de las probabilidades Binomiales, constituye un buen modelo para experimentos donde Y representa el número de veces que ha ocurrido un evento en una unidad dada de tiempo o de espacio. Por ejemplo:

Número de llamadas recibidas en un conmutador durante un día, conociendo el promedio por día.

Número de reclamaciones contra una empresa de seguros por semana, conociendo el prom. Sem.

Número de llegadas a una estación de servicio durante un minuto dado, conociendo el prom./min.

Número de ventas hechas por un agente de ventas en un día, conociendo el promedio por día.

Sólo se requiere que los eventos sean independientes.

Distribuciones de probabilidad. Variable continua.

Distribuciones de probabilidad de una variable continua.

Una distribución de probabilidad de una variable continua nos mide el área bajo la curva de la función de densidad de la variable continua.

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

Mientras que en una variable discreta la p(X=a) suele tener un valor no nulo, en una variable continua es cero. Por ejemplo, si se mide la anchura de una hoja de roble, el resultado 3,5 cm es posible, pero tiene probabilidad cero porque hay infinitos valores posibles entre 3 cm y 4 cm. Cada uno de esos valores individuales tiene probabilidad cero, aunque la probabilidad de ese intervalo no lo es. Esta aparente paradoja se resuelve por el hecho de que la probabilidad de que *X* tome algún valor en un conjunto infinito como un intervalo, no puede calcularse mediante la adición simple de probabilidades de valores individuales. Formalmente, cada valor tiene una probabilidad infinitesimal que estadísticamente equivale a cero.

Distribución normal. Manejo de la tabla de la función de distribución N(0,1).

Distribución normal.

Esta es, con seguridad, la más importante de las distribuciones de probabilidad.
El primero que la describe es Moivre (en 1733), pero no fue hasta cincuenta años más tarde cuando el matemático alemán Gauss, la redescubrió. De ahí que también se le llama "campana de Gauss".
Ejemplo de poblaciones con distribución de frecuencias prácticamente de tipo Normal:
1.-caracteres morfológicos (tallas, pesos…)
2.-caracteres fisiológicos (efectos de una misma dosis de un fármaco)
3.-caracteres sociológicos (consumo de ciertos productos por un mismo grupo humano)
En general: Cualquier característica que se obtenga como suma de muchos factores

La curva Normal usada con más frecuencia (por estar tabulada) es la llamada Normal Estándar o N (0,1), se le llama así precisamente por tener como parámetros: media cero ($\mu = 0$), y desviación típica uno ($\sigma = 1$), se llama distribución estándar o normal tipificada y suele designarse por la letra Z.

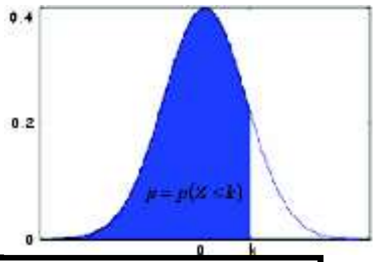
En esta distribución los valores de las probabilidades para los distintos valores de Z, están tabulados, según muestra la tabla anterior, por lo que para conocerlos debemos aprender a manejar las tablas.
La tabla nos da las probabilidades $p(z \leq k)$ para valores de k de 0 hasta 4, de centésima en centésima.

Las características más importantes de la curva normal son las siguientes:

1. **El dominio de la variable normal es todo $\Re$**
2. ***f(x)* es simétrica respecto a la media de la distribución μ**
3. **El máximo de *f(x)* se alcanza en *x* = μ**

4. **Tiene dos puntos de inflexión con abscisas: μ-σ y μ+σ .**
5. **El eje OX es una asíntota de *f(x)***

Tabla de áreas bajo la curva Normal estándar: N (0,1)



El valor de k se busca así: unidades y décimas: columna de la izquierda, las centésimas: en la fila superior.

