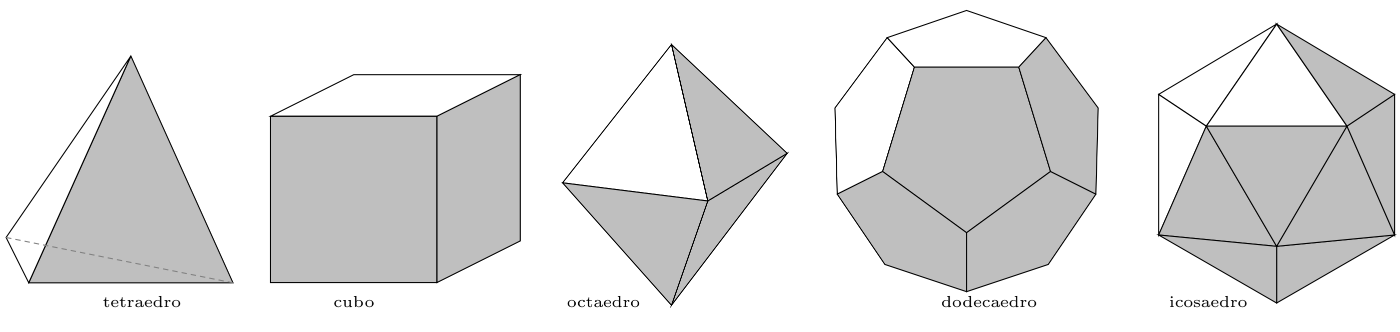
2ºBACHILLERATO

2ºBACHILLERATO

2ºBACHILLERATO

ESTADÍSTICA



7 de enero de 2016

Índice general

1. ESTADISTICA **1**
- 1.1. Introducción . . . 1
- 1.2. Variable estadística . . . 1
- 1.3. Medidas de centralización . . . 3
- 1.4. Medidas de dispersión . . . 3
- 1.5. Observaciones: . . . 5
- 1.6. Problemas . . . 10

2. REGRESION. CORRELACION **13**
- 2.1. Variables estadísticas bidimensionales . . . 13
- 2.2. Cáculo de los parámetros de una variable estadística bidimensional . . . 14
- 2.3. Correlación . . . 14
- 2.4. Recta de regresión de y sobre x . . . 15
- 2.5. Series temporales . . . 17
- 2.6. Números índice . . . 17
- 2.7. Problemas . . . 19

3. PROBABILIDAD **21**
- 3.1. Introducción . . . 21
- 3.2. Sucesos . . . 21
- 3.3. Frecuencia de un suceso . . . 22
- 3.4. Probabilidad . . . 22
- 3.5. Probabilidad con combinatoria . . . 24
  - 3.5.1. Variaciones con repetición . . . 24
  - 3.5.2. Variaciones . . . 24
  - 3.5.3. Permutaciones . . . 25
  - 3.5.4. Combinaciones . . . 25
- 3.6. Sucesos dependientes e independientes . . . 27
- 3.7. Problemas . . . 30

4. VARIABLES ALEATORIAS. DISTRIBUCIÓN DE PROBABILIDAD **39**
- 4.1. Variable aleatoria. Función de distribución de probabilidad . . . 39
- 4.2. Tabla de probabilidades de una variable aleatoria discreta. Histograma de Probabilidad . . . 40
- 4.3. Relación entre variables estadísticas y aleatorias . . . 40

4.4. Parámetros de una variable aleatoria discreta 40
4.5. Distribución binomial 41
4.6. Variable aleatoria continua 42
4.7. Función de densidad de probabilidad de una v.a. continua 42
4.7.1. Parámetros de una variable aleatoria continua: 43
4.8. Distribución normal 43
4.8.1. Aproximación normal de la distribución binomial 45
4.9. Problemas 47

5. DISTRIBUCIÓN MUESTRAL. ESTIMACIÓN ESTADÍSTICA 51
5.1. Muestreo 51
5.2. Distribución muestral de medias. Teorema Central del Límite. 52
5.3. Estimación estadística 54
5.4. Estimas por intervalos de confianza 54
5.5. Decisiones estadísticas. Hipótesis estadísticas 56
5.6. Distribución muestral de proporciones 58
5.7. Problemas 62

Tema 1

ESTADISTICA

1.1. Introducción

Fenómeno aleatorio es aquel en el cual es imposible predecir el resultado en cada realización u observación; ej: lanzar una moneda, extraer una carta de una baraja, número de nacimientos de una ciudad en un mes, etc.

Estadística Descriptiva es la parte de las Matemáticas que se ocupa de proporcionar métodos para recoger, organizar, analizar y resumir listas de datos numéricos de fenómenos aleatorios.

Colectivo o población es el conjunto de elementos con caracteres comunes.

Muestra es un subconjunto o parte representativa de un colectivo.

1.2. Variable estadística

Variable estadística es el carácter común que se considera en los elementos del colectivo. Puede ser:

Variable estadística cualitativa, cuando el carácter que se considera no es numérico, ej: colectivo: alumnos de un instituto, variable cualitativa color del pelo

Variable estadística cuantitativa, cuando el carácter que se considera es numérico. Se suele representar por x_i;
ej: colectivo: alumnos de un instituto, variable cuantitativa la estatura.

Frecuencia de un dato es el número de veces que aparece ese dato. Se suele representar por n_i.Tambien se llama frecuencia absoluta. La suma de las frecuencias es igual al número de datos.

Frecuencia relativa de un dato es la frecuencia dividida por el número de datos. Se suele representar por f_i. La suma de las frecuencias relativas es igual a 1.

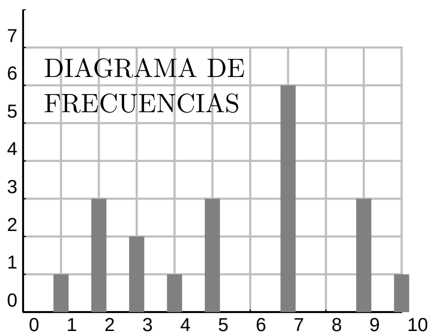
Frecuencia acumulada hasta un dato es la suma de las frecuencias de ese dato y de los anteriores. Se suele representar por la misma letra mayúscula por ejemplo para las fecuencias absolutas N_i.

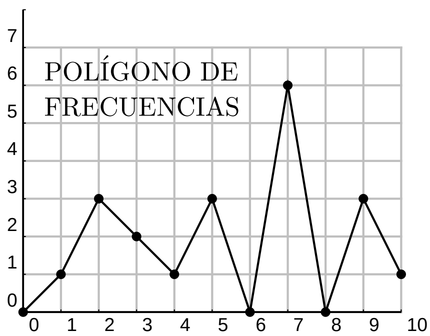
Es decir N_i = frecuencia de x_i o menor = frecuencia$(x \leq x_i)$.

Ejemplo * Supongamos que las calificaciones de 20 alumnos vienen dadas por la serie estadística:

2,4,5,9,9,10,7,3,2,5,7,3,7,7,5,1,2,7,7,9

var.est	frecuencias	frec. rel	frec. acum.	frec .rel. acum.
x_i	n_i	f_i	N_i	F_i
0	0	0	0	0
1	1	0'05	1	0'05
2	3	0'15	4	0'20
3	2	0'10	6	0'30
4	1	0'05	7	0'35
5	3	0'15	10	0'50
6	0	0	10	0'50
7	6	0'30	16	0'80
8	0	0	16	0'80
9	3	0'15	19	0'95
10	1	0'05	20	1
	$\Sigma n_i = 20$			





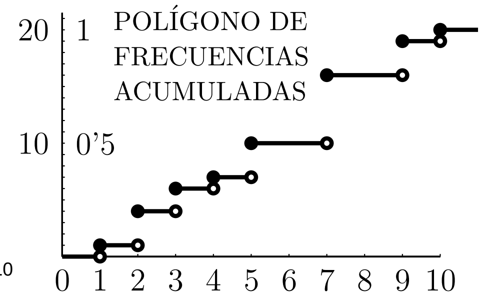


Diagrama de sectores Sea trata de repartir un círculo en sectores proporcionales a las frecuencias:

Por ejemplo para mostrar la proporción de suspensos y aprobados:

Para hacer el diagrama de sectores se plantea la regla de tres: si todo el círculo 360^0 corresponde con 20 notas, a los 7 suspensos le corresponde x, $x = 126^0$

Normalmente interesa dar un resumen numérico de los datos de un fenómeno aleatorio. Para ello se requieren dos números: uno que dé un valor medio representativo y otro que indique lo alejados que están los datos entre sí.

Tenemos entonces las medidas de **centralización** que indican valores medios representativos y las de **dispersión** que indican lo separados que están los datos.

1.3. Medidas de centralización

Moda es el valor de la variable estadística que tiene mayor frecuencia. En el ejemplo* de las notas de clase: 7.

Mediana es el valor central del conjunto ordenado de datos x_i, el que deja a la izquierda la mitad de los datos cuando los datos están ordenados. En el ejemplo* de las notas de clase: 1 2 2 2 3 3 4 5 5 5*7 7 7 7 7 7 9 9 9 10 la mitad está entre $N_i = 10$ y 11, o sea entre 5 y 7, (pasa cuando es par el número de datos) y se toma la semisuma: mediana $= \dfrac{5+7}{2} = 6$.

Media es la media aritmética: se suman todos los datos y se divide por el número de datos.

$$\boxed{\text{media sin frecuencias: } \bar{x} = \frac{\Sigma x_i}{N}}$$

Si conviene considerar las frecuencias, como cada dato se sumaría un número de veces igual a su frecuencia resulta:

$$\boxed{\text{media con frecuencias: } \bar{x} = \frac{\Sigma x_i n_i}{\Sigma n_i}}$$

En el ejemplo* de las notas de clase: $\bar{x} = \dfrac{111}{20} = 5'55$

1.4. Medidas de dispersión

Rango o recorrido es la diferencia entre los valores más grande y más pequeño, en el ejemplo: $10 - 1 = 9$.

Desviación media Desviación de un valor respecto de la media es $x_i - \bar{x}$.

Se llama desviación media a la media de los valores absolutos de las desviaciones. Como los valores absolutos se trabajan mal con calculadora en la práctica se usa:

Varianza es la media aritmética de los cuadrados de las desviaciones.

Desviación típica es la raíz cuadrada de la varianza, es decir, la raíz cuadrada de la media aritmética de los cuadrados de las desviaciones, se representa por σ:

$$\boxed{\text{Desviación típica sin frecuencias: } \sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}}$$

$$\boxed{\text{Desviación típica con frecuencias: } \sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 n_i}{\Sigma n_i}}}$$

Ejemplos: (sin calculadora estadística)

1. (Datos sin frecuencias) Dados los números: 3 6 12
   a) Hallar la media. b) Hallar la desviación típica.

   $\boxed{\text{Media: } \bar{x} = \frac{3+6+12}{3} = 7}$

   Cálculo de la desviación típica:

   Desviaciones $3-7=-4, \quad 6-7=-1, \quad 12-7=5.$

   Cuadrado de las desviaciones: 16, 1, 25

   Varianza: media de los cuadrados de las desviaciones: Varianza: $\sigma^2 = \frac{16+1+5}{3} = 14$

   Desviación típica: Raíz cuadrada de la varianza: $\boxed{\text{Desviación típica: } \sigma = \sqrt{14} = 3'74}$

2. (Datos con frecuencias) Dados los datos y sus frecuencias. a) Hallar la media. b) Hallar la desviación típica.

x_i	n_i	$x_i \cdot n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
6	8	48	-4	16	128
8	20	160	-2	4	80
16	12	192	6	36	432
	$\Sigma n_i = 40$	$\Sigma x_i n_i = 400$			$\Sigma(x_i - \bar{x})^2 \cdot n_i = 640$

   Media: $\bar{x} = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{400}{40} = 10$

   Varianza: $\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2 n_i}{\Sigma n_i} = \frac{640}{40} = 16$

   Desviación típica: $\sigma = \sqrt{16} = 4$

Ejercicios: (sin calculadora estadística)

1. (Datos sin frecuencias) Dados los números: 6 8 6 9 6
   a) Hallar la media. b) Hallar la desviación típica. c) Hallar la mediana. d) Hallar la moda.

2. (Datos con frecuencias) Dados los datos y sus frecuencias. a) Hallar la media. b) Hallar la desviación típica.

x_i	n_i
5	4
9	10
25	2

Ejercicios:

1. Dados los números: 9 3 8 10 1 9 5 6 8 2 3 10 10 1 10 2 9 5
   a) Hallar la media. b) Hallar la desviación típica. c) Hallar la mediana. d) Hallar la moda.
   media = 6,17 des. Tip. = 3,34 num. Dat= 18 mediana= 7 moda= 10

2. Dados los datos y sus frecuencias:

x_i	2	3	5	7	9	12
n_i	13	12	18	16	14	13

   a) Hallar la media. b) Hallar la desviación típica. media = 6,35 des. Tip. = 3,29 num. Dat= 86

3. Dados los números: 2 2 6 6 8 5 6 4 10 3 10 1 8 5 5 6 6 1 6 9 10 4 4 3 6 2 1 3 8 10 9 6 3 3 5 5 3 7 6 9 2 6 8 6 3 3 5 8 8 8
   a) Hallar la media. b) Hallar la desviación típica. c) Hallar la mediana. d) Hallar la moda.
   media = 5,46 des. Tip. = 2,59 num. Dat= 50 mediana= 5,5 moda= 6

4. Dados los datos y sus frecuencias:

x_i	1	2	4	5	8	9	11	12	14	15	18	19	21
n_i	19	13	12	12	11	11	14	19	18	14	16	18	13

   a) Hallar la media. b) Hallar la desviación típica. media = 10,98 des. Tip. = 6,39 num. Dat= 190

Media y desviación típica Son las dos medidas más importantes

En el ejemplo de las veinte notas se obtiene: $\sigma = 2'67$. Recordemos que la media era 5'55. Nos dicen que si tomamos un alumno al azar lo más probable es que haya obtenido una nota próxima a 5'55 con una diferencia de $\pm 2'67$.

Pero sobre todo sirve para comparar dos variables; si otro curso tiene como media 6'5 y desviación típica 1'2, podríamos afirmar con total seguridad que estos últimos alumnos han sacado mejores notas y que éstas son más uniformes.

1.5. Observaciones:

1. **Agrupamiento en clases:**

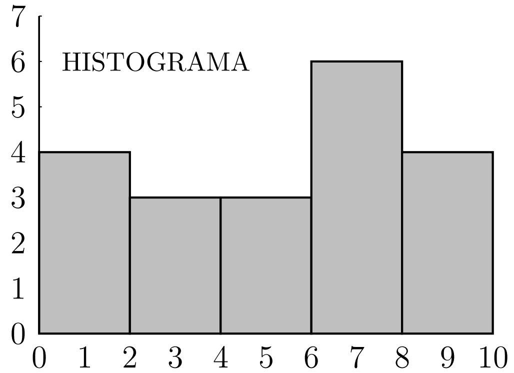
   Si interesa porque hay muchos valores distintos, se suelen agrupar los valores en **intervalos de clase** por ej. las tallas de 5 cm en 5 cm, el centro de cada intervalo se llama **marca de clase** y se considera éste como el valor de la variable estadística.

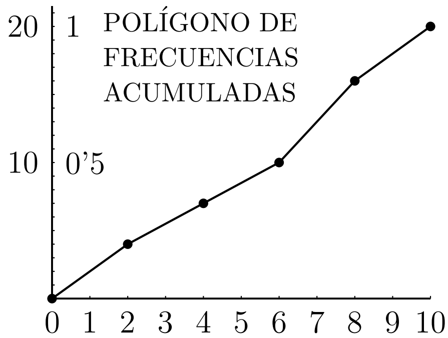
   Un criterio para decidir el número de intervalos de clase puede ser el de Norcliffe:

$$\text{n}^0 \text{ de clases} \approx \sqrt{\text{n}^0 \text{ de datos}}$$

   En el ejemplo * n^0 clases $\approx \sqrt{20} \approx 5$ intervalos iguales, el intervalo total es 10, la longitud de cada intervalo de clase es $10/5 = 2$

int.clase	marca clase				
	x_i	n_i	f_i	N_i	F_i
[0,2]	1	4	0'20	4	0'20
(2,4]	3	3	0'15	7	0'35
(4,6]	5	3	0'15	10	0'50
(6,8]	7	6	0'30	16	0'80
(8,10]	9	4	0'20	20	1
	$\Sigma n_i = 20$				





2. **Cuantiles:**

   Análogamente a como la mediana ocupa el lugar medio de la serie estadística, el 1^{er} cuartil Q_1 deja a la izquierda $\frac{1}{4}$ del total de la serie de datos ordenada, o sea de 100 deja 25, el 3^{er} decil D_3 el 30 %, el percentil 99 P_{99} el 99 %, etc.

   También se considera: rango intercuartílico $Q_3 - Q_1$, rango interdecílico $D_9 - D_1$, rango intercentílico $P_{99} - P_1$.

   Cuando hay que hallar varios compensa hacer la columna de frecuencias absolutas acumuladas e incluso las de los % acumulados.

   Ejemplo: Hallar Q_1, Q_3, P_{30}, P_{77}

   $Q_1 : \quad \dfrac{N}{4} = \dfrac{40}{4} = 10$; deja a la izda 10; $Q_1 = 4$

   $Q_3 : \quad \dfrac{3N}{4} = \dfrac{120}{4} = 30$; deja a la izda 30; $Q_3 = \dfrac{6+7}{2} = 6'5$

   $P_{30} :$ por regla de tres $\begin{matrix} 100 & - & 30 \\ 40 & - & y \end{matrix}$ $y = 12$; deja a la izquierda 12; $P_{30} = 4$

   $P_{77} :$ $\begin{matrix} 100 & - & 77 \\ 40 & - & y \end{matrix}$ $y = 30'8$; deja a la izquierda 30'8; $P_{77} = 7$

x_i	n_i	N_i
1	2	2
2	2	4
3	4	8
4	5	13
5	8	21
6	9	30
7	3	33
8	4	37
9	3	40
	40	

3. **Diagrama de caja:**

   Un Diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes".

   Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q_1, mediana y Q_3, y sobre la existencia de valores atípicos y la simetría de la distribución.

Los bigotes, las líneas que se extienden desde la caja, se extienden hasta los valores máximo y mínimo de la la serie o hasta 1'5 veces el Rango Inter Cuartilico RIC ($Q_3 - Q_1$).

Cuando los datos se extienden más allá de esto, significa que hay valores atípicos en la serie. Se representa el punto por un asterisco.

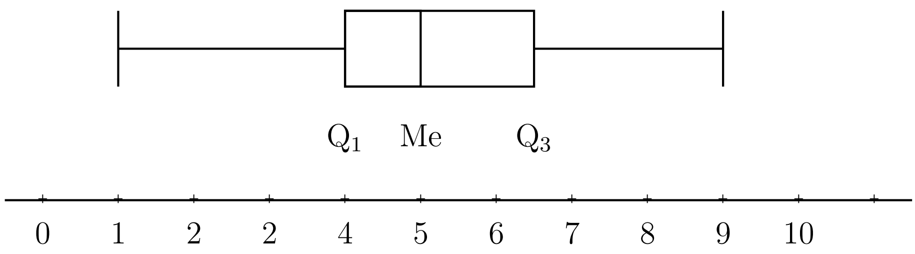
Para ello, se consideran atípicos los valores son aquellos inferiores a $Q_1 - 1'5 \cdot RIC$ o superiores a $Q_3 + 1'5 \cdot RIC$.

Además, se pueden considerar valores extremadamente atípicos aquellos que exceden $Q_1 - 3 \cdot RIC$ o $Q_3 + 3 \cdot RIC$.

Proporcionan una visión general de la simetría de la distribución de los datos; si la mediana no está en el centro del rectángulo, la distribución no es simétrica.

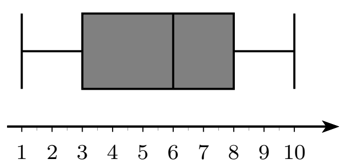
Son útiles para ver la presencia de valores atípicos también llamados outliers.

Pertenece a las herramientas de las estadística descriptiva. Permite ver como es la dispersión de los puntos con la mediana, los percentiles 25 y 75 y los valores máximos y mínimos.

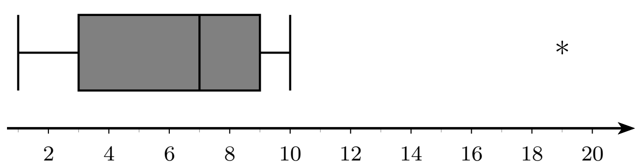


Ejemplo:

datos: 2 4 5 9 9 10 10 3 2 5 7 3 7 7 5 1 2 7 7 9



Si añadimos el dato: 19, Resulta:



4. Si tenemos varios grupos de medias y número de datos $\bar{x}_A, N_A$; $\bar{x}_B, N_B$; $\bar{x}_C, N_C \dots$ respectivamente la media de la unión de las distribuciones es:
$\bar{x} = \dfrac{\bar{x}_A \cdot N_A + \bar{x}_B \cdot N_B + \bar{x}_C \cdot N_C + \dots}{N_A + N_B + N_C + \dots}$ es lo que se llama media ponderada

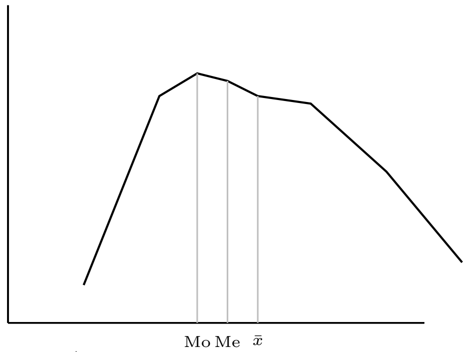
Ejemplo: Un granjero tiene una explotación con dos establos de vacas. Cada uno de los 13 animales del primero produce una media de 30 litros de leche por día, mientras que en otro hay 17 animales y la media es de 28 litros. ¿Cual es la producción media por vaca y día de la explotación?.

$\bar{x} = \dfrac{30 \times 13 + 28 \times 17}{30} = 28'86$ litros

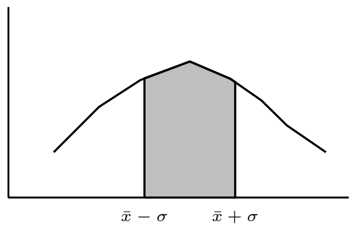
nota: no hay fórmula análoga para la desv. típica.

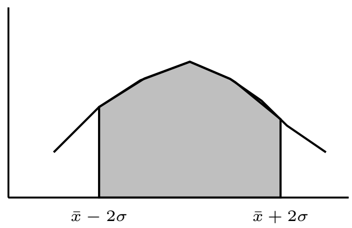
5. Para polígonos de frecuencias unimodales y aproximadamente simétricos se tiene la relación empírica:

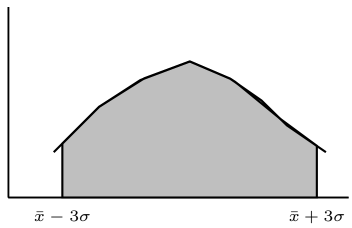
   media $-$ moda $\approx$ 3(media - mediana)



6. Para polígonos de frecuencias unimodales y aproximadamente simétricos se tienen las relaciones:







   en el intervalo: $(\bar{x}-\sigma, \bar{x}+\sigma)$ se encuentra aproximadamente el 68 % de los datos

   en el intervalo: $(\bar{x}-2\sigma, \bar{x}+2\sigma)$ se encuentra aproximadamente el 95 % de los datos

   en el intervalo: $(\bar{x}-3\sigma, \bar{x}+3\sigma)$ se encuentra aproximadamente el 99 % de los datos

7. **Tipificación de variables**

   Dada una serie estadística x_i, se llaman puntuaciones típicas a los valores: $\dfrac{x_i - \bar{x}}{\sigma}$

   Sirven para comparar puntuaciones de un individuo en distintas distribuciones.

   Ejemplo Un alumno ha contestado a dos tests, obteniendo las siguientes puntuaciones: Test A: 50 puntos, Test B: 32 puntos. La puntuación media y las desviaciones típicas del curso en los dos tests han sido:

Test A: $\bar{x}_A = 45, \quad \sigma_A = 6$

Test B: $\bar{x}_B = 26, \quad \sigma_B = 2$

¿En cuál de los dos tests ha obtenido, comparativamente con el grupo, mejor resultado el alumno?

Test A puntuación típica: $\dfrac{50-45}{6} = 0'83$

Test B puntuación típica $\dfrac{32-26}{2} = 3$

Comparado con el resto del grupo el alumno ha obtenido mejor puntuación en el segundo test.

8. La desviación típica viene dada también por:

   Desviación típica sin frecuencias: $\sigma = \sqrt{\dfrac{\Sigma x_i^2}{N} - \bar{x}^2}$

   Desviación típica con frecuencias: $\sigma = \sqrt{\dfrac{\Sigma x_i^2 n_i}{\Sigma n_i} - \bar{x}^2}$

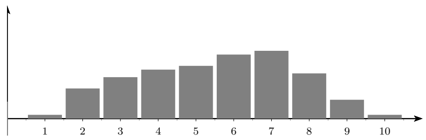
9. En todos los cálculos en vez de n_i podríamos utilizar frecuencias relativas f_i, pues es dividir numerador y denominador por Σn_i:

   Media: $\bar{x} = \Sigma x_i f_i$

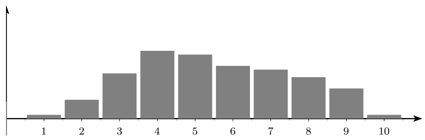
   Desviación típica: $\sigma = \sqrt{\Sigma(x_i - \bar{x})^2 f_i} = \sqrt{\Sigma x_i^2 f_i - \bar{x}^2}$

10. **ASIMETRIA**

    ASIMETRÍA POSITIVA

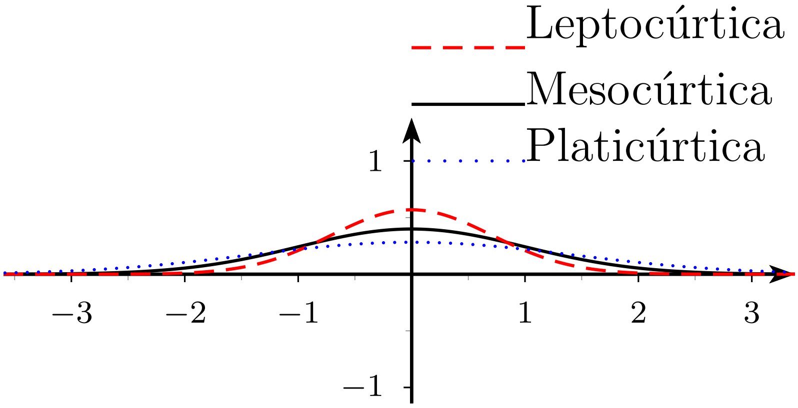


    ASIMETRÍA NEGATIVA



11. **CURTOSIS**

    Sr refiere al grado de apuntamiento y aplastamiento de la curva del polígono de frecuencias.



1.6. Problemas

1. Dada la distribución de frecuencias :

x_i	n_i
1	1
2	3
3	0
4	2
5	4
6	0

a) Constrúyase una tabla en la que aparezcan frecuencias absolutas, relativas y absolutas acumuladas. b) Represéntese mediante un diagrama de barras la distribución dada y su correspondiente polígono de frecuencias.

2. (Sin calculadora estadística) Dados los datos y sus frecuencias. a) Hallar la media. b) Hallar la desviación típica.

x_i	n_i
5	7
7	13
10	12
13	2

3. (Sin calculadora estadística) Dados los datos y sus frecuencias. a) Hallar la media. b) Hallar la desviación típica.

x_i	n_i
5	8
8	16
13	12

4. El número de hijos de 10 familias, seleccionadas aleatoriamente, es el siguiente: 5, 2, 0, 6, 3 ,1, 2, 3, 1, 4. Hallar la mediana y la varianza.

Solución: media = 2'7, des.tip. = 1'79, mediana = 2'5, var = 3'21

5. Se efectúan 10 series de 5 tiradas de esa moneda. Se considera la variable estadística "número de caras en cada serie", resultando: 3,4,5,1,2,3,2,3,4,2

a) Hacer la tabla de frecuencias relativas.

b) Dibujar diagrama de barras de anchura 1 de frecuencias relativas.

d) Hallar la media y la desviación típica.

6. Una variable estadística tiene las siguientes frecuencias relativas:

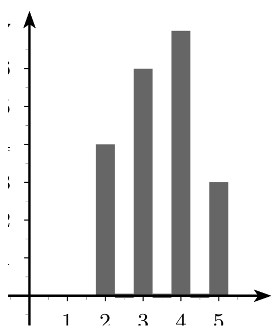
x_i	0	1	2	3
f_i	0'4	0'3	0'2	0'1

a) Dibujar el polígono de frecuencias relativas.

b) Hallar la media y la desviación típica.

c) Hallar la frecuencia relativa acumulada del valor $x_i = 2$

7. Dado el diagrama de barras



a) Hallar las frecuencias relativas.

b) Hallar la media y la desviación típica.

8. En una bolsa hay 8 bolas blancas y 5 azules. Se hacen 10 series de 3 extracciones

con devolución. Consideramos el número de bolas blancas que salen en cada serie.

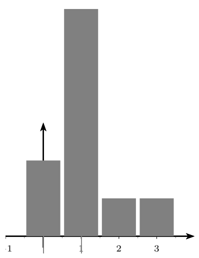
El número de bolas blancas en cada serie ha sido: 3, 1, 0, 1, 0, 1, 1, 1, 2, 1

a) Hallar las frecuencias relativas y hacer un diagrama de barras de ancho uno.

b) Hallar la media y la desviación típica.

Solución:

x_i	f_i
0	0,2
1	0,6
2	0,1
3	0,1



$\mu = 1,1, \quad \sigma = 0,83$

9. En un reclutamiento militar se ha tomado una muestra de dieciseis jóvenes obteniéndose las siguientes estaturas en cms. : 172, 161, 168, 182, 167, 179, 175, 198, 180, 166, 164, 174, 185, 177, 191, 173 Agrupar los datos en intervalos de 10 cms. Escribir la tabla estadística y calcular la media y la desviación típica: a) directamente, b) agrupando los datos.

   nota: aunque no lo concreta el problema tomar como extremo más pequeño 160 para unificar: $[160 - 170) \ldots$.

   Solución: a)media = 175'75, des.tip. = 9'38 b) media = 176'625, des.tip. = 9'66

10. Durante el mes de julio, en una determinada ciudad de la costa levantina, se han registrado las siguientes temperaturas máximas: 32, 31, 28, 28, 33, 32, 31, 30, 31, 27, 28, 29, 29, 30, 32, 31, 31, 30, 30, 29, 29, 29, 30, 31, 30, 34, 33, 33, 32, 33, 32 Hallar, la moda y los percentiles de orden 30 y 70. Hacer un diagrama de caja.

    Solución: moda = 30, 31, P_{30} = 30, P_{70} = 32

11. En el departamento de selección de personal de una empresa se ha aplicado un test de inteligencia a los mandos intermedios, obteniéndose los siguientes resultados: 63, 69, 71, 56, 58, 68, 73, 67, 65, 72, 78, 56, 68, 65, 72, 58, 68, 71, 63, 71, 65, 77, 51, 81, 67, 67, 65, 66, 68, 69, 61, 65, 48.

    a) Hallar los cuartiles y el recorrido intercuartílico.

    b) Los percentiles de orden 90 y 10, y el recorrido interdecílico.

    c) Hacer diagrama de caja.

    Solución: Q_1 = 65, Q_2 = 67, Q_3 = 71, Q_3 - Q_1 = 6, P_{90} = 73 , P_{10} = 58 , P_{90} - P_{10} = 15

12. Un tirador hace 60 series de 5 disparos. La frecuencia relativa acumulada de número de aciertos en cada serie viene dada por la tabla:

x_i	0	1	2	3	4	5
F_i	0'12	0'31	0'59	0'77	0'93	1

a) Hallar las frecuencias relativas y hacer un diagrama de barras de ancho uno.

b) Hallar la media y la desviación típica.

Solución: a) media= 2'28 des. tip.= 1'428

Tema 2

REGRESION. CORRELACION

2.1. Variables estadísticas bidimensionales

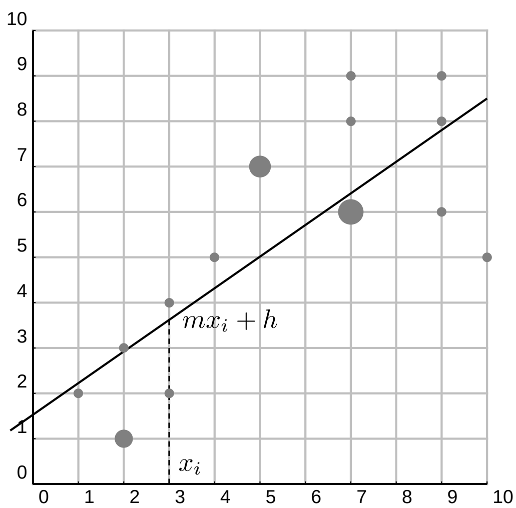
Cuando estudiamos dos variables estadísticas puede interesar ver si están relacionados sus valores, por ejemplo en las calificaciones en dos asignaturas, Física y Matemáticas, de 20 alumnos, cabe esperar que a una nota alta en Física corresponda otra alta en Matemáticas.

Para ello se consideran simultáneamente las dos variables estadísticas, se tiene entonces una **variable estadística bidimensional**.

Consideremos en el ejemplo anterior las calificaciones:

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

Podemos representar en el plano cada pareja de valores, obtenemos así los **diagramas de dispersión** llamados también **nube de puntos**. Estos puntos no se situarán sobre una línea determinada (a diferencia de las funciones, en los que cada valor de una variable determina el valor de la otra), pero cuando hay dependencia entre los valores sí aparece cierta forma en la nube.



Se llama ajuste de la nube de puntos, al problema de encontrar la línea que mejor se adapta a la nube de puntos. Nos limitaremos a encontrar rectas. Una vez halladas nos darán el valor más probable para una de las variables correspondiente a un valor dado de la otra.