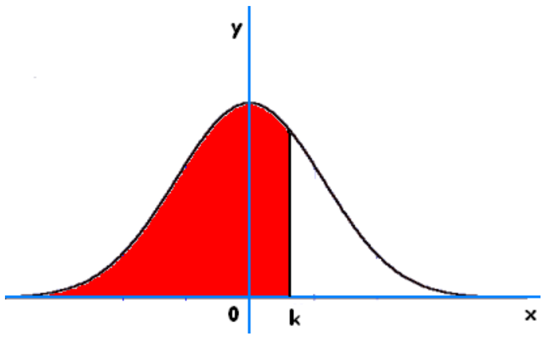
$p(Z \leq k) = p = \textit{Área coloreada}$

N(0,1) K	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Manejo de tablas

Vamos a ver como se calculan probabilidades utilizando la tabla de la Normal Estándar (tipificada) $Z \in N(0{,}1)$ **con algunos casos particulares.**

La tabla anterior da los valores de la probabilidad acumulada hasta el valor "k", es decir: $p(Z \leq k)$



Por la simetría de la función de probabilidad de una Normal estándar tenemos:

a) $p(Z \le k) + p(Z \ge k) = 1$	**b)** $p(Z \ge -k) = p(Z \le k) \quad k > 0$
c) $p(Z \le -k) = p(Z \ge k) \quad k > 0$	**d)** $p(t \le Z \le k) = p(Z \le k) - p(Z \le t)$

Los ejemplos que tienes a continuación están resueltos manejando esa tabla.

1. **Probabilidad de que Z tome valores menores o iguales que 1,45**

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

2. **Probabilidad de que Z tome valores menores o iguales que -1,45**

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

3. **Probabilidad de que Z tome valores entre 1´25 y 2'57**

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

4. **Probabilidad de que Z tome valores entre –2'57 y –1'25**

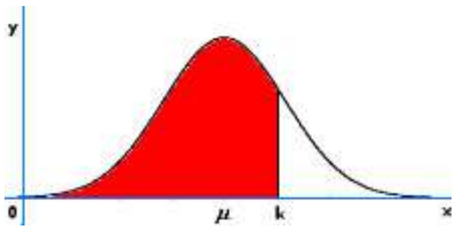
$$p(-2'57 \le Z \le -1'25) = (1'25 \le Z \le 2'57) = 0'1005$$

5. **Probabilidad de que Z tome valores entre –0'53 y 2'46**

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

Cálculo de probabilidades en distribuciones normales. Tipificación

Para calcular la probabilidad de una variable normal X: N (μ,σ) no tipificada, es decir, que no toma los valores $\mu = 0 \quad y \quad \sigma = 1$**, se transforma en una variable normal tipificada mediante el cambio:** $Z = \dfrac{X - \mu}{\sigma}$ **que sigue una distribución de media 0 y desviación típica 1 (tipificada):** $Z \in N(0,1)$



En una Normal de media 6 y desviación típica 4, X: N (6,4) calcula:

a) **Probabilidad de que X tome valores menores o iguales que 3**

$$p(x \le 3) = p\left(z \le \frac{3-6}{4}\right) = p(z \le -0,75) = 1 - p(z \le 0,75) = 1 - 0,7734 = 0,2266$$

b) **Probabilidad de que X tome valores entre 5 y 8.**

$$p(5 \le x \le 8) = p\left(\frac{5-6}{4} \le z \le \frac{8-6}{4}\right) = p(-0,25 \le z \le 0,5) = 0,6915 - 1 + 0,5987 = 0,2902$$

Cálculo de probabilidades en la distribución normal

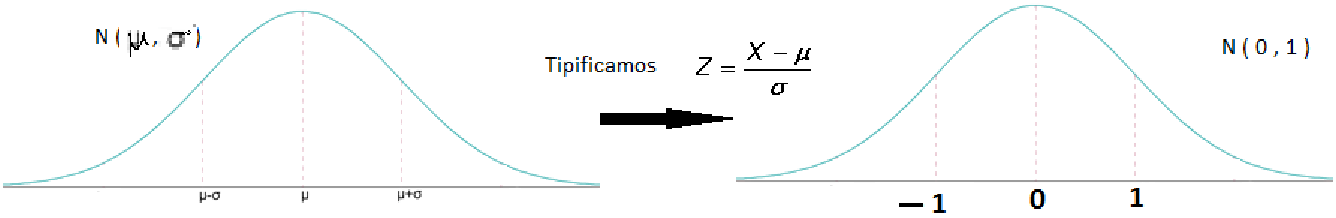
La **tabla** nos da las **probabilidades de P(z ≤ k)**, siendo **z** la variable tipificada.