Recta de regresión de y sobre x: Es la recta $y = mx+h$, de manera que el error cometido al tomar como valor y_i correspondiente a x_i, el dado por la recta: $y = mx_i + h$ sea mínimo, o sea la recta que hace mínimas las diferencias $y_i - (mx_i + h)$.

m se llama **coeficiente de regresión de y sobre x**

2.2. Cáculo de los parámetros de una variable estadística bidimensional

x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\cdot(y_i-\bar{y})$
9	9	0	0	2	4	0
7	7	-2	4	0	0	0
8	7	-1	1	0	0	0
12	5	3	9	-2	4	-6
$\Sigma x_i=36$	$\Sigma y_i=28$		$\Sigma(x_i-\bar{x})^2=14$		$\Sigma(y_i-\bar{y})^2=8$	$\Sigma(x_i-\bar{x})\cdot(y_i-\bar{y})=-6$

Media de x: $\bar{x}=\dfrac{\Sigma x_i}{N}=\dfrac{36}{4}=9$

Varianza de x: $\sigma_x^2=\dfrac{\Sigma(x_i-\bar{x})^2}{N}=\dfrac{14}{4}=3'5$ Desviación típica de x: $\sigma_x=\sqrt{3'5}=1'87$

Media de y: $\bar{y}=\dfrac{\Sigma y_i}{N}=\dfrac{28}{4}=7$

Varianza de y: $\sigma_y^2=\dfrac{\Sigma(y_i-\bar{y})^2}{N}=\dfrac{8}{4}=2$ Desviación típica de y: $\sigma_y=\sqrt{2}=1'41$

Covarianza. Se llama **covarianza** a la media de los productos de las desviaciones de las dos componentes de la variable bidimensional, $\boxed{\sigma_{xy}=\dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}=\dfrac{\Sigma x_i\cdot y_i}{N}-\bar{x}\cdot\bar{y}}$

$$\sigma_{xy}=\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}=\frac{-6}{4}=-1'5$$

Coeficiente de correlación. Viene dado por la covarianza dividida por el producto de las desviaciones típicas: $\boxed{r=\dfrac{\sigma_{xy}}{\sigma_x.\sigma_y}}$

$$r=\frac{\sigma_{xy}}{\sigma_x.\sigma_y}=\frac{-1'5}{\sqrt{3'5}\cdot\sqrt{2}}=-0'56$$

Recta de regresión y/x: $\boxed{y-\bar{y}=\dfrac{\sigma_{xy}}{\sigma_x^2}.(x-\bar{x})}$

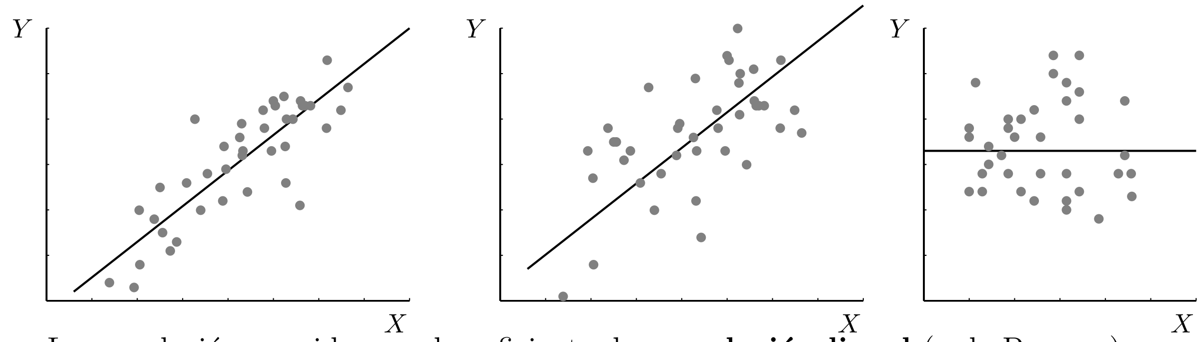
$$y-7=\frac{-1'5}{3'5}(x-9)$$

2.3. Correlación

Es el grado de mutua dependencia entre las dos variables estadísticas que componen la variable bidimensional.

Cuanto mayor es la correlación más estrecha es la banda en la que se sitúan los puntos de la nube.

CORRELACIÓN CORRELACIÓN PEQUEÑA INCORRELACIÓN



La correlación se mide por el coeficiente de **correlación lineal** (o de Pearson).

Se tiene que $r \in [-1, 1]$:

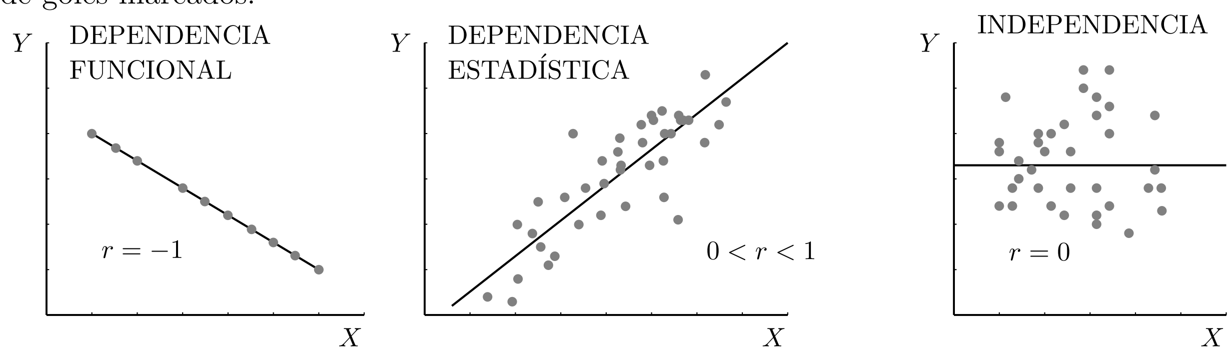
Cuanto más próximo a 1 está $|r|$ mayor es la correlación, más estrecha es la banda en que están los puntos alrededor de la recta de regresión.

Si $r = \pm 1$ entonces hay dependencia funcional, los puntos están en la recta.

Cuanto más próximo a 0 está r menor es la correlación, más redonda es la nube de puntos. Si es 0 hay independencia lineal.

Si $r > 0$ es correlación positiva la recta es creciente Si $r < 0$ es correlación negativa la recta es decreciente

ejemplo de correlación negativa: puesto de calificación en un campeonato de liga y número de goles marcados.



De todas formas para valorar la correlación hay que tener en cuenta el contexto: así por ejemplo una correlación $r = 0'6$ entre "estaturas" y "pesos" de los soldados de un regimiento es baja; una correlación $r = 0'6$ entre "la nota de matemáticas" y "el número total de horas de estudio a la semana" de los alumnos de una clase es notablemente alta.

2.4. Recta de regresión de y sobre x

Cuando la correlación es suficientemente alta, tiene sentido considerar la recta de regresión de y sobre x "y/x" que pasa por el punto de coordenadas las medias $(\bar{x}, \bar{y})$:

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

la pendiente es el **coeficiente de regresión de y sobre x** y es igual a la covarianza dividida por la varianza de x:

Ejemplo En las notas de Física y Matemáticas de los 20 alumnos.

x_i	2	4	5	9	9	10	7	3	2	5	7	9	7	3	7	7	5	1	2	7
y_i	3	5	7	9	6	5	6	4	1	7	6	8	6	2	8	6	7	2	1	9

Las medias son: $\bar{x} = 5'55, \bar{y} = 5'40$, resulta: $\sigma_{xy} = 4'98$

El coeficiente de correlación lineal de la Física y las Matemáticas, cuyas desviaciones típicas son $\sigma_x = 2'67, \sigma_y = 2'43$, resulta: $r = \dfrac{4'98}{2'67 \cdot 2'43} = 0'76$

La varianza de la Física es: $\sigma_x^2 = 7'15$ resulta:

recta de regresión de y sobre x: $y - 5'4 = \dfrac{4'98}{7'15}(x - 5'55)$

El **valor esperado** de y_0 para un valor dado x_0, obtenido a partir de la recta de regresión y/x es más fiable cuanto mayor sea $|r|$ y más próximo a la media de x esté x_0. En el ejemplo, el valor esperado para una nota de Física de 5 es de: $y - 5'40 = 0'7(5 - 5'55)$; resulta $y = 5'03$, valor de alto grado de fiabilidad.

Ejemplo Hallar el coeficiente de correlación y el valor esperado para $x = 10$ en la variable bidimensional:

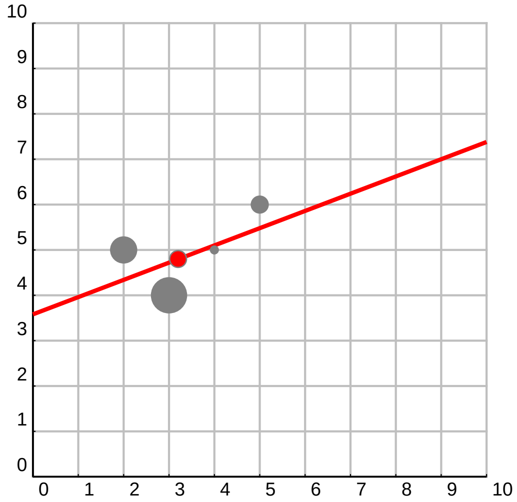
x_i	y_i	n_i
5	6	2
3	4	4
4	5	1
2	5	3

x_i	y_i	n_i	$x_i n_i$	$y_i n_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2 n_i$	$(y_i - \bar{y})^2 n_i$
5	6	2	10	12	1'8	1,2	3,24	1'44	6'48	2'88
3	4	4	12	16	-0'2	-0'8	0'04	0'64	0'16	2'56
4	5	1	4	5	0'8	0'2	0'64	0'04	0'64	0'04
2	5	3	6	15	-1'2	0'2	1'44	0'04	4'32	0'12

$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y}) n_i$
2'16	4'32
0'16	0'64
0'168	0'16
-0'24	-0'72

$\Sigma x_i n_i = 32$; $\Sigma y_i n_i = 48$; $\Sigma (x_i - \bar{x})^2 n_i = 11'6$; $\Sigma (y_i - \bar{y})^2 n_i = 5'6$; $\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) n_i = 4'4$

$\bar{x} = 3'2$; $\bar{y} = 4'8$; $\sigma_x^2 = 1'16$; $\sigma_x = 1'07$; $\sigma_y^2 = 0'56$; $\sigma_y = 0'75$; $\sigma_{xy} = 0'44$



coeficiente de correlación: $r = 0'54$

coeficiente de regresión y/x: $0'379$

recta de regresión y/x: $y - 4'8 = 0'379(x - 3'2)$

Para $x = 10$: $y - 4'8 = 0'379(10 - 3'2)$; $\quad y = 7'38$

x	$3'2$	10
y	$4'8$	$7'38$

2.5. Series temporales

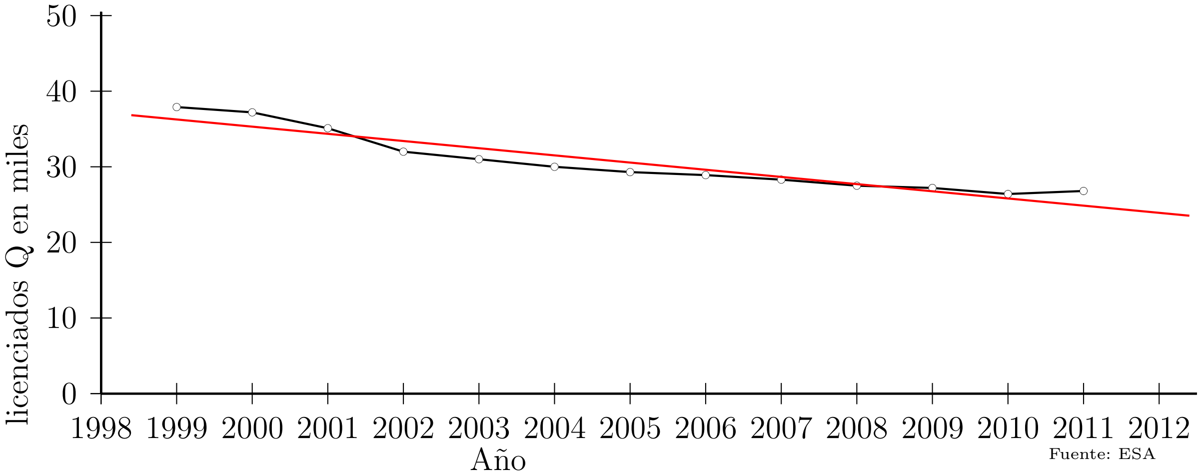
Una **serie temporal** es una variable estadística cuyas observaciones están ordenadas temporalmente. Por ejemplo el número de alumnos matriculados cada año en Selectividad en la Universidad de Murcia, el volumen de precipitaciones mensuales en la Región.

Resulta una variable bidimensional

El principal objetivo de las series de tiempo es hacer proyecciones o pronósticos sobre una actividad futura, suponiendo estables las condiciones y variaciones registradas hasta la fecha.

Ejemplo: Licenciados en Ciencias Q en Qurrilandia en miles:

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
$37,9$	$37,2$	$35,1$	32	31	30	$29,3$	$28,9$	$28,3$	$27,5$	$27,2$	$26,4$	$26,8$



2.6. Números índice

Número índice es una medida estadística que sirve para comparar una magnitud en distintos momentos del tiempo con respecto a uno que se toma como referencia.

Un **índice simple** es el cociente entre la magnitud en el período corriente y la magnitud en el período base. Generalmente se multiplica por cien y se lee en porcentaje.

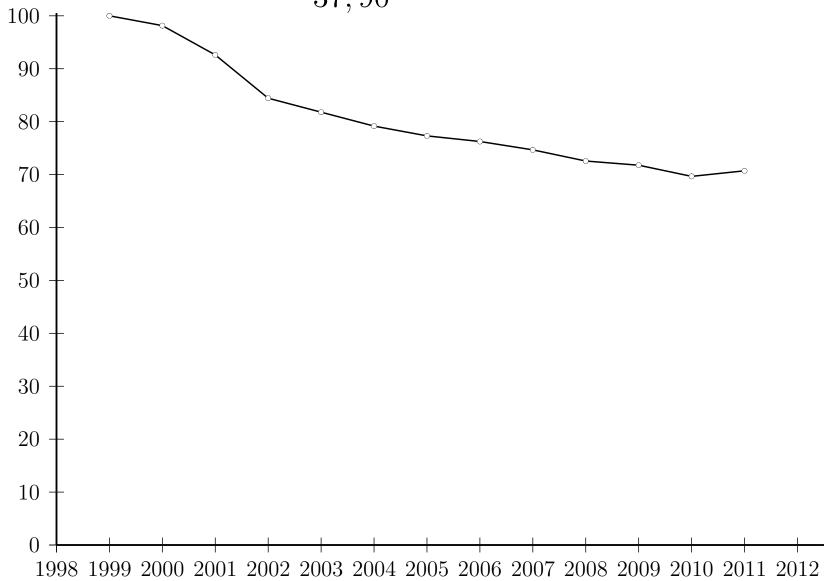
Período base es la situación inicial o el periodo tomado como referencia, se representa: p_0.

$$I_{t/0}(p) = \frac{p_t}{p_0} \times 100$$

Ejemplo: Licenciados en Ciencias Q en Qurrilandia en miles:

Año	lic. miles	Índice simple $I_{t/0}(p)$
1998	$37,90$	$100,00$
1999	$37,20$	$98,15$
2000	$35,10$	$92,61$
2001	$32,00$	$84,43$
2002	$31,00$	$81,79$
2003	$30,00$	$79,16$
2004	$29,30$	$77,31$
2005	$28,90$	$76,25$
2006	$28,30$	$74,67$
2007	$27,50$	$72,56$
2008	$27,20$	$71,77$
2009	$26,40$	$69,66$
2010	$26,80$	$70,71$

Cálculo año 2004: $\dfrac{29,30}{37,90} \times 100 = 77,31$



Indice complejo pretende hacer comparaciones sobre una magnitud compleja, consistente en la agregación de varias magnitudes simples. Por ejemplo el Índice de Precios al Consumo, IPC.

2.7. Problemas

Sin calculadora estadística:

a) Hallar los parámetros de la variable estadística bidimensional: $(13,12),(17,17),(19,15),(23,24)$

x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\cdot(y_i-\bar{y})$
13	12	-5	25	-5	25	25
17	17	-1	1	0	0	0
19	15	1	1	-20	4	-2
23	24	5	25	7	49	35
$\Sigma x_i=72$	$\Sigma y_i=68$		$\Sigma(x_i-\bar{x})^2=52$		$\Sigma(y_i-\bar{y})^2=78$	$\Sigma(x_i-\bar{x})\cdot(y_i-\bar{y})=58$

$\bar{x}=\dfrac{\Sigma x_i}{N}=\dfrac{72}{4}=18 \quad \sigma_x^2=\dfrac{\Sigma(x_i-\bar{x})^2}{N}=\dfrac{52}{4}=13 \quad \sigma_x=\sqrt{13}=3'61$

$\bar{y}=\dfrac{\Sigma y_i}{N}=\dfrac{68}{4}=17\ \sigma_y^2=\dfrac{\Sigma(y_i-\bar{y})^2}{N}=\dfrac{78}{4}=19'5\ \sigma_y=\sqrt{19'5}=4'42$

$\sigma_{xy}=\dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}=\dfrac{58}{4}=14'5$

$r=\dfrac{\sigma_{xy}}{\sigma_x.\sigma_y}=\dfrac{14'5}{\sqrt{13}\cdot\sqrt{19'5}}=0'91$

recta y/x: $y-\bar{y}=\dfrac{\sigma_{xy}}{\sigma_x^2}.(x-\bar{x});\quad y-17=\dfrac{14'5}{13}(x-18)$

b) Hallar los parámetros de la variable estadística bidimensional: $(4,12),(12,16),(20,20),(24,28)$

$r=0'94625 \quad y-19=\dfrac{43}{59}(x-15)$

1. El cambio de la moneda de dos naciones respecto al marco alemán ha sufrido las siguientes fluctuaciones:

   1’3; 2’5; 1’2; 1’1; 0’9;

   1’1; 2’3; 0’9; 1’0; 0’8.

   Indica la dependencia comercial y económica de esas dos naciones.

   Solución: mediax = 1’40, varx = 0’32, mediay = 1’22, vary = 0’30, covar = 0’31, r = 0’99, hay correlación muy grande, al ser positiva indica que crecen a la vez, las economías son complementarias de intensa relación comercial

2. Si en el problema anterior se obtuviera un coeficiente de correlación igual a -0’61 ¿como se interpretaría?

   Solución: Hay correlación negativa, no muy grande pero sí significativa. Al ser negativa indica que las economías están en competición: cuando una crece la otra decrece

3. Las estaturas y pesos, en centímetros y kilogramos respectivamente, de un grupo de 6 personas están dadas por:

Estatura (cm)	168	174	180	175	158	162
peso (kg)	65	70	73	68	55	62

   i) Hallar la recta de regresión que sirve para predecir la altura conocido el peso y el coeficiente de correlación entre ambas medidas.

   ii) Predecir la etatura de una séptima persona, afín a las anteriores, que pesa 71 kg. ¿Es fiable la predicción?.