La probabilidad de la variable X dependerá del área del recinto sombreado en la figura. Y para calcularla utilizaremos una tabla.

Tipificación de la variable

Para poder utilizar la tabla tenemos que transformar la variable **X** que sigue una distribución **N(μ, σ)** en otra variable **Z** que siga una distribución **N(0, 1)**.

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

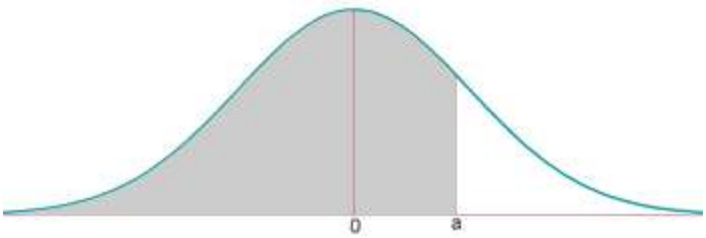


Búsqueda en la tabla de valor de k

Unidades y décimas en la columna de la izquierda.

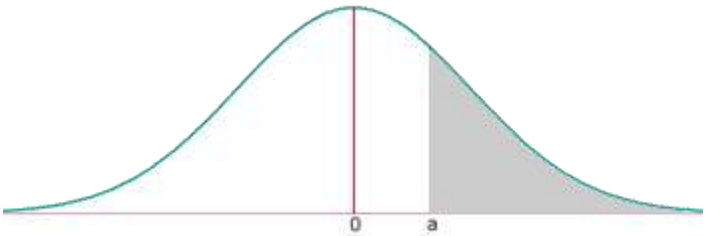
Céntesimas en la fila de arriba.

P(Z ≤ a)



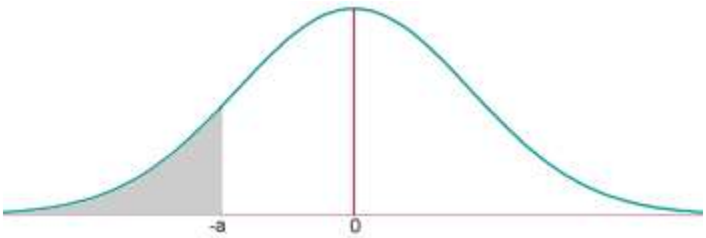
P(Z ≤ 1.47) = 0.9292

P(Z > a) = 1 - P(Z ≤ a)



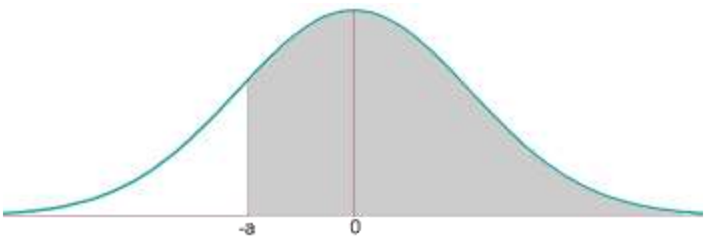
P(Z > 1.47) = 1 − P(Z ≤ 1.47) = 1 − 0.9292 = 0.0708

P(Z ≤ −a) = 1 − P(Z ≤ a)



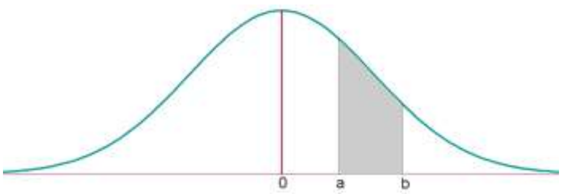
P(Z ≤ −1.47) = 1 − P(Z ≤ 1.47) = 1 − 0.9292 = 0.0708

P(Z > −a) = P(Z ≤ a)



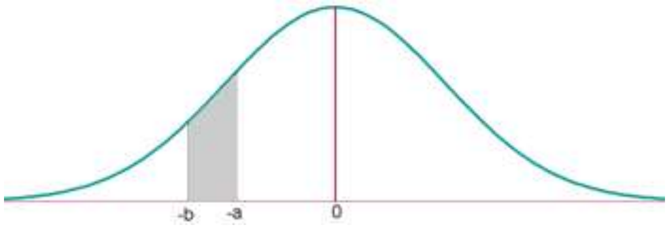
p(Z > −1.47) = p(Z ≤ 1.47) = 0.9292

P(a < Z ≤ b) =
=P(Z ≤ b) − P(Z ≤ a)



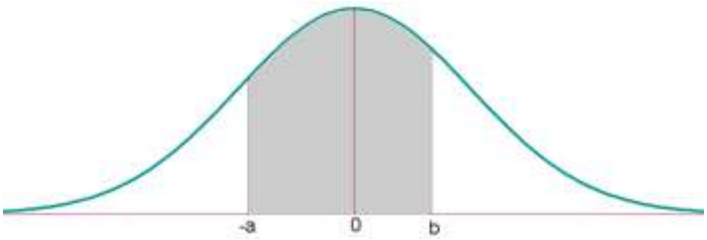
P(0.45 <Z ≤ 1.47) =
= P(Z ≤ 1.47) − P(Z ≤ 0.45) =
= 0.9292 − 0.6736 = 0.2556

P(−b < Z ≤ −a) = =
P(a < Z ≤ b)



P(−1.47 <Z ≤ − 0.45) =
=P(0.45 <Z ≤ 1.47) =
= P(Z ≤ 1.47) − P(Z ≤ 0.45) = =
0.9292 − 0.6736 = 0.2556

P(−a < Z ≤ b) =
P(Z ≤ b) –
[1 − P(Z ≤ a)]



P(-1.47 < Z ≤ 0.45) =
P(Z ≤ 0.45) − [1 − P(Z ≤ 1.47)]=
= 0.6736 − (1 − 0.9292) = 0.6028

Ejercicio 1: Si X es una variable aleatoria de una distribución N(μ, σ), hallar: p(μ−3σ ≤ X ≤ μ+3σ)
Solución:

$$p(\mu-3\sigma\le X\le\mu+3\sigma)=p\left(\frac{(\mu-3\sigma)-\mu}{\sigma}\le z\le\frac{(\mu+3\sigma)-\mu}{\sigma}\right)=$$
$$p(-3\le z\le 3)=p(z\le 3)-p(z\le -3)=$$
$$=p(z\le 3)-(1-p(z\le 3))=$$
$$=0.9987-1+0.9987=0.9974$$

Es decir, que aproximadamente el 99.74% de los valores de X están a menos de tres desviaciones típicas de la media.

Ejercicio 2: En una distribución normal de media 4 y desviación típica 2, calcular el valor de a para que:
P(4−a ≤ x ≤ 4+a) = 0.5934
Solución:

$$p\left(\frac{(4-a)-4}{2}\le z\le\frac{(4+a)-4}{2}\right)=0.5934$$
$$p\left(\frac{-a}{2}\le z\le\frac{a}{2}\right)=p\left(z\le\frac{a}{2}\right)-p\left(z\le-\frac{a}{2}\right)=$$
$$=p\left(z\le\frac{a}{2}\right)-p\left(z\ge\frac{a}{2}\right)=p\left(z\le\frac{a}{2}\right)-p\left(1-p\left(z\le\frac{a}{2}\right)\right)$$
$$2\cdot p\left(z\le\frac{a}{2}\right)-1=0.5934 \qquad p\left(z\le\frac{a}{2}\right)=0.7969$$
$$\frac{a}{2}=0.83 \qquad a=1.66$$