   Solución: mediax = 65’50, varx = 34’25, mediay = 169’50, vary = 58’58, covar = 43’32, r = 0’97, $y-169'50=1'27(x-65'50)$, $y(71)=176'3$. Es fiable porque la correlación es alta y el valor 71 está cerca de la media

4. El puesto de clasificación y los goles marcados en una temporada de liga vienen dados por los pares: (1,75),(2,77),(3,72),(4,63),(5,69),(6,75), (7,62),(8,61),(9,63),(10,47),(11,49),(12,43) (13,51),(14,48),(15,44),(16,57),(17,47), (18,51), (19,47),(20,55),(21,37),(22,53) . Hallar la recta de regresión y el coeficiente de correlación interpretando el re-

sultado. ¿Cuántos goles serían necesarios para quedar 8^0?

Solución: coef correl -0,797968258 covar -57,40909091

recta y/x $y = -1,426312818x + 73,03896104$ valor esperado f(8)= 61,6284585

5. Hallar el coeficiente de correlación y si es adecuado, en la recta de regresión y/x, el valor esperado para x = 10, de la variable bidimensional:

x_i	9	11	14	12
y_i	5	9	12	12

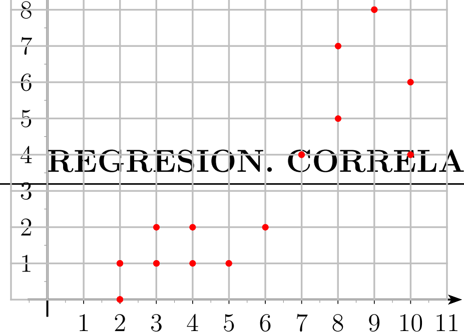
Dibujar la nube de puntos y la recta de regresión.

Solución: mediax = 11'5, mediay = 9'5, covar = 4'75, r = 0'9173, $f(10) = 7'3$

6. Hallar el coeficiente de correlación y si es adecuado, en la recta de regresión y/x, el valor esperado para x = 8, en la variable bidimensional de la que se conoce: $\Sigma x_i = 253, \Sigma y_i = 1171, \Sigma(x_i - \bar{x})^2 = 885,5, \Sigma(y_i - \bar{y})^2 = 2829,09, \Sigma(x_i - \bar{x}) \cdot (y_i - \bar{y}) = -1263, N = 22$

Solución: $r = -0,7979, \quad f(8) = 61,62$

7. Hallar el coeficiente de correlación y si es adecuado, en la recta de regresión y/x, el valor esperado para x = 8, en la variable bidimensional de la que se conoce:



Solución: . coef correl 0'85103036; covar 7,2544 ; recta y/x: $y = 0'8105x - 1,3843$ valor esperado $f(8) = 5,0997$

8. Las notas de Matemáticas y de Física de un grupo de alumnos están dadas por los pares (3,4) (7,6) (5,3) (5,4) (8,7) (7,5) (2,3) (2,2) (8,6). Hallar las rectas de regresión Física/Matemáticas y el coeficiente de correlación entre ambas notas interpretando el resultado.

Solución: mediax = 5'22, varx = 5'28, mediay = 4'44, vary = 2'47, covar = 3'23, recta y/x : $y - 4'44 = 0'61(x - 5'22)$, r = 0'90

9. Dada la variable bidimensional:

x_i	2	4	4	6	8
y_i	3	5	7	5	8
n_i	2	4	6	4	4

Hallar el coeficiente de correlación y el valor esperado para $x = 10$

$r = 0,630, \quad f(10) = 8,65$

$numdat = 20, \Sigma(x_i - \bar{x})(y_i - \bar{y}) = 36, \sigma_{xy} = 1,8, y/x: \quad y = 0,529x + 3,35$

$\Sigma x_i = 100, \bar{x} = 5, \Sigma(x_i - \bar{x})^2 = 68, var_x = 3,4, \sigma_x = 1,844$

$\Sigma y_i = 120, \bar{y} = 6, \Sigma(y_i - \bar{y})^2 = 48, var_y = 2,4, \sigma_y = 1,549$

Tema 3

PROBABILIDAD

3.1. Introducción

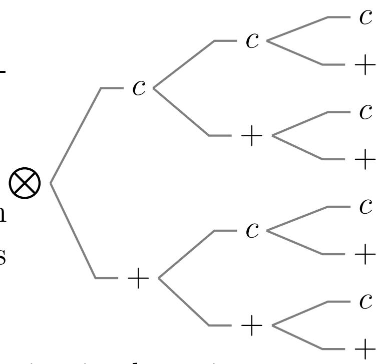
Fenómeno aleatorio es aquel en el cual es imposible predecir el resultado en cada realización u observación; ej: lanzar una moneda, extraer una carta de una baraja, número de nacimientos de una ciudad en un mes, etc.

Cálculo de probabilidades es el modelo teórico de las regularidades que se observan en los resultados de los fenómenos aleatorios cuando crece el número de pruebas.

3.2. Sucesos

El conjunto de todos los resultados asociados a un experimento aleatorio se llama **espacio muestral** y se suele representar por E

Ejemplo Escribir el espacio muestral del lanzamiento de una moneda tres veces a) por extensión, b) mediante diagrama en árbol.
a) $E = \{ccc, cc+, c+c, +cc, c++, +c+, ++c, +++\}$



Suceso es todo subconjunto del espacio muestral. Por ejemplo, en el experimento lanzar un dado $E = \{1, 2, 3, 4, 5, 6\}$, son sucesos "salir par", "salir menos de 3".

Se dice que un suceso se ha verificado cuando al realizar la experiencia aleatoria correspondiente, el resultado es uno de los elementos de ese suceso. Si al tirar el dado sale un 6 se han verificado, entre otros, los sucesos $\{6\}, \{\textit{salir par}\}, \{5, 6\}, E$.

Los sucesos formados por un solo elemento se llaman **sucesos elementales**, por ejemplo $\{6\}$.

El espacio muestral se llama también **suceso seguro**, el suceso $\emptyset$ se llama suceso imposible.

Hemos considerado los sucesos como conjuntos, por tanto hablaremos de:

inclusión $\subset$: $A \subset B$ (se lee A contenido en B), si todos los elementos de A están en B
unión $\cup$: $A \cup B$ se forma juntando los elementos de A y de B

intersección $\cap$: $A \cap B$ está formado por los elementos comunes a los dos
complementario $\bar{A}$: los elementos restantes que no están en A.

Existen también denominaciones propias del lenguaje de sucesos:

$A \subset B$ es $A \Longrightarrow B$ (se lee A implica B), la verificación del suceso A implica la del suceso B; ej $A =$ salir múltiplo de 3, $B =$ salir más de 2.

$A \cup B$ se verifica el suceso A **o** el suceso B, se verifica **al menos** uno de los dos
$A \cap B$ se verifica el suceso A **y** el suceso B
El complementario $\bar{A}$ del suceso A se llama suceso **contrario**.
Dos sucesos disjuntos, sin ningún elemento común: $A \cap B = \emptyset$ se llaman **incompatibles**.

3.3. Frecuencia de un suceso

Prueba es cada realización de un experimento aleatorio. Sea un experimento aleatorio del que se han realizado N pruebas. Si el suceso A aparece n veces se dice que en la referida muestra de N pruebas la frecuencia relativa del suceso A es $fr(A) = \dfrac{n}{N}$.

Observamos que: (podemos pensar en el lanzamiento 20 veces de un dado: A =salir par)
1) La frecuencia relativa de un suceso está comprendida entre 0 y 1.
2) La frecuencia relativa del suceso seguro es 1.
3) La frecuencia relativa de la unión de dos sucesos incompatibles es la suma de las respectivas frecuencias: si$A \cap B = \emptyset$, $\quad fr(A \cup B) = fr(A) + fr(B)$

Por otro lado si por ejemplo se lanza una moneda 50 veces y salen 28 caras, no tiene por qué ocurrir que al repetir las 50 tiradas vuelvan a salir 28 caras, o sea, las frecuencias relativas suelen variar en cada serie de pruebas.

No obstante al aumentar el número de pruebas se tiene el siguiente resultado práctico llamado **ley del azar** : las frecuencias relativas de los sucesos tienden a estabilizarse alrededor de ciertos números, a estos números se les suele llamar probabilidad de los respectivos sucesos.

3.4. Probabilidad

Es el modelo teórico de las frecuencias relativas. Por tanto la probabilidad de un suceso es un número entre 0 y 1 y cumple las condiciones:
1) $p(E) = 1$, la probabilidad del suceso seguro es 1.
2) dados A, B sucesos incompatibles : $p(A \cup B) = p(A) + p(B)$, es decir la probabilidad de la unión de sucesos incompatibles es la suma de las probabilidades.

Probabilidad de **Laplace** es la que asigna a cada suceso elemental la misma probabilidad, por tanto la probabilidad de un suceso elemental es $\dfrac{1}{N}$ siendo N el número de sucesos elementales.

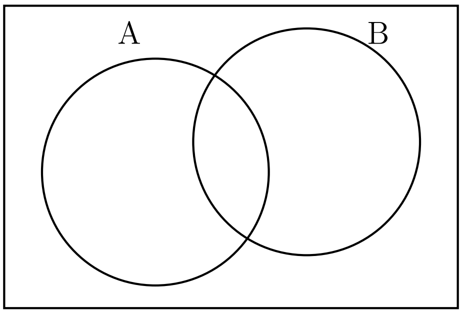
Entonces si el suceso A es la unión de n sucesos elementales tendremos:
$p(A) = \dfrac{n}{N}$ o en otras palabras $p(A) = \dfrac{\text{casos favorables}}{\text{casos posibles}}$

Por ejemplo en la extracción de una carta de una baraja española, la probabilidad de que salga un basto es $p(B) = \dfrac{10}{40}$

Probabilidad **estimada**, empírica o a posteriori de un suceso es la frecuencia relativa de la aparición del suceso cuando el número de observaciones es muy grande.

Por ejemplo a la vista de la producción de un gran número de piezas, una fábrica encuentra que el 20 % de los cerrojos producidos por una determinada máquina son defectuosos para unos ciertos requerimientos. Parece lógico asignar una probabilidad 0'2 de obtener un cerrojo defectuoso.

Propiedades de una probabilidad:
Las demostraciones se deducen de las condiciones de la definición de probabilidad.



1. La probabilidad del suceso imposible es 0:
$$p(\emptyset) = 0,$$
2. Para el suceso complementario se cumple:
$$p(\bar{A}) = 1 - p(A)$$
3. Para la unión de dos sucesos cualesquiera se tiene:
$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

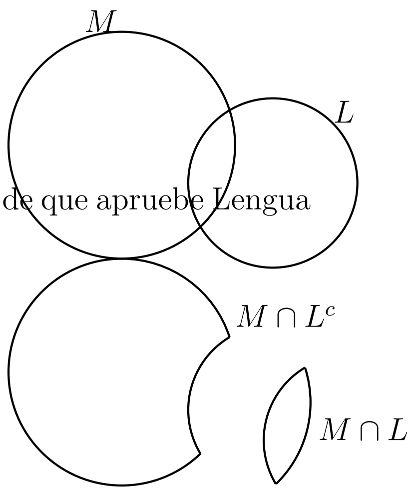
Ejemplos

1. Hallar la probabilidad de que salga bastos o figura al sacar una carta de una baraja española (40 cartas).

   $A =$ salir bastos, $p(A) = \dfrac{10}{40}$

   $B =$ salir figura (sota, caballo, rey), $p(B) = \dfrac{12}{40}$

   $p(A \cup B) = p(A) + p(B) - p(A \cap B) = \dfrac{10}{40} + \dfrac{12}{40} - \dfrac{3}{40} = \dfrac{19}{40}$



2. La probabilidad de que un alumno apruebe Matemáticas es $0'6$ y la de que apruebe Lengua es $0'5$ y la de que apruebe las dos es $0'2$.

   a) Hallar la probabilidad de que apruebe alguna (es decir, al menos una).
   b) Hallar la probabilidad de que no apruebe ninguna.
   c) Hallar la probabilidad de que apruebe Matemáticas y no Lengua.

   a) $p(M \cup L) = p(M) + p(L) - p(M \cap L) = 0'6 + 0'5 - 0'2 = 0'9$

   b) $p[(M \cup L)^c] = 1 - 0'9 = 0'1$

   c) $M = (M \cap L^c) \cup (M \cap L)$ disjunta; $p(M \cap L^c) = p(M) - p(M \cap L) = 0'6 - 0'2 = 0'4$

3. Una urna contiene 25 bolas blancas de madera, 36 blancas de cristal, 39 bolas rojas en total, y 32 de madera en total.

a) Hallar el número total de bolas.

Si se elige al azar una bola:

b) ¿Cuál es la probabilidad de que sea blanca?.

c) ¿Cuál es la probabilidad de que sea roja y de madera?.

d) ¿Cuál es la probabilidad de que sea blanca o de cristal?.

a) Completamos el cuadro:

	rojas	blancas	
madera	7	25	32
cristal	32	36	68
	39	61	100

Consideremos los sucesos B = extraer bola blanca, M = extraer bola de madera, R = extraer bola roja. Entonces:

b) $p(B) = 61/100 = 0'61$

c) $p(R \cap M) = 7/100 = 0'07$

d) $p(B \cup C) = p(B) + p(C) - p(B \cap C) = 0'93$

3.5. Probabilidad con combinatoria

3.5.1. Variaciones con repetición

Por ejemplo consideremos las cuatro letras a, b, c, d Cada grupo con tres de estas letras repetidas o no es una variación con repetición de los cuatro elementos a, b, c, d de orden 3,

$$\text{Ejemplos: } \begin{matrix} aab & ccc \\ abc & bca \\ baa & dad \end{matrix}$$

Se llama **variación con repetición** de m elementos de orden h a cada uno de los grupos de h elementos que se pueden formar con los m elementos pudiendo repetirse un mismo elemento, son distintas dos variaciones con repetición si difieren en algún elemento o en el orden de colocación.

Ejemplo: Una quiniela de fútbol es una variación con repetición de orden 14 de los elementos 1 X 2.

El **número de variaciones** con repetición distintas de m elementos de orden h es $RV_m^h = m^h$

Así para esos 4 elementos el número de variaciones con repetición de orden 3 distintas viene dado por $RV_4^3 = 4^3 = 64$

3.5.2. Variaciones

Consideremos las 4 letras a, b, c, d. Cada grupo de 3 de estas letras sin repetir es una variación de orden 3 de esos 4 elementos.

	abc	*dac*
Ejemplos:	*cba*	*bda*
	acd	*dab*

Se llama **variación** de m elementos de orden h a cada uno de los grupos de h elementos que se pueden formar con los m elementos sin repetirse un mismo elemento, son distintas dos variaciones si difieren en algún elemento o en el orden de colocación.

El **número de variaciones** distintas de m elementos de orden h es $V_m^h = m^{(h} = m(m-1)(m-2)(m-3)\dots(m-h+1)$, es decir h factores consecutivos decrecientes a partir de m.

Es útil la fórmula: $V_m^h = \dfrac{m!}{(m-h)!}$

Para esos 4 elementos el número de variaciones de orden 3 distintas es $V_4^3 = 4^{(3}$

3.5.3. Permutaciones

Son las variaciones cuando el orden es el número total de elementos o sea, cuando en cada grupo entran todos los elementos.

	abcd	*dacb*
Ejemplos:	*cbad*	*bdac*
	acdb	*dabc*

Dos permutaciones se distinguen por ser distinto el orden de los elementos.

Se llama **permutación** de m elementos a cada ordenación de los m elementos, (son variaciones de m elementos de orden m)

El **número de permutaciones** distintas de m elementos es $P_m = m! = m(m-1)(m-2)(m-3)\dots 3\cdot 2\cdot 1$

3.5.4. Combinaciones

Consideremos las 4 letras a, b, c, d. Cada subconjunto de 3 de estas letras (por tanto el orden no importa) es una combinación de orden 3 de esos 4 elementos

Ejemplos:	*abc*	*dab*	*cdb* sería la misma combinación que la última
	acd	*dcb*	

Se llama **combinación** de m elementos de orden h a cada uno de los grupos de h elementos que se pueden formar con los m elementos sin repetirse un mismo elemento, son distintas dos combinaciones solo si difieren en algún elemento, (son pues los posibles subconjuntos del conjunto formado por los m elementos).

Los boletos de la loto son combinaciones de orden 6.

El **número de combinaciones** distintas de m elementos de orden h es: $C_m^h = \dbinom{m}{h} = \dfrac{m^{(h}}{h!}$

Para esos 4 elementos el número de combinaciones de orden 3 distintas es $C_4^3 = \binom{4}{3} = \frac{4^{(3}}{3!} = 4$

Problemas en los que se utiliza combinatoria para contar los casos Se utiliza cuando el diagrama en árbol resultaría muy grande:

1. Se tiene una urna con 9 bolas numeradas del 1 al 9. Cual es la probabilidad de que al extraer tres bolas sucesivamente las tres lleven número par, a) si no se reemplaza la bola tras cada extracción; b) si se reemplaza la bola tras cada extracción.

   Llamemos A el suceso salir par en las tres extracciones:

   a) Los resultados son del tipo 143, 987, variaciones de 9 elementos de orden 3.

   los casos posibles son $V_9^3 = 9^{(3} = 9 \cdot 8 \cdot 7 = 504$

   hay 4 bolas con número par, los casos favorables son $V_4^3 = 4^{(3} = 24$,

   luego $p(A) = \frac{24}{504} = \frac{1}{21}$.

   Puesto que según el enunciado el orden no parece influir también se puede considerar que se trata de combinaciones, la probabilidad que resulta es la misma,

   los casos posibles son $C_9^3 = \binom{9}{3} = \frac{9^{(3}}{3!} = 84$

   hay 4 pares, los casos favorables son $C_4^3 = \binom{4}{3} = \frac{4^{(3}}{3!} = 4$

   luego $p(A) = \frac{4}{84} = \frac{1}{21}$.

   b) Los resultados ahora pueden ser 143, 144, 298, variaciones con repetición de 9 elementos de orden 3.

   los casos posibles son: $RV_9^3 = 9^3 = 729$

   hay 4 pares, los casos favorables son $RV_4^3 = 4^3 = 64$

   luego $p(A) = \frac{64}{729}$.

2. Se elige al azar un número de 8 cifras, ¿Cuál es la probabilidad de que el número elegido presente únicamente cuatro dígitos distintos?.