Ejercicio 3: En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°
Solución:

$$p[21<X\le 27]=p\left(\frac{21-23}{5}<Z\le\frac{27-23}{5}\right)=$$
$$=p(-0.4<Z\le 0.8)=p(Z\le 0.8)-[1-p(Z\le 0.4)]=$$
$$=0.7881-(1-0.6554)=0.4425\cdot 30=13$$

Ejercicio 4: La media de los pesos de 500 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, halla cuántos estudiantes pesan:

1 Entre 60 kg y 75 kg
2 Más de 90 kg
3 Menos de 64 kg
4 64 kg
5 64 kg o menos

Soluciones: 1 Entre 60 kg y 75 kg

$$p[60 < X \le 75] = p\left(\frac{60-70}{3} < Z \le \frac{75-70}{3}\right) =$$
$$= p(-3.33 < Z \le 1.67) = p(Z \le 1.67) - [1 - p(Z \le 3.33)] =$$
$$= 0.9525 - (1 - 0.9996) = 0.9521 \cdot 500 = 476$$

2 Más de 90 kg

$$p(X > 90) = p\left(Z > \frac{90-70}{3}\right) = p(Z > 6.67) =$$
$$= 1 - p(Z < 6.67) = 1 - 1 = 0 \cdot 500 = 0$$

3 Menos de 64 kg

$$p(X < 64) = p\left(Z < \frac{64-70}{3}\right) = p(Z < -2) = 1 - p(Z \le 2) =$$
$$= 1 - 0.7772 = 0.02128 \cdot 500 = 11$$

4 64 kg

$$p(X = 64) = p\left(Z = \frac{64-70}{3}\right) = p(Z = -2) = 0 \cdot 500 = 0$$

5 64 kg o menos

$$p(X \le 64) = p(X < 64) = 11$$

Ejercicio 5: Se supone que los resultados de un examen siguen una distribución normal con media 78 y desviación típica 36. Se pide:

1 ¿Cuál es la probabilidad de que una persona que se presenta el examen obtenga una calificación superior a 72?
2 Calcula la proporción de estudiantes que tienen puntuaciones que exceden por lo menos en cinco puntos de la puntuación que marca la frontera entre el Apto y el No-Apto (son declarados No-Aptos el 25% de los estudiantes que obtuvieron las puntuaciones más bajas)
3 Si se sabe que la calificación de un estudiante es mayor que 72 ¿cuál es la probabilidad de que su calificación sea, de hecho, superior a 84?

Soluciones:

1 ¿Cuál es la probabilidad de que una persona que se presenta el examen obtenga una calificación superior a 72?

$$p(X > 72) = p\left(Z > \frac{72-78}{36}\right) =$$
$$= p(Z > -0.16) = p(Z < 0.16) = 0.5636$$

2 Calcula la proporción de estudiantes que tienen puntuaciones que exceden por lo menos en cinco puntos de la puntuación que marca la frontera entre el Apto y el No-Apto (son declarados No-Aptos el 25% de los estudiantes que obtuvieron las puntuaciones más bajas)

$$p(X \le N) = 0.25 \Rightarrow p\left(Z \le \frac{N-78}{36}\right) = 0.25$$

$$\frac{N-78}{36} < 0 \qquad 1 - p\left(Z \le -\frac{N-78}{36}\right) = 0.25$$

$$p\left(Z \le \frac{-N+78}{36}\right) = 0.75 \Rightarrow \frac{-N+78}{36} = 0.68 \qquad N = 54$$

$$p(X > 54+5) = p(X > 59) = p\left(Z > \frac{59-78}{36}\right) =$$

$$p(Z > -0.53) = p(Z < 0.53) = 0.7019 = 70.19\%$$

3 Si se sabe que la calificación de un estudiante es mayor que 72 ¿cuál es la probabilidad de que su calificación sea, de hecho, superior a 84?