   Casos Posibles: 10 números y usamos 8 con repetición: $RV_{10}^8 = 10^8$

   Casos Favorables: 10 números y usamos 4 sin repetición: $V_{10}^4 = 5040$

   Probabilidad de que tenga 4 dígitos distintos: $5'04 \cdot 10^{-5}$

3. Dados diez puntos del plano tales que no hay 3 alineados, se nombra a cuatro de ellos con las letras A,B,C,D. De todos los triángulos que se pueden dibujar con ese conjunto de

puntos se elige uno. ¿Cuál es la probabilidad de que el triángulo elegido tenga rotulado todos sus vértices con letras?

Posibles triángulos con 10 puntos: $C_{10}^3 = 120$ Posibles triángulos cuyos vértices estén marcados con letras: $C_4^3 = 4$

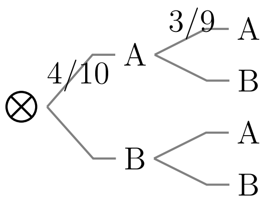
Probabilidad de que los triángulos estén rotulados: 1/30

3.6. Sucesos dependientes e independientes

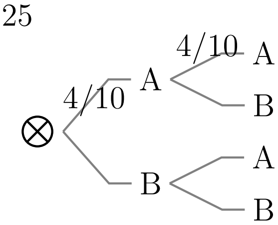
Ejemplo Una caja contiene 10 piezas, de las cuales 4 son defectuosas.
I) Hallar la probabilidad de extraer dos defectuosas consecutivas
a) sin devolver la primera.
b) devolviendo la primera.
II) Sin devolver la primera, hallar la probabilidad de obtener una de cada tipo.

A = extraer pieza defectuosa ; B = extraer pieza no defectuosa
I) Para hallar la probabilidad de una rama se multiplican las probabilidades de la rama:

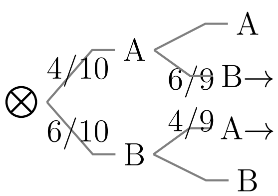
a) Sin devolución, sucesos dependientes:
$$p(A_1 \cap A_2) = p(A_1).p(A_2/A_1) = \frac{4}{10}.\frac{3}{9} = \frac{2}{15}$$



b) Con devolución, sucesos independientes:
$$p(A_1 \cap A_2) = p(A_1).p(A_2) = \frac{4}{10}.\frac{4}{10} = \frac{4}{25}$$



II) Como es la unión de varias ramas, se suman las probabilidades de las ramas favorables:



$$p[(A_1 \cap B_2) \cup (B_1 \cap A_2)] = p(A_1 \cap B_2) + p(B_1 \cap A_2) = \frac{4}{10}.\frac{6}{9} + \frac{6}{10}.\frac{4}{9} = \frac{24}{45}$$

Dos sucesos A y B son **independientes** si la realización de uno no varía la probabilidad de la realización del otro;

Si se lanza una moneda y un dado, el salir cara en la moneda es independiente de que salga par en el dado. Si lanzo una moneda la primera vez la probabilidad de salir cara es 1/2, si la lanzo la segunda vez la probabilidad de cara sigue siendo 1/2. En cambio si extraigo una carta de una baraja la probabilidad de salir espada la primera vez es 10/40, si no devuelvo la carta, evidentemente la probabilidad de salir espada en la segunda no es 10/40, pues ha cambiado la composición de la baraja.

Para sucesos independientes la probabilidad de la intersección es el producto de las probabilidades: $p(A \cap B) = p(A).p(B)$

Dados dos sucesos A, B , se llama suceso B **condicionado al** A y se representa B/A, al suceso realizarse el suceso B supuesto realizado el suceso A".

Para sucesos dependientes la probabilidad de la intersección es el producto de la probabilidad del primero por la probabilidad del segundo condicionado al primero: $p(A\cap B) = p(A).p(B/A)$

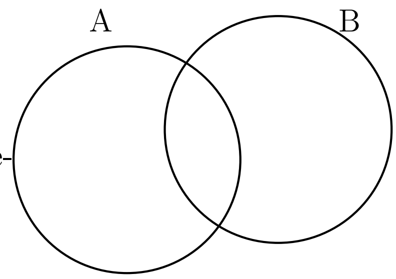
Ejemplos

1. Para no confundir la velocidad con el tocino se estudió una muestra de 100 casos y se obtuvieron estos datos:

	Tocino T	No tocino
Velocidad V	32	48
No velocidad	8	12

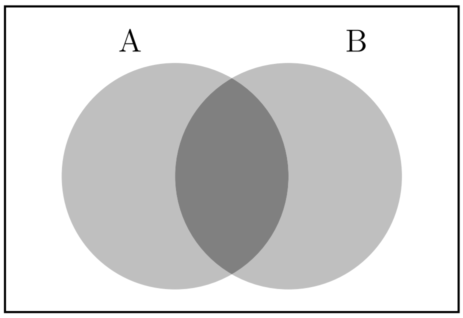
   Según estos datos, ¿son independientes los sucesos T y V?
   $p(V).p(T) = \dfrac{80}{100}.\dfrac{40}{100} = 0'32$
   $p(V\cap T) = \dfrac{32}{100} = 0'32$
   efectivamente la velocidad y el tocino , V y T son independientes.

2. Sean A y B dos sucesos independientes de un espacio de probabilidades. Sean $0'3$ y $0'6$ sus probabilidades respectivas. Hallar las probabilidades de cada uno de los sucesos siguientes:
   S_1 acontece exactamente uno de los sucesos A o B, uno de los dos pero no los dos.
   S_2 acontecen los dos A y B.

   $p(S_1) = p(A\cup B - A\cap B) = p(A) + p(B) - 2p(A\cap B)$
   necesitamos $p(A\cap B)$ que es el 2^0 apartado, como son independientes:
   $p(A\cap B) = p(A).p(B) = 0'3{,}0'6 = 0'18 = p(S_2)$
   luego $p(S_1) = 0'3 + 0'6 - 2{,}0'18 = 0'54$

3. Sean A y B dos sucesos, tales que $P(A) = \frac{3}{4}$, $\quad P(B) = \frac{1}{2}$, $\quad P(\bar{A}\cap\bar{B}) = \frac{1}{20}$. Calcular:
   a) $P(A\cup B)$
   b) $P(A\cap B)$
   c) $P(\bar{A}/B)$
   Nota: $\bar{A}$ representa el suceso complementario de A.

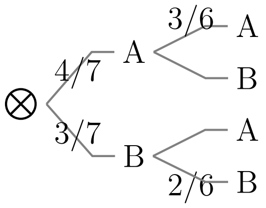
   

   a) Como vemos en el dibujo $A\cup B$ es lo contrario de $\bar{A}\cap\bar{B}$ por tanto $P(A\cup B) = 1 - p(\bar{A}\cap\bar{B}) = 1 - \dfrac{1}{20} = \dfrac{19}{20}$
   b) Partiendo de la probabilidad de la unión:
   $P(A\cup B) = P(A) + p(B) - P(A\cap B)$,
   sustituyendo: $\dfrac{19}{20} = \dfrac{3}{4} + \dfrac{1}{2} - P(A\cap B)$ y despejando queda: $P(A\cap B) = \dfrac{19}{20} - \dfrac{3}{4} - \dfrac{1}{2} = \dfrac{3}{10}$

   c) $P(\bar{A}/B) = \dfrac{P(\bar{A}\cap B)}{p(B)} = \dfrac{P(B) - p(A\cap B)}{p(B)} = \dfrac{\frac{1}{2} - \frac{3}{10}}{\frac{1}{2}} = \dfrac{\frac{2}{10}}{\frac{1}{2}} = \dfrac{2}{5}$

4. En una urna hay bolas: 4 azules y 3 blancas. Se extraen dos bolas simultáneamente. Hallar la probabilidad de que sean las dos blancas sabiendo que han salido de igual color.

   Llamamos "cc" a igual color, piden $p(BB/cc)$

   $$p(BB) = \frac{3}{7}.\frac{2}{6} = \frac{1}{7}$$

   $$p(cc) = \frac{4}{7}.\frac{3}{6} + \frac{3}{7}.\frac{2}{6} = \frac{3}{7}$$

   Para la intersección tenemos que $BB \subset cc$ luego $p(BB \cap cc) = p(BB)$:

   Despejando en la expresión: $p(BB \cap cc) = p(BB/cc) \cdot p(cc)$

   $$p(BB/cc) = \frac{p(BB \cap cc)}{p(cc)} = \frac{\frac{1}{7}}{\frac{3}{7}} = \frac{1}{3}$$

Observaciones:

1. Resumiendo:

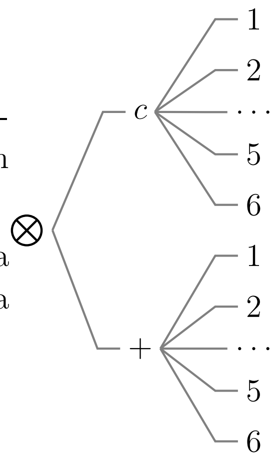
   independientes $p(A \cap B) = p(A).p(B)$ **dependientes** $p(A \cap B) = p(B/A).p(A)$

2. No confundir sucesos incompatibles (la probabilidad de la unión es la suma de las probabilidades), con sucesos independientes (la probabilidad de la intersección es el producto de las probabilidades).Por eso:

   Dos sucesos compatibles pueden ser dependientes o independientes. Dos sucesos incompatibles necesariamente son dependientes.

3. En la extracción de, por ejemplo, dos bolas de una urna es lo mismo: extracción simultánea de las dos, que extracciones sucesivas sin devolución.

4. Experimentos independientes simultáneos es situación análoga a extracción sucesiva con devolución, esto permite utilizar diagrama en árbol. Por ejemplo se lanza un dado y una moneda.



Ejercicio Una urna A contiene 3 bolas blancas y una negra y otra urna B contiene 5 bolas negras y 7 blancas. 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.

a) ¿Cuál es la probabilidad de que esa bola sea negra?

b) Si la bola extraída ha sido negra, cuál es la probabilidad de que las dos bolas pasadas de A a B fueran blancas.

3.7. Problemas

1. Escribir el espacio muestral correspondiente al lanzamiento de un dado dos veces. a) Mediante diagrama en árbol. b) Por extensión.

2. Escribir el espacio muestral correspondiente a la suma de puntos en el lanzamiento de un dado dos veces. ¿Tiene la misma probabilidad el 8 que el 3?.

   Solución: $p(tres) = 2/36$, $p(ocho) = 5/36$

3. Tres cajas tienen las siguientes composiciones: A = 5 bolas blancas y 2 negras, B = 7 bolas blancas y 1 negra y C = 2 bolas blancas y 8 negras. Se escoge al azar una caja y se extraen dos bolas sin reemplazamiento. Escribir el espacio muestral.

4. Se tiran un dado y una moneda. Hallar la probabilidad de obtener cruz y número primo.

   Solución: 0'3333

5. En una urna hay 3 bolas blancas, 4 negras, 5 rojas y 6 azules. Hallar: a) Probabilidad de que al sacar una bola sea azul. b) Probabilidad de que al sacar dos bolas sean blancas. c) Probabilidad de que al sacar dos bolas sean, la primera negra y la segunda roja.

   Solución: a) 0'3333 b) 0'0196 c) 0'0653

6. Hallar la probabilidad de que al sacar dos cartas de una baraja española: a) sean 2 oros, sin devolver la primera carta. b) sean 2 figuras, devolviendo la primera carta.

   Solución: a) 0'0576 b) 0'09

7. En una clase mixta hay 30 alumnas; 15 estudiantes repiten curso de los que 10 son alumnos y hay 15 alumnos que no repiten curso. a) Justificar que el número de estudiantes de esa clase es 55. b) Si se elige al azar un estudiante de esa clase: b_1) ¿Cuál es la probabilidad de sea alumno?. b_2) ¿Cuál es la probabilidad de que repita curso y sea alumna?. c) Si se eligen dos estudiantes al azar ¿cuál es la probabilidad de que ninguno repita curso?.

   Solución: a) 55 estudiantes, b_1 25/55, b_2 5/55, c)52/99

8. La caja C_1 contiene 5 fichas azules y 3 rojas, la caja C_2 contiene 4 fichas azules y 6 rojas. Se traslada una ficha de la caja C_1 a la caja C_2; a continuación se extrae una ficha de C_2. ¿Cuál es la probabilidad de que la ficha extraída sea roja?.

   Solución: p(roja extracción 2^a caja) = 51/88

9. Hallar 5 resultados posibles al tirar un dado 7 veces. ¿Cuántos resultados hay en total?. ¿Cual es la probabilidad de que todos los números que salgan sean primos?

   Solución: $279936, p = 0'058$

10. Poner 4 ejemplos de casos posibles en una mano de mus: 4 cartas. ¿Cuántos resultados hay en total?. ¿Cual es la probabilidad de que todos sean bastos?

    Solución: $91390, p = 0,0023$

11. Hallar el número de productos diferentes que se pueden formar tomando tres cifras de 1,2,3,5,7 sin que haya factores repetidos.

    Solución: 10

12. En un campeonato de dardos participan 6 países, cuántas quinielas hay que hacer para acertar con seguridad los tres primeros.

    Solución: 120

13. De una baraja de 48 cartas se extraen 10 al azar. Calcular la probabilidad de que 6 de ellas sean copas.

    Solución: $\frac{C_{12}^{6} \cdot C_{36}^{4}}{C_{48}^{10}}$

14. Se lanzan simultáneamente tres monedas al aire. ¿Cuál es la probabilidad de que todas queden en el suelo del mismo modo?.

    Solución: $p(c) + p(+) = 1/4$

15. Se extraen 3 cartas de una baraja española (40 cartas). Hallar la probabilidad de que sean 3 bastos; a) sin reemplazamiento; b) con reemplazamiento.

    Solución: a) $P[(1B) \cap (2B) \cap (3B)] = 10/40{,}9/39{,}8/38 = 0'012$, b) $P[(1B) \cap (2B) \cap (3B)] = 10/40{,}10/40{,}10/40 = 0'015$

16. De una baraja de 40 cartas se toman dos. Hallar la probabilidad: a) De que las dos sean oros. b) De que las dos sean espadas o figuras. c) Al menos una sea sea bastos.

    Solución: a) $p(OO) = 10/40{,}9/39 = 0'0576$, b) X salir espadas o figura $p(XX) = 19/40{,}18/39 = 0'21$, c) árbol $p(almenosunbasto) = 1 - \frac{30}{40}.\frac{29}{39} = 0'442$

17. Se lanzan 6 monedas simultáneamente. Calcular la probabilidad de que al menos salga una cara.

    Solución: 63/64

18. Tres cajas tienen las siguientes composiciones: A = 5 bolas blancas y 2 negras, B = 7 bolas blancas y 1 negra y C = 2 bolas blancas y 8 negras. Se escoge al azar una caja y se extraen dos bolas sin reemplazamiento. Calcula la probabilidad de que las bolas sean del mismo color.

    Solución: $1/3(11/21 + 3/4 + 29/45)$

19. Sean A y B dos sucesos, tales que $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{5}$, $P(\bar{A} \cup \bar{B}) = \frac{3}{4}$

    1. $P(B/A)$
    2. $P(\bar{A}/B)$

    Nota: $\bar{A}$ representa el suceso complementario de A.

    Solución: a) 1/2, b) 3/8

20. Sobre los sucesos A y B se conocen las siguientes probabilidades:

    $P(A) = 0'7$; $P(B) = 0'5$; $P(A \cap B) = 0'45$

    Calcular:

    1. $P(B/A)$
    2. $P(A^c \cap B^c)$

    Nota: A^c representa el suceso complementario de A.

    Solución: a) $0'6428$, b) $0'25$

21. Se lanza un dado y, a continuación, una moneda. ¿Cuál es la probabilidad de obtener:

    i) Cuatro y cara.

    ii) Cruz e impar.

    iii) Cara o un número mayor que 1.

    Solución: i) 1/12, ii) 3/12, iii) 11/12

22. Sean A y B dos sucesos de un experimento aleatorio, tales que $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$, $P(\bar{A} \cup \bar{B}) = \frac{11}{12}$

    1. ¿Son A y B dos sucesos independientes? Razónese.
    2. $P(\bar{A}/\bar{B})$

    Nota: $\bar{A}$ representa el suceso complementario de A.

    Solución: a) son independientes, b) 3/4

23. Se tienen dos urnas A y B, en la primera hay 6 bolas negras y 4 rojas; en la segunda hay 3 bolas negras, 2 rojas y 5 blancas. Se lanza un dado y si sale múltiplo de 3 se extrae una bola de la urna A

y en caso contrario de la B. ¿Cuál es la probabilidad de que al extraer una bola sea roja?.

Solución: 4/15

24. En una clase, el 40 % aprueban Filosofía y el 50 % Matemáticas. Además, la probabilidad de aprobar la Filosofía habiendo aprobado las Matemáticas es 0'8. Prueba que la mitad de la clase suspende ambas asignaturas y calcula el porcentaje de alumnos que teniendo aprobada la Filosofía aprueban también las Matemáticas.

    Solución: a) 0'5 b) el 100 %

25. De una baraja española de 40 cartas se extraen 4 sucesivamente sin reemplazamiento. Calcular la probabilidad de que sean del mismo palo.

    Solución: 4(10/40)(9/39)(8/38)(7/37) = 0'009

26. En un cierto edificio se usan dos ascensores; el primero lo usan el 45 % de los vecinos y el resto usan el segundo. El porcentaje de fallos del primero es del 5 % mientras que el del segundo es del 8 %. Si en un cierto día un inquilino queda "atrapado" en un ascensor, hallar la probabilidad de que haya sido en el primero.

    Solución: $\frac{225}{225+440} = 0'34$

27. Dos personas A y B organizan el siguiente juego: Tiran un dado tres veces. Si sale algún 1, gana A. Si no sale ningún 1, gana B. ¿Cuál de las dos personas tiene más probabilidades de ganar?

    Solución: $p(B) = (\frac{5}{6})^3 = 0'5787 > 0'5$ gana B

28. El 45 % de los habitantes de una determinada ciudad son del Barça y los demás son del Madrid. Un 27 % de los del Barça y el 38 % de los del Madrid además juegan al fútbol. Calcular la probabilidad de que al elegir un habitante: a) Juegue al fútbol b) Sea del Barça sabiendo que no juega al fútbol.

    Solución: a) 0'33, b) 0'4906

29. Ana, Pedro y Juan se reparten los problemas que tienen que resolver. Se quedan respectivamente con el 23 %, 44 %, y 33 %. Sabemos que Ana resuelve correctamente el 60 % de los problemas que intenta, Pedro el 20 % y Juan el 40 %. a) Hallar la probabilidad de que al elegir un problema al azar esté mal hecho. b) Hallar la probabilidad de que al elegir un problema al azar y que resulta que está mal resuelto sea de los hechos por Juan.

    Solución: a) 0'642, b) 0'308

30. Los datos de votantes en unas elecciones muestran que votó el 73'5 % de los hombres censados y que no votó el 42'9 % de las mujeres. El censo era de 48 % hombres y el 52 % mujeres.

    De entre todas las personas censadas, escogemos una al azar. Calcular la probabilidad de que esta persona: a) Haya votado. b) Haya votado y sea hombre. c) Sabiendo que ha votado, sea mujer.

    Solución: a) 0'649, b) 0'352, c) 0'457

31. Dos profesores comparten un número de teléfono. De las llamadas que llegan, 2/5 son para A y 3/5 son para B. Sus ocupaciones les alejan de este teléfono, de modo que A está fuera el 50 % del tiempo y B el 25 %. Calcular la probabilidad de que no esté ninguno para responder al teléfono. Llaman por teléfono y no lo cogen, cuál es la probabilidad de que llamen a A.

    Solución: a) 0'35, b) 0'57

32. El despertador de Pepe no suena el 20 % de las veces. Cuando no suena el despertador llega tarde a clase el 84 % de los

días, en cambio cuando suena llega tarde solo el 12 %. Hoy Pepe ha llegado puntual, cuál es la probabilidad de que haya sonado el despertador.

Solución: 0'956

33. La fabricación de cierto tipo de objetos se hace en dos fases, la probabilidad de que resulte defectuoso en la primera fase es del 4 % mientras que en la segunda es del 1 %. ¿Cuál es la probabilidad de que un objeto tomado al azar no tenga defectos?

    Solución: por árbol en dos fases $p(nodef) = 0'96{,}0'99 = 0'9504$

34. Tenemos tres bolsas iguales, la A con 13 bolas negras y 15 blancas, la B con 16 bolas negras y 12 blancas y la C con 7 bolas negras y 13 blancas

    a) Se coge una bola de una bolsa al azar y resulta negra, ¿cuál es la probabilidad de que provenga de la bolsa A.

    b) Hallar la probabilidad de que la bola extraída sea blanca.

    Solución: a) Bayes (vuelta atrás de árbol) $p(A/n) = \dfrac{0'1518}{0'4554} = 0'33$

    b)árbol normal $p(b) = 1 - 0'4554 = 0'53$

35. El test para detectar una sustancia contaminante en agua, presenta los siguientes resultados: si el agua no está contaminada, suceso que ocurre con una probabilidad igual a 0,99, el resultado del test es que el agua está contaminada con una probabilidad igual a 0,05. Cuando el agua está contaminada, el test lo detecta con una probabilidad igual a 0,99. Se ha realizado una prueba y el test indica que hay contaminación. Calcular la probabilidad de que el agua no esté realmente contaminada. Interpretar el valor numérico obtenido.

    El test detecta que el agua está contaminada, cuando en realidad no lo está el 83,33 % de las veces. Se trata de un mal producto.

36. En química clínica son particularmente interesantes los llamados coeficientes falso-positivo y falso-negativo de un test. Tales coeficientes son probabilidades condicionadas. El coeficiente falso-positivo α es la probabilidad de que el contraste resulte positivo cuando de hecho el sujeto no padece la dolencia. El coeficiente falso-negativo β se define de manera análoga. Cada una de estas probabilidades es una probabilidad de error; por tanto, cabe esperar que los valores obtenidos en la práctica sean próximos a cero.

    Los resultados siguientes se obtuvieron en un estudio diseñado con el fin de averiguar la capacidad de un cirujano patólogo para clasificar correctamente las biopsias quirúrgicas como malignas o benignas (T^+ = diagnóstico es positivo; R^+ = la biopsia es en realidad maligna)

	T^+	T^-
R^+	79	19
R^-	7	395

Determinar α y β a partir de estos datos.

$\alpha = p(T^+/R^-) = 0{,}017$; $\beta = p(R^-/T^+) = 0{,}194$.

	T^+	T^-
R^+		falso-negativo $\beta = p(R^-/T^+)$
R^-	$\alpha = p(T^+/R^-)$ falso-positivo	

37. En una clase hay 40 estudiantes de los que 10 son chicos.

a) En la elección de delegado y subdelegado, ¿cuantas posibilidades distintas hay?. ¿Cuál es la probabilidad de que sean los dos chicos?

b) Se hacen comités de dos estudiantes, ¿cuantas posibilidades distintas hay?. ¿Cuál es la probabilidad de que sean los dos chicos?

c) Los 40 estudiantes echan una carrera, puntúan los 9 primeros, ¿cuantas posibilidades distintas hay?. ¿Cuál es la probabilidad de que sean todos chicos?

d) Se hacen comités de nueve estudiantes, ¿cuantas posibilidades distintas hay?. ¿Cuál es la probabilidad de que sean los dos chicos?

38. Con las cifras que son número primo 1,2,3,5,7

a) ¿Cuantos números de 8 cifras se pueden formar?. ¿Cuál es la probabilidad de que resulte par?

b) ¿Cuantos números de 3 cifras distintas se pueden formar?. ¿Cuál es la probabilidad de que resulte par?

c) ¿Cuantos números de 5 cifras distintas se pueden formar?. ¿Cuál es la probabilidad de que resulte par?

d) ¿Cuantos productos distintos se pueden formar con 3 de esas cifras sin que se repitan los factores? ¿Cuál es la probabilidad de que resulte par?

39. Con las cifras 0,1,2,3,4,5,6,7,8,9

a) ¿Cuantos números de 4 cifras, pudiendo repetirse, se pueden formar?. ¿Cuál es la probabilidad de que resulte par?

b) ¿Cuantos números de 4 cifras distintas se pueden formar?. ¿Cuál es la probabilidad de que resulte par?

a) el primer dígito no puede ser 0, hay 9 para elegir; los siguientes se eligen entre 10 $9 \cdot 10^3 = 9000$

b) $9 \cdot V_9^3 = 4536$

40. Un alumno tiene que elegir 7 de las 10 preguntas de un examen. ¿De cuántas maneras puede elegirlas? ¿Y si las 4 primeras son obligatorias?

El orden en que elija las preguntas, que además no podrán repetirse, es irrelevante. Así, puede elegir las preguntas de $C_{10}^7 = 120$ maneras.

Por otra parte, si las 4 primeras son obligatorias, debe escoger 3 preguntas entre las 6 restantes para completar las 7 necesarias, resultando un total de $C_6^3 == 20$ maneras.

41. En la síntesis de proteínas hay una secuencia de tres nucleótidos sobre el ADN que decide cuál es el aminoácido a incorporar. Existen cuatro tipos distintos de nucleótidos según la base, que puede ser A (adenina), G (guanina), C (citosina) y T (timina). ¿Cuántas secuencias distintas se podrán formar si se pueden repetir nucleótidos?

Ya que importa el orden de los nucleótidos en la secuencia, y además éstos pueden repetirse, entonces existen $RV_4^3 = 64$ secuencias distintas.

42. Una mano de póker consiste en cinco cartas seleccionadas sin reemplazamiento de una baraja de 52 (sin comodines). Determinar la probabilidad de obtener las siguientes combinaciones:

1. Escalera de color: las cinco cartas consecutivas y del mismo palo.

2. Escalera de color real: escalera de color con el As como carta mayor, detrás de la K.

3. Póker: cuatro cartas con la misma numeración.

4. Póker de ases.

5. Full: tres cartas con una numeración y las otras dos con otra.

6. Escalera: las cinco cartas consecutivas (el As puede ir al comienzo o al final).

7. Color: las cinco cartas del mismo palo.

8. Dobles parejas.

9. Trío.

10. Pareja.

Para introducir un espacio muestral denotemos cada carta mediante un par (n, e), donde n representa el número en la carta (es decir, $n \in \{1, 2, \ldots 13\}$ y e representa el palo (es decir, $e \in \{A, B, C, D\}$. Entonces el espacio muestral es:

$\Omega = \{w = \{w_1, w_2, w_3, w_4\} : \forall i \neq j; w_i = (n, e), n \in \{1, 2, \ldots 13\}, e \in \{A, B, C, D\}; w_i \neq w_j\}$

Claramente este espacio es equiprobable y hay C_{52}^5 resultados posibles.

1. Definamos el suceso $A =$ "Se obtiene una escalera de color". Cada palo de la baraja tiene $52/4 = 13$ cartas, con las que se pueden formar $13 - 5 + 1 = 9$ escaleras de color. Por tanto, ya que hay cuatro palos distintos, se tiene que:

A comprende $4 \cdot 9 = 36$ resultados favorables.

2. Sea el suceso $B =$ "Se obtiene una escalera de color real". Por cada palo de la baraja sólo hay una escalera de color real posible. Por tanto:

B comprende 4 resultados favorables.

3. Sea C el suceso "Se obtiene un póker". Hay 13 numeraciones diferentes. Una vez escogidas 4 cartas con la misma numeración se elige entre las $52 - 4 = 48$ restantes la que falta para completar la mano, obteniéndose que

C comprende $13 \cdot 48$ resultados favorables.

4. Definamos el suceso $D =$ "Se obtiene un póker de ases". Hay $52 - 4 = 48$ cartas posibles para añadir a los 4 ases y completar la mano, por lo que

D comprende 48 resultados favorables.

5. Sea el suceso $E =$ "Se obtiene un full". Fijada una numeración, pueden formarse $C_4^3 = 4$ conjuntos de tres cartas, ya que hay 4 palos distintos. Por lo tanto, como hay 13 posibles numeraciones distintas, en total se tienen $13 \cdot 4 = 52$ posibilidades para escoger las tres cartas iguales del full.

Para las dos cartas restantes hay que tener en cuenta que no pueden ser de la misma numeración anterior, luego, procediendo análogamente al caso anterior, hay en total$12C_4^2 = 72$ combinaciones posibles. Finalmente, se calcula:

E comprende $52 \cdot 72$ resultados favorables.

6. Sea el suceso $F =$ "Se obtiene una escalera". Hay 13-5+1 $= 9$ numeraciones posibles de las escaleras, a las que hay que añadir una más que corresponde a la escalera con el As al final. Si fijamos una numeración $i, i+1, i+2, i+3, i+4; con i = 1, \ldots 9, 10$, tendremos, para cada valor de i, 4^5 escaleras (incluyendo las de color, y las de color real si $i = 10$). Si eliminamos las 4 escaleras de color correspondientes a esa numeración (una por cada palo), quedan $4^5 - 4$ escaleras y, dado que hay 10 numeraciones posibles. Entonces:

F comprende $(4^5 - 4) \cdot 10$ resultados favorables.

7. Representemos por G al suceso "Se obtiene color". Para cada palo, hay C_{13}^5 combinaciones posibles de 5 cartas. De ellas, como vimos en los apartados (a) y (b), $9 + 1 = 10$ corresponden a escaleras de color y a escaleras de color reales. Por lo tanto se eliminan, resultando:

G comprende $4 \cdot (C_{13}^5 - 10)$ resultados favorables.

8. Definamos el suceso $H =$ "Se obtienen dobles parejas". Hay C_{13}^2 formas distintas de elegir los palos con los que se forman las dos parejas, C_4^2 de crear la pareja para cada uno de esos palos. Para la quinta carta quedan $52 - 4 - 2 - 2 = 44$ posibilidades, puesto que se restan, además de las cuatro cartas ya escogidas, las cuatro cartas que quedan con la misma numeración que cada una de las parejas. De este modo se evita obtener un full.

H comprende $C_{13}^2 \cdot C_4^2 \cdot C_4^2 \cdot 44$ resultados favorables.

9. Denotemos por I al suceso "Se obtiene un trío". Hay $C_{13}^1 \cdot C_4^3$ combinaciones posibles de tres cartas con la misma numeración. Para las dos que completan la mano se debe tener en cuenta que ninguna de ellas puede tener la misma numeración que las tres cartas anteriores, ya que se obtendría un póker, y además ambas no pueden ser de la misma numeración, pues se formaría un full. Luego, una vez fijadas las 3

primeras, se escoge la cuarta carta de un conjunto de $52-4=48$ cartas (se descartan las 4 cartas que hay con la numeración de las tres ya elegidas), y para la última quedan, finalmente, $48-4=44$ posibilidades (se descartan las de la misma numeración que la cuarta carta). Además, como no se tiene en cuenta el orden en que se elijan estas dos últimas cartas, dividimos $48\cdot 44$ por $2!$ y resulta:

I comprende $C_{13}^1\cdot C_4^3\cdot\dfrac{48\cdot 44}{2!}$ resultados favorables.

10. Sea J el suceso "Se obtiene una pareja". Las dos cartas que forman la pareja pueden escogerse de un total de $13\cdot C_4^2$ parejas posibles. Para las tres que faltan deben descartarse aquellas combinaciones que, junto a las dos primeras cartas, formarían un trío, un póker o un full. Por lo tanto, y procediendo de forma similar al caso del trío, fijadas las dos primeras hay $52-4=48$ posibilidades para la tercera carta, $48-4=44$ para la cuarta y $44-4=40$ para la última. Análogamente al apartado anterior, se dividen las $48\cdot 44\cdot 40$ combinaciones de las tres últimas cartas por $3!=6$, ya que no importa el orden en que éstas se elijan. Así:

J comprende $C_{13}^1\cdot C_4^2\cdot\dfrac{48\cdot 44\cdot 40}{6!}$ resultados favorables.

43. Un examen de oposición consta de 14 temas. Se debe escoger un tema de entre dos tomados al azar. Calcular la probabilidad de que a un alumno que ha preparado 5 temas le toque al menos uno que sabe. ¿Cuál es el número mínimo de temas que debe preparar para que tenga una probabilidad superior a $1/2$ de superar el examen?

Definimos el suceso A = "Le toca al menos un tema que ha preparado". Entonces: $p(A)=1-p(A^c)=1-\dfrac{\binom{14-5}{2}}{\binom{14}{2}}=\dfrac{55}{91}$ que es la probabilidad que se pide calcular.

Finalmente, supongamos que i = "número de temas preparados por el alumno". Para superar el examen le debe tocar al menos un tema que haya preparado. Por lo tanto, la probabilidad de aprobar el examen sería

$$p(A)=1-\frac{\binom{14-i}{2}}{\binom{14}{2}}>\frac{1}{2}$$

y resolviéndola se concluye que el alumno debe preparar como mínimo 4 temas.

44. Problema de Buffon. Se tiene una mesa rayada con líneas paralelas separadas una distancia $2b$. Se lanza una aguja de longitud $2a$ para que caiga sobre la mesa. Hallar la probabilidad de que la aguja corte a alguna línea si $a\le b$.

Un suceso elemental de este problema puede describirse mediante un par de números, $w=(w_1;w_2)$, donde el primero w_1 representa la distancia del centro de la aguja (tras caer sobre la mesa) a la línea más próxima, y el segundo w_2 representa el ángulo que la inclinación de la aguja tiene respecto de las líneas en la mesa. Nótese que por tanto $w_1\in[0,b)$ y que $w_2\in[0,\pi)$. En efecto, dado el objetivo que se persigue en el problema, es suficiente mirar sólo la posición respecto a la línea en la mesa más próxima y no interesa distinguir entre los dos extremos de la aguja (da igual la punta que el ojo de la aguja).

El conjunto de todos los sucesos elementales anteriores configura un espacio muestral $\Omega=[0,b[\times[0,\pi[$ claramente equiprobable. Por ello, la probabilidad de cualquier suceso $A\subseteq\Omega$ se obtiene dividiendo la integral sobre A entre $b\pi$. En concreto, si A es el suceso que representa "la aguja toca una línea" y, asumiendo que $a\le b$, entonces

$A=\{w\in\Omega: w_1\in[0,a\cos(w_2)[$ para cada $w_2\in[0,\pi[\}$

Consecuentemente

$$P(A)=\frac{1}{b\pi}\int_0^{\pi}\left(\int_0^{a\cos y}\partial x\right)\partial y=\frac{2a}{b\pi}$$

45. Se consideran dos números aleatorios elegidos uniformemente y con independencia dentro del intervalo $[0,1]$.

1. Calcular la probabilidad de que su diferencia sea mayor que $1/6$.

2. Calcular la probabilidad de que su suma sea mayor que 1.

Un espacio muestral viene dado por $\Omega=\{w=(w_1,w_2):0\le w_1\le 1;0\le w_2\le 1\}$

Según la hipótesis del enunciado, se trata de un espacio equiprobable y por tanto se aplica la Ley de Laplace. Consecuentemente, el cálculo de la probabilidad de un suceso se limita al cálculo del área del correspondiente trozo en Ω, ya que el área del total es la unidad.

1. Sea A el suceso "su diferencia es mayor que 1/6", es decir:

$A = \{w \in \Omega : w_1 - w_2 > 1/6 o w_2 - w_1 > 1/6\}$

Gráficamente es fácil ver que el área de A es la de un cuadrado con lado 5/6, es decir, $P(A) = (5/6)^2 = 0'69444$.

2. Sea B el suceso "su suma es mayor que 1", es decir:

$B = \{w \in \Omega : w_1 + w_2 > 1\}$

Gráficamente se deduce que el área coincide con la de medio cuadrado, es decir: $P(B) = \dfrac{1}{2}$

46. Problema de encuentro. Se conoce que en un intervalo de tiempo de 30 minutos llegan a un mismo punto de encuentro y de forma aleatoria dos personas. ¿Qué probabilidad existe de que una de las personas espere por la otra al menos 10 minutos?