$$p(X > 84) = p\left(Z > \frac{84-78}{36}\right) = p(Z > 0.16) =$$

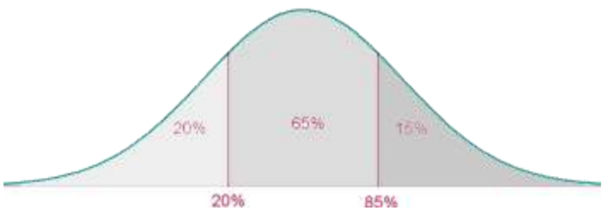
$$= 1 - p(Z < 0.16) = 1 - 0.5636 = 0.4364$$

$$p(x > 84 / x > 72) = \frac{p[x > 84 \cap x > 72]}{p(x > 72)} =$$

$$= \frac{p(x > 84)}{p(x > 72)} = \frac{0.4364}{0.5636} = 0.774$$

Ejercicio 6: Tras un test de cultura general se observa que las puntuaciones obtenidas siguen una distribución una distribución N(65, 18). Se desea clasificar a los examinados en tres grupos (de baja cultura general, de cultura general aceptable, de excelente cultura general) de modo que hay en el primero un 20% la población, un 65% el segundo y un 15% en el tercero. ¿Cuáles han de ser las puntuaciones que marcan el paso de un grupo al otro?

Solución:



$$p(Z \le z_1) = 0.2 \qquad p(Z \le -z_1) = 0.8$$

$$-z_1 = 0.84 \qquad z = -0.84$$

$$\frac{X_1 - 65}{18} = -0.84 \qquad X_1 = 49.88$$

$$p(Z \le z_2) = 0.2 \qquad z_2 = 1.04$$

$$\frac{X_2 - 65}{18} = 1.04 \qquad X_2 = 83.72$$

Baja cultura hasta 49 puntos.
Cultura aceptable entre 50 y 83.
Excelente cultura a partir de 84 puntos.

Ejercicio 7: Varios test de inteligencia dieron una puntuación que sigue una ley normal con media 100 y desviación típica 15

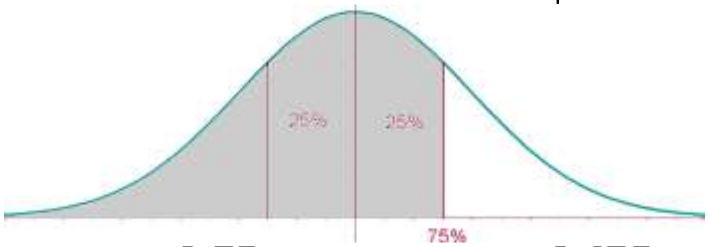
1 Determina el porcentaje de población que obtendría un coeficiente entre 95 y 110
2 ¿Qué intervalo centrado en 100 contiene al 50% de la población?
3 En una población de 2500 individuos ¿cuántos individuos se esperan que tengan un coeficiente superior a 125?

Soluciones:

1 Determina el porcentaje de población que obtendría un coeficiente entre 95 y 110

$$p(95 < X \leq 110) = p\left(\frac{95-100}{15} < Z \leq \frac{110-100}{15}\right) =$$
$$= p(0.33 < Z \leq 0.67) = p(Z \leq 0.67) - [1 - p(Z \leq 0.33)] =$$
$$= 0.7486 - (1 - 0.6293) = 0.3779$$

2 ¿Qué intervalo centrado en 100 contiene al 50% de la población?



$$p = 0.75 \qquad z = 0.675$$
$$\frac{X-100}{15} = 0.675 \qquad X = 110$$
$$(90,\ 110)$$

3 En una población de 2500 individuos ¿cuántos individuos se esperan que tengan un coeficiente superior a 125?

$$p(X > 125) = p\left(Z > \frac{125-100}{15}\right) = p(Z > 1.67) =$$
$$= 1 - p(Z < 1.67) = 1 - 0.9525 = 0.0475 \cdot 2500 = 119$$

Ejercicio 8: En una ciudad una de cada tres familias posee teléfono. Si se eligen al azar 90 familias, calcula la probabilidad de que entre ellas haya por lo menos 30 tengan teléfono