Este clásico problema en cualquier asignatura relacionada con las Probabilidades es una mera reformulación del problema anterior.

En efecto, un espacio muestral equiprobable es $\Omega = \{w = (w_1, w_2) : 0 \leq w_1 \leq 30; 0 \leq w_2 \leq 30\}$

Si A representa los sucesos en los que "el encuentro sucede después de 10 minutos de espera por parte de algunas de las personas", es decir:

$A = \{w \in \Omega : w_1 - w_2 > 10 o w_2 - w_1 > 10\}$

y de forma análoga a como se hizo en el problema anterior: $P(A) = \dfrac{5}{9}$

Tema 4

VARIABLES ALEATORIAS. DISTRIBUCIÓN DE PROBABILIDAD

4.1. Variable aleatoria. Función de distribución de probabilidad

Es el modelo matemático de la variable estadística. Se dice que hemos definido una variable aleatoria X (v.a.) para un experimento aleatorio cuando hemos asociado un valor numérico a cada resultado del experimento.

Ejercicio Imagínese un juego de apuestas con estas normas: Se lanza un dado normal y se cobra 3 euros si sale 1 o 2, 1 euro si sale 4, 5 o 6 y se pagan 5 euros si sale un 3. Se lanza el dado 60 veces y se obtienen los siguientes resultados:

3, 4, 6, 1, 3, 1, 1, 5, 6, 6, 1, 1, 6, 1, 5, 6, 2, 2, 3, 2, 6, 4, 6, 2, 5, 6, 1, 1, 3, 2, 4, 5, 5, 3, 2, 5, 6, 5, 3, 5, 2, 6, 1, 4, 6, 1, 5, 5, 5, 5, 2, 4, 3, 3, 1, 4, 5, 2, 2, 6

Se considera la variable estadística que dé las ganancias y pérdidas:

1) Hacer la tabla de frecuencias absolutas y relativas.

2) Dibujar el diagrama de frecuencias relativas y el polígono de frecuencias relativas.

número	1	2	3	4	5	6
frecuencia	11	10	8	6	13	12

número	var. estad. X	frecuencia n_i	frec. relativa f_i
{ 3 }	-5	8	0'13
{ 4,5,6 }	1	31	0'51
{ 1,2 }	3	21	0'35
		$\Sigma N_i = 60$	

Ejemplo 1) Considérese el juego anterior: Se lanza un dado normal y se cobra 3 euros si sale 1 o 2, 1 euros si sale 4, 5 o 6 y se pagan 5 euros si sale un 3. La v.a. que describe las posibles ganancias en este juego es $X(1) = 3, X(2) = 3, X(3) = -5, X(4) = 1, X(5) = 1, X(6) = 1$.

4.2. Tabla de probabilidades de una variable aleatoria discreta. Histograma de Probabilidad

A cada valor que toma la variable le asociamos la probabilidad del suceso que representa así obtenemos la tabla de probabilidades de una variable aleatoria discreta:

–5–4–3–2–1 | 1 2 3

x_i	-5	1	3
p_i	$1/6$	$3/6$	$2/6$

Tomando intervalos de longitud uno con centro en los valores de la v.a. x_i tenemos el **histograma de probabilidad** de la v.a. X.

En el histograma de probabilidad la suma de las áreas de los rectángulos hasta un valor x_i (incluido el suyo) da la probabilidad $p(X \leq x_i)$.

Función de distribución de la v.a. X es la función que a cada número le asigna la probabilidad acumulada hasta ese número, se suele expresar: $F(x) = p(X \leq x)$

En el ejemplo: $F(2'5) = p(X \leq 2'5) = p(X = -5) + p(X = 1) = \frac{1}{6} + \frac{3}{6}$

4.3. Relación entre variables estadísticas y aleatorias

Para muestras grandes las frecuencias relativas tienden a las correspondientes probabilidades, lo cual nos permite considerar a las funciones de probabilidad como el modelo teórico de las frecuencias relativas, que son las que se pueden obtener en la práctica. Es lo que llamábamos probabilidad empírica.

Así por ejemplo en el problema que veremos más adelante:

”En la fabricación de automóviles de una determinada marca de cada 1.000 fabricados 10 resultan defectuosos por término medio. ¿Cuál es la probabilidad de que en un lote de cuatro automóviles más de la mitad sean defectuosos?”

Se toma como probabilidad de que un automóvil resulte defectuoso $p = 10/1000 = 0'01$.

4.4. Parámetros de una variable aleatoria discreta

Se corresponden con los de una variable estadística, por ejemplo la media de una variable estadística es: media $\bar{x} = \dfrac{\Sigma x_i n_i}{\Sigma n_i} = \Sigma x_i f_i$

y la desviación típica: $(\text{des. tip.})^2 = \dfrac{\Sigma x_i^2 . n_i}{\Sigma n_i} - \bar{x}^2 = \Sigma x_i^2 . f_i - \bar{x}^2$

Para una variable aleatoria discreta:

Esperanza matemática o media: $\mu = \Sigma x_i p_i$

Varianza: $\sigma^2 = \Sigma(x_i - \mu)^2 p_i = \Sigma x_i^2 p_i - \mu^2$

Desviación típica: $\sigma = \sqrt{\text{varianza}}$

Intuitivamente, si la variable aleatoria describe las ganancias y pérdidas de un determinado juego, la esperanza indica la ganancia media por partida que puede esperar un jugador. Si la esperanza es cero se dice que el juego es equitativo; en caso contrario, es favorable o desfavorable al jugador según que la esperanza sea positiva o negativa.

La desviación típica determina, junto con la esperanza, el intervalo $[\mu - \sigma, \mu + \sigma]$ en el que se espera se produzcan "la mayoría de los resultados".

En el ejemplo resultaría:

$$E(X) = \frac{1}{6}(-5) + \frac{3}{6}1 + \frac{2}{6}3 = \frac{4}{6} = 0'666$$

$$\sigma^2 = \frac{1}{6}(-5)^2 + \frac{3}{6}1^2 + \frac{2}{6}3^2 - \left(\frac{4}{6}\right)^2 = \frac{260}{36} = 7'222; \quad \sigma = \sqrt{7'222} = 2'68$$

4.5. Distribución binomial

Ejemplo En una bolsa hay 2 bolas blancas y 3 negras. Hacemos extracciones con devolución.

VARIABLE ESTADÍSTICA

Se hacen 10 series de 3 extracciones con devolución de una bolsa con 2 bolas blancas y 3 negras. Consideramos el número de bolas blancas que salen en cada serie.

Supongamos que el número de bolas blancas en cada serie ha sido respectivamente: 1, 1, 1, 3, 1, 2, 2, 0, 2, 0.

a) Hallar las frecuencias relativas y hacer un diagrama de barras de ancho uno.

b) Hallar la media y la desviación típica.

VARIABLE ALEATORIA

Se hacen 3 extracciones con devolución de una bolsa con 2 bolas blancas y 3 negras. Consideramos el número de bolas blancas que pueden salir.

a) Hacer el diagrama en árbol de la experiencia aleatoria.

b) Hacer la tabla de probabilidades y el histograma de probabilidad.

c) Calcular la media y la desviación típica

Solución:

VARIABLE ESTADÍSTICA

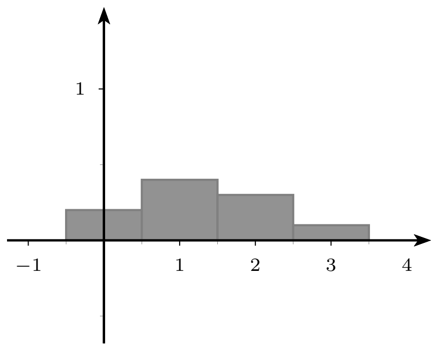
a)

frecuencias absolutas

x_i	0	1	2	3
n_i	2	4	3	1

frecuencias relativas

x_i	0	1	2	3
f_i	0'2	0'4	0'3	0'1



b)

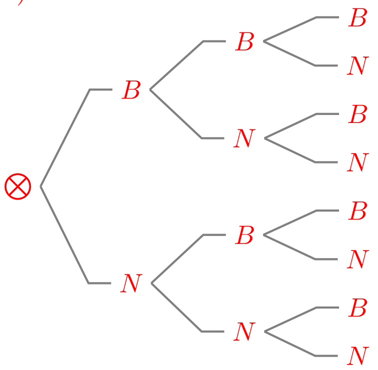
x_i	f_i	$x_i.f_i$	x_i^2	$x_i^2.f_i$
0	0'2	0	0	0
1	0'4	0'4	1	0'4
2	0'3	0'6	4	1'2
3	0'1	0'3	9	0'9
		$\Sigma x_i f_i = 1'3$		$\Sigma x_i^2.f_i = 2'5$

Media: $\bar{x} = \Sigma x_i f_i = 1'3$

Desviación típica: $\sigma = \sqrt{\Sigma x_i^2 f_i - \bar{x}^2} = \sqrt{2'5 - 1'3^2} = \sqrt{0'81} = 0'9$

VARIABLE ALEATORIA

a)



b)

$p(3 \text{ blancas}) = \left(\frac{2}{5}\right)^3 = 0'064$

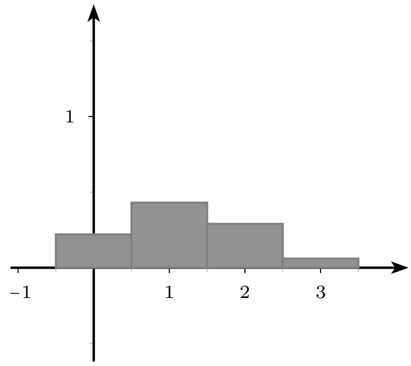
$p(2 \text{ blancas}) = 3 \cdot \left(\frac{2}{5}\right)^2 \cdot \frac{3}{5} = 0'288$

$p(1 \text{ blancas}) = 3 \cdot \frac{2}{5} \cdot \left(\frac{3}{5}\right)^2 = 0'432$

$p(0 \text{ blancas}) = \left(\frac{3}{5}\right)^3 = 0'216$

probabilidad

x_i	0	1	2	3
p_i	0'216	0'432	0'288	0'064



c)

x_i	p_i	$x_i.p_i$	x_i^2	$x_i^2.p_i$
0	0'216	0	0	0
1	0'432	0'432	1	0'432
2	0'288	0'576	4	1'152
3	0'064	0'192	9	0'576
		$\Sigma x_i p_i = 1'2$		$\Sigma x_i^2.p_i = 2'16$

Media: $\mu = \Sigma x_i p_i = 1'2$

Desviación típica: $\sigma = \sqrt{\Sigma x_i^2 p_i - \mu^2} = \sqrt{2'16 - 1'2^2} = \sqrt{0'72} = 0'84852$

Ejemplo En el lanzamiento de un dado se considera éxito obtener 5 o más puntos y fracaso lo contrario, por tanto probabilidad de éxito: $p = \frac{2}{6} = \frac{1}{3}$, probabilidad de fracaso: $q = \frac{2}{3}$. Supongamos que se hacen 10 pruebas.

Se trata de la distribución binomial $B(10, \frac{1}{3})$, consideremos la variable aleatoria: $X =$ número de éxitos en las 10 pruebas

Hallemos la probabilidad de tener 4 éxitos (y por tanto 6 fracasos), o sea de $X = 4$:

La probabilidad de tener 4 éxitos y 6 fracasos en un orden determinado, como los lanzamientos son independientes, es: $p.p.p.p.q.q.q.q.q.q = p^4.q^6$; como el orden no nos importa el suceso tener cuatro éxitos es la unión de los sucesos del tipo anterior, hay $\binom{10}{4}$ de estos sucesos (que son las posibilidades de "escoger" las cuatro tiradas con éxito entre las 10) por tanto la probabilidad buscada de $X = 4$, es sumar $\binom{10}{4}$ veces la cantidad $p^4.q^6$:

$$p(X = 4) = \binom{10}{4} p^4 q^6 = 210. \left(\frac{1}{3}\right)^4 . \left(\frac{2}{3}\right)^6 = 0'7868$$

En general:

Distribución binomial $B(n, p)$: Es la distribución de probabilidad de una variable aleatoria discreta que tiene las características:

1) En el experimento aleatorio hay dos resultados posibles "éxito" de probabilidad p y su contrario "fracaso" de probabilidad $q = 1 - p$. Las probabilidades no cambian en las sucesivas pruebas.

2) La variable aleatoria discreta es: X = número de éxitos en n pruebas,

Entonces la probabilidad viene dada por:

$$\boxed{p(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \text{ con } q = 1 - p}$$

Los parámetros de la binomial son: $\mu = n.p, \quad \sigma^2 = n.p.q$

Ejemplo En la fabricación de automóviles de una determinada marca de cada 1.000 fabricados 10 resultan defectuosos por término medio. Cuál es la probabilidad de que en un lote de seis automóviles

a) Haya 2 defectuosos.
b) Haya tres o menos defectuosos.
c) Hallar la media y la desviación típica

Sea $p = 0'01$ la probabilidad de ser defectuoso; $B(6, 0'01)$

a) $p(X = 2) = \binom{6}{2} p^2 q^4 = 15 \cdot 0'01^2 \cdot 0'99^4 = 0'0014$

b) $p(X \leq 3) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) =$

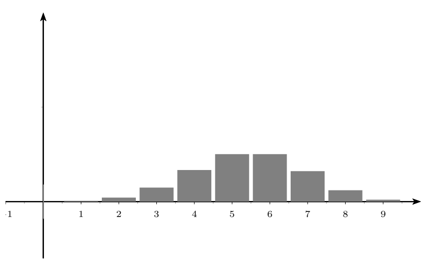
c) $\mu = 6 \cdot 0'01 = 0'06, \quad \sigma^2 = 6 \cdot 0'01 \cdot 0'99 = 0'0594 \quad \sigma = \sqrt{0'0594}$

Ejercicios

1. En una bolsa hay 12 bolas blancas y 8 azules. Se hacen 9 extracciones **con reemplazamiento** y se considera el número de bolas blancas que pueden salir. a) Hacer la tabla de probabilidad. b) Hacer el histograma de probabilidad. c) Hallar la media y la desviación típica.

   Es $B(9; 0,6)$.

x_i	p_i
0	0
1	0,004
2	0,021
3	0,074
4	0,167
5	0,251
6	0,251
7	0,161
8	0,06
9	0,01

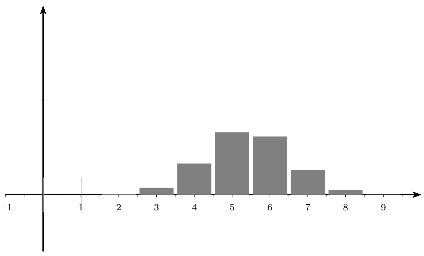


   $\mu = 5,4, \quad \sigma^2 = 2,16, \quad \sigma = 1,470$

2. En una bolsa hay 12 bolas blancas y 8 azules. Se hacen 9 extracciones **sin reemplazamiento** y se considera el número de bolas blancas que pueden salir. a) Hacer la tabla de probabilidad. b) Hacer el histograma de probabilidad. c) Hallar la media y la desviación típica. (En este caso se dice que la variable sigue una distribución de probabilidad HIPERGEOMETRICA)

x_i	p_i
0	0
1	0
2	0,003
3	0,037
4	0,165
5	0,33
6	0,308
7	0,132
8	0,024
9	0,001

$\mu = 5,4, \quad \sigma^2 = 1,248, \quad \sigma = 1,117$



4.6. Variable aleatoria continua

Hasta ahora hemos visto casos en los que la variable aleatoria toma unos valores concretos. En estos casos se llama **variable aleatoria discreta**.

Pero hay otra posibilidad:

Ejemplo Lugar de rotura de una cuerda de 3 m al tirar de un extremo estando el otro fijo. El espacio muestral es $E =$ conjunto de lugares de rotura $= [0, 3]$. Consideramos la variable aleatoria:

$X =$longitud del punto de corte al punto fijo.

Vemos que la variable aleatoria puede tomar cualquier valor del intervalo $[0, 3]$. En este caso se llama **variable aleatoria continua**

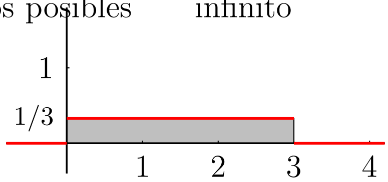
4.7. Función de densidad de probabilidad de una v.a. continua

Ejemplo Lugar de rotura de una cuerda de 3 m al tirar de un extremo estando el otro extremo fijo.

$X =$longitud del punto de rotura al extremo fijo, puede tomar cualquier valor entre 0 y 3.

Consideremos: probabilidad $= \dfrac{\text{casos favorables}}{\text{casos posibles}}$; la probabilidad de que se rompa en un punto determinado, $X = x_0$, es cero pues en este caso $\dfrac{\text{casos favorables}}{\text{casos posibles}} = \dfrac{1}{\text{infinito}} = 0$. Por ello:

Lo que podemos considerar es la probabilidad de que la v.a. tome un valor menor o igual que uno dado, por ejemplo que se rompa antes de 2'5 metros.



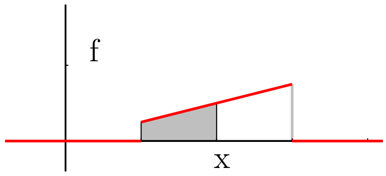
$p(X \leq 2'5) = \dfrac{\text{longitud favorable}}{\text{longitud posible}} = \dfrac{2'5}{3}$

Para una v.a. continua no tiene sentido hablar de probabilidad de que la variable tome un determinado valor porque habría que dividir por "infinitos" casos posibles

Entonces como modelo teórico del polígono de frecuencias relativas, se introduce el concepto de **función de densidad de probabilidad** f:

La **función de densidad de probabilidad** $f(x)$ indica la cantidad de probabilidad en esa zona:

La probabilidad viene dada por $p(X \leq x) =$ **área bajo la función densidad entre el inicio de la gráfica y el valor** x**.**

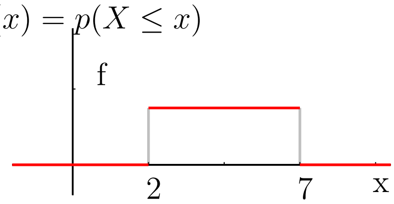


Por tanto se cumple que una función de densidad siempre es positiva y además el área bajo la función densidad vale 1.