Solución:

$$n = 90 \qquad p = \frac{1}{3} \qquad q = \frac{2}{3}$$
$$n \cdot p > 5 \qquad n \cdot q > 5$$
$$B\left(90, \frac{1}{3}\right) \rightarrow N\left(90 \cdot \frac{1}{3}, \sqrt{90 \cdot \frac{1}{3} \cdot \frac{2}{3}}\right) = N(30,\ 4.47)$$
$$p(X > 30) = p\left(Z > \frac{30-30}{4.47}\right) = p(Z > 0) = 1 - p(Z \leq 0) = 0.5$$

Ejercicio 9: En un examen tipo test de 200 preguntas de elección múltiple, cada pregunta tiene una respuesta correcta y una incorrecta. Se aprueba si se contesta a más de 110 respuestas correctas. Suponiendo que se contesta al azar, calcula la probabilidad de aprobar el examen

Solución:

$$n = 200 \qquad p = 0.5 \qquad q = 0.5$$

$$n \cdot p > 5 \qquad n \cdot q > 5$$

$$B(200, 0.5) \to N\left(200 \cdot 0.5,\ \sqrt{200 \cdot 0.5 \cdot 0.5}\right) = N(100,\ 7.07)$$

$$p(X > 110) = p\left(Z > \frac{110 - 100}{7.07}\right) = p(Z > 1.41) =$$

$$= 1 - p(Z < 1.41) = 1 - 0.92073 = 0.07927$$

Ejercicio 10: Un estudio ha mostrado que, en un cierto barrio, el 60% de los hogares tienen al menos dos televisores Se elige al azar una muestra de 50 hogares en el citado barrio. Se pide:

1 ¿Cuál es la probabilidad de que al menos 20 de los citados hogares tengan cuando menos dos televisores?

2 ¿Cuál es la probabilidad de que entre 35 y 40 hogares tengan cuando menos dos televisores?

Soluciones:

1 ¿Cuál es la probabilidad de que al menos 20 de los citados hogares tengan cuando menos dos televisores?

$$n = 50 \qquad p = 0.6 \qquad q = 0.4$$

$$n \cdot p > 5 \qquad n \cdot q > 5$$

$$B(50, 0.6) \to N\left(50 \cdot 0.6, \sqrt{50 \cdot 0.6 \cdot 0.4}\right) = N(30,\ 3.46)$$

$$p(X > 20) = p\left(Z > \frac{20 - 30}{3.46}\right) =$$

$$p(Z > -2.89) = p(Z \le 2.89) = 0.9981$$

2 ¿Cuál es la probabilidad de que entre 35 y 40 hogares tengan cuando menos dos televisores?

$$p(35 < X \le 40) = p\left(\frac{35 - 30}{3.46} < Z \le \frac{40 - 30}{3.46}\right) =$$

$$= 0.9981 - 0.9265 = 0.0716$$

Aproximación de la distribución binomial por la normal

Teorema de Moivre

Si: **n·p ≥ 0** y **n·q ≥ 0**, la **distribución binomial B(n, p)** se puede aproximar mediante una **distribución normal**:

$$N\left(np, \sqrt{npq}\right)$$

$$N(np, \sqrt{npq})$$

$$B(n, p) \quad \nearrow \quad \searrow \qquad \downarrow$$

$$Z = \frac{X - np}{\sqrt{npq}} \to N(0, 1)$$

Ejemplo: En una ciudad una de cada tres familias posee teléfono. Si se eligen al azar 90 familias, calcula la probabilidad de que entre ellas haya por lo menos 30 tengan teléfono.

$$n = 90 \qquad p = \frac{1}{3} \qquad q = \frac{2}{3}$$

$$n \cdot p > 5 \qquad n \cdot q > 5$$

$$B\left(90, \frac{1}{3}\right) \to N\left(90 \cdot \frac{1}{3}, \sqrt{90 \cdot \frac{1}{3} \cdot \frac{2}{3}}\right) = N(30,\ 4.47)$$

$$p(X > 30) = p\left(Z > \frac{30 - 30}{4.47}\right) = p(Z > 0) = 1 - p(Z \le 0) = 0.5$$