Función de distribución de la v.a. X es la función que a cada número le asigna la probabilidad acumulada hasta ese número, se suele expresar: $F(x) = p(X \leq x)$

Ejercicio Dada la función de densidad de gráfica.
a) Hallar su expresión analítica.

b) Hallar $p(-1 \leq x < 3'5)$



4.7.1. Parámetros de una variable aleatoria continua:

Si tenemos una variable aleatoria continua X con función de densidad f:

Función de distribución $F(x) = p(X \leq x) = \displaystyle\int_{-\infty}^{x} f(t)dt$, o sea la función de distribución es el área bajo la curva $f(t)$ entre el inicio de la gráfica y el valor x.

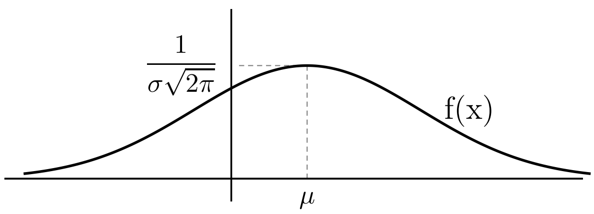
Media: $\mu = \displaystyle\int_{-\infty}^{\infty} x.f(x)dx$

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

Desviación típica: $\sigma = \sqrt{\textbf{varianza}}$

4.8. Distribución normal

La variable aleatoria continua más utilizada es la normal su función de densidad de probabilidad tiene de gráfica:

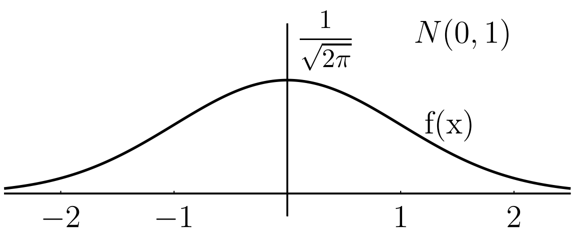


Se suele expresar $N(\mu, \sigma)$; los parámetros μ y σ son respectivamente el valor medio y la desviación típica

La curva se llama campana de Gauss.

La normal $N(0,1)$ tiene de función densidad:

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

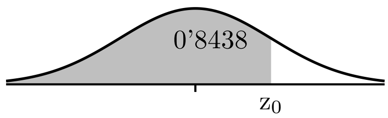


cuyos parámetros son $\mu = 0, \sigma = 1$, y tiene las probabilidades acumuladas por $f(x)$ tabuladas.

Cálculo de probabilidades en la normal Las instrucciones de la hoja de cálculo dan $p(X \leq x)$, para buscar otras probabilidades hay que utilizar la simetría y el complementario.

=DISTR.NORM.ESTAND(x): probabilidad acumulada, $p(Z \leq z)$ en $N(0,1)$
=DISTR.NORM.ESTAND.INV(p): inversa: dada la probabilidad hallar z en $N(0,1)$
=DISTR.NORM(x;μ;σ;VERDADERO) probabilidad acumulada
=DISTR.NORM(x;μ;σ;FALSO) valor de la función de densidad

Ejercicios: Hallar: a) $p(Z \leq 0'34) =$
b) $p(Z < -2'85) =$ c) $p(Z \geq 2'1) =$



Proceso inverso: dada la probabilidad, hallar el valor de la variable aleatoria

Ejercicio: En $N(0,1)$ hallar z_0 tal que $p(Z \leq z_0) = 0'8438$, resulta:

Ejercicio: Hallar en $N(8,3)$ el valor de $p(X \leq 9'6)$

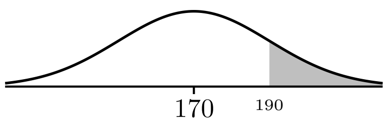
Tipificación: Para relacionar las probabilidades de una normal cualquiera $N(\mu, \sigma)$ con la normal $N(0,1)$ se hace el cambio de variable (se llama tipificar) $z = \dfrac{x-\mu}{\sigma}$ que la transforma en la normal $N(0,1)$.

Ejemplos

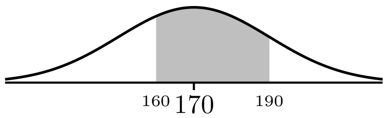
1. Se eligió una muestra de 1000 personas de una determinada población y resultó que su talla media era de 170 cm, con una desviación típica de 10 cm. Suponiendo que las tallas se distribuyen normalmente, calcúlese cuantas personas de esa muestra miden: a) Más de 190 cm; b) Entre 160 y 190 cm.

   La v.a. X que describe las tallas de la población es del tipo $N(170,10)$.

   a)
   $p(X > 190) = 1 - 0'9772 = 0'0228$
   Es de esperar que haya $0'0228. \cdot 1000 = 22'8 \approx 23$ personas de más de 190 cms.

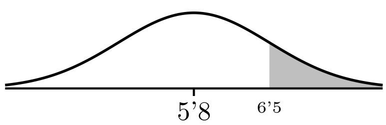
   

   b)
   $p(160 < X < 190) = 0'9772 - 0'1587 = 0'8185$
   O sea 818 personas aproximadamente medirán entre 160 y 190 cm.

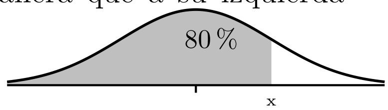
2. En una prueba de selectividad se ha obtenido de nota media 5’8 y la desviación típica es 1’75. Suponemos que las notas están distribuidas normalmente. Todos los alumnos que sobrepasen la nota 6’5 serán admitidos en la universidad. ¿Qué porcentaje de admitidos cabe esperar?

   $p(X \geq 6'5) = 1 - 0'6554 = 0'3446$
   Este valor es el tanto por uno, el tanto por ciento será 34’46 % de admitidos.

Ejercicio: Proceso inverso

3. En una normal $N(23, 12)$, hallar el valor de la variable de manera que a su izquierda esté el 80 % de la probabilidad.

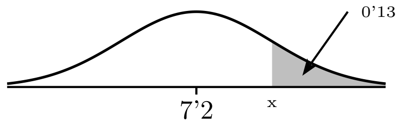
   

   Al contrario que antes buscamos un x concreto tal que $p(X \leq x) = 0'8$

   En la $N(0, 1)$ tenemos que si $p(Z \leq z) = 0'8$, el valor que corresponde es $z = 0'84$.

   sustituyendo en la tipificación: $z = \frac{x-\mu}{\sigma}$, $\quad x = \sigma z + \mu = 12z + 23 = 12{,}0'84 + 23 = 33'08$

4. En una oposición la puntuación media del último examen fue 7'2 y la desviación típica 0'9. Hay plazas para un 13 % de los presentados. ¿Cuál es la puntuación mínima que un estudiante debe tener para conseguir plaza en la oposición?.

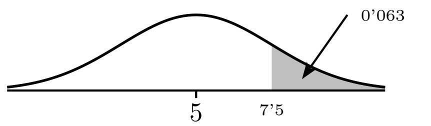
   

   Buscamos un x concreto tal que $p(X \geq x) = 0'13$

   Sabemos que $p(X \geq x) = 0'13$, en la $N(0, 1)$ para buscar en la tabla tenemos: $p(Z \geq z) = 0'13$, corresponde con $p(Z \leq z) = 0'87$ que corresponde con $z = 1'13$.

   sustituyendo en la tipificación: $z = \frac{x-\mu}{\sigma}$, $\quad x = \sigma z + \mu = 0'9z + 7'2 = 0'9{,}1'13 + 7'2 = 8'21$

5. Las puntuaciones de un examen calificado entre 0 y 10 puntos siguen una distribución normal de media $\mu = 5$. El 6'3 por ciento de los alumnos tiene una puntuación por encima de 7'5, ¿qué tanto por ciento de los alumnos es de esperar que tengan una puntuación por debajo de 4 puntos?

   Primero hemos de hallar σ :
   $p(X \geq 7'5) = 0'063$
   $p(Z \geq z) = 0'063 \longrightarrow p(Z \leq z) = 0'937$
   se obtiene $z = 1'53$
   el cambio $z = \dfrac{x - \mu}{\sigma}$, $\quad 1'53 = \dfrac{7'5 - 5}{\sigma}$, $\quad$ despejando $\sigma = 1'63$
   Piden $p(X \leq 4) = 0'2709$, luego aproximadamente el 27'1 % de los alumnos sacará menos de 4.

4.8.1. Aproximación normal de la distribución binomial

La aproximación normal de la distribución binomial es válida cuando $n.p > 5$ y $n.q > 5$.

Ejemplo En un proceso de control de calidad se sabe que el 3 % de los artículos son defectuosos. Si estos se colocan en cajas de 300, se pide:

a) Probabilidad de que una caja contenga 10 o más artículos defectuosos.

b) Probabilidad de que el número de defectuosos esté comprendido entre 15 y 20 ambos inclusive.

c) Si se rechazan todas las cajas con más de 10 defectuosos y se examinan 125 cajas, ¿cuántas de ellas se rechazarán?

Solución:

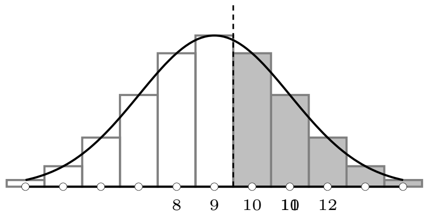
La variable X que es $B(300, 0'03)$, podemos aproximarla por la variable X' normal:

$\mu = n.p = 9, \quad \sigma = \sqrt{n.p.q} = 2'95, \qquad N(9, 2'95)$

a) nota [a] $p(X \geq 10) = p(X' \geq 9'5) = 0'4325$

b) $p(15 \leq X \leq 20) = p(14'5 \leq X' \leq 20'5) = 0'0314$

c) Puesto que la probabilidad de más de 10 defectuosas en cada caja es $0'4325$, en 125 cajas habrá que rechazar $125{,}0'4325 = 54$ cajas.



---

[a] para mayor precisión como la binomial toma valores enteros en la normal se toma valor intermedio para repartir con el complementario

4.9. Problemas

1. Se tiene un dado correcto, pero de tal manera que tres caras tienen el número 2, dos caras el número 1 y una cara el número 3. Se considera la variable aleatoria X que asigna a cada resultado del dado el número obtenido.

   a) Hacer una tabla con las probabilidades.

   b) Representar el histograma de probabilidad.

   c) Hallar la media y la desviación típica.

   Solución:

x_i	1	2	3
p_i	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

   $\mu = 11/6, \sigma = \sqrt{17}/6$

2. En una caja donde hay dos bolas blancas y tres negras se efectúa el siguiente experimento: se sacan dos bolas consecutivas sin reponer. Una bola blanca vale un punto y una negra, dos puntos. A cada extracción se asigna la suma de los puntos obtenidos obteniéndose así la variable aleatoria X.

   a) Espacio muestral.

   b) Hacer una tabla con las probabilidades.

   c) Representar el histograma de probabilidad.

   d) Hallar la media y la desviación típica.

   e) El mismo ejercicio reponiendo la bola cada vez.

   Solución: a) $E = \{bb, bn, nb, nn\}$ b) $R = \{2, 3, 4\}$,

   c)

x_i	2	3	4
p_i	$\frac{2}{20}$	$\frac{12}{20}$	$\frac{6}{20}$

   $\mu = 16/5, \sigma = 3/5$

3. Un tirador olímpico da en el blanco una media de 3 veces cada 5 disparos. Una competición es a tres disparos. Hallar la tabla de distribución aleatoria que considera el número de blancos. Representar la función de probabilidad. Hallar la probabilidad de hacer algún blanco. Hallar la media y la desviación típica.

   Solución:

x_i	0	1	2	3
p_i	0'064	0'288	0'432	0'216

   $\mu = 1'8, \sigma^2 = 0'72, p(\text{algún blanco}) = 0'936$

4. En la fabricación de automóviles de una determinada marca de cada 1.000 fabricados 10 resultan defectuosos por término medio. Se consideran lotes de 4 automóviles. Hallar la tabla de la distribución aleatoria que considera el número de defectuosos en un lote. Representar la función de probabilidad. ¿Cuál es la probabilidad de que en un lote de cuatro automóviles más de la mitad sean defectuosos?. Hallar la media y la desviación típica.

   Solución:

x_i	0	1	2	3	4
p_i	0'96	0'038	0'0005	0'000004	$\left(\frac{1}{100}\right)^4$

   $\mu = 0'03912, \sigma = 0'19$

5. La probabilidad de que un hombre al disparar pegue en el blanco es 1/3. Hallar y representar la función de probabilidad de la variable aleatoria "número de blancos en cinco disparos".

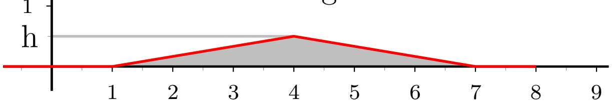
   Solución: $B(5, \frac{1}{3})$

x_i	0	1	2	3	4	5
p_i	0'12	0'31	0'31	0'15	0'03	0'0039

6. En una prueba de selectividad se suspende al 15 % de los estudiantes. a) Hallar el número esperado (o media) de los alumnos suspendidos y la desviación típica si, entre los estudiantes presentados se eligen 2.000. b) Hallar la probabilidad de que suspendan de un grupo de 6 alumnos: I) como máximo 2; II) por lo menos la mitad.

Solución: a) Bin(2000,0’15) prob susp 0’15, media np = 300, destip $\sqrt{npq} = \sqrt{255}$ b) B(6,0’15) I) $p(X \leq 2) = p(X = 0) + p(X = 1) + p(X = 2) = \sum_{x=0}^{2} \binom{6}{x} 0'15^x , 0'85^{6-x} = 0'9526$, $p(X \geq 3) = 1 - p(X \leq 2) = 1 - 0'9526 = 0'04735$

7. Dada la función de gráfica:



a) Hallar h para que cumpla las condiciones de función de densidad de probabilidad.

b) Hallar las siguientes probabilidades:

$P(X \leq 2)$

$P(X \leq 6)$

$P(X \geq 10)$

$P(3 \leq X \leq 6)$

8. Calcular las siguientes probabilidades en la normal $N(0,1)$ a) $p(z \leq 2'78)$; b) $p(z \leq -0'94)$; c) $p(z \leq -1'7)$; d) $p(-1'24 \leq z \leq 2'16)$

Solución: a) 0’9973, b) 0’1736, c) 0’0446, d) 0’8771

9. Calcular las siguientes probabilidades en la normal $N(3,5)$ a) $p(x \leq 4'3)$; b) $p(x < -1)$; c) $p(2 \leq x \leq 10)$

Solución: a) 0’6026, b) 0’2119, c) 0’9192-0’4207=0’4985

10. Se supone que la estancia de los enfermos en un hospital sigue una distribución normal de media 8 días y desviación típica 3. Calcular la probabilidad de que la estancia de un enfermo, a) sea inferior a 7 días; b) sea superior a 3 días; c) esté comprendida entre 10 y 12 días.

Solución: a) 0’3708, b) 0’9515, c) 0’1628

11. Se llama cociente intelectual al cociente entre la edad mental y la edad real. Se sabe que la distribución de los cocientes intelectuales de 2.000 reclutas sigue una distribución normal de media 0’80 y desviación típica 0’50. a) Número de reclutas con cociente intelectual comprendido entre 0’7 y 1’2. b) Id. inferior a 0’3. c) Id. inferior a 0’9. d) Id. superior a 1’4.

Solución: a) 0’3674.2000 ≈ 735, b) 0’1587.2000 ≈ 318, c) ≈ 1159, d) ≈ 230

12. La media de las calificaciones obtenidas en las pruebas de acceso a la Universidad en cierta convocatoria fue $\mu = 4'7$ con una desviación típica $\sigma = 1'3$. Suponiendo que las calificaciones siguen una distribución normal, calcular: i) El porcentaje de aprobados. ii) El porcentaje de alumnos que obtuvo entre 4 y 6 puntos. iii) El porcentaje de alumnos que obtuvo menos de 3 puntos iv) El porcentaje de alumnos que obtuvo más de ocho puntos.

Solución: N(4’4,1’3) i) $p(X \geq 5) = 40'9\,\%$ ii) $p(4 \leq X \leq 6) = 54'32\,\%$ iii) $p(X \leq 3) = 9'68\,\%$ iv) $p(X \geq 8 = 0'57\,\%$

13. Las estaturas de 500 reclutas están distribuidas normalmente con una media de 169 cms y una desviación típica de 7 cms. Calcular el número de reclutas cuya altura, i) está entre 165 y 175 cms ii) es mayor de 180 cms.

Solución: N(169,7) i) $p(X \leq 175) = 0'823$, $p(X \leq 165 = 0'2843$, $p(165 \leq x \leq 175) = 0'518$ ii) $p(X > 180) = 0'0582$

14. Un profesor realiza un test de cien items a un curso con doscientos cincuenta alumnos. Suponiendo que las puntuaciones obtenidas por los alumnos siguen una distribución normal de media 64 puntos y desviación típica 10 puntos y denotando con $p(X \leq n)$ la probabilidad de obtener n puntos como máximo y con $p(X \geq n)$ la probabilidad de obtener al

menos n puntos. Calcular: i) $p(X \geq 60)$, $p(X \leq 75)$, $p(30 \leq X \leq 60)$ ii) Número de alumnos que se espera que tengan al menos 45 puntos.

Solución: i) $p(X \geq 60) = 65'5\,\%$, $p(X \leq 75) = 86'43\,\%$, $p(30 \leq X \leq 60) = 34'43\,\%$ $\quad ii) 0'9713{,}250 \approx 243$ alumnos

15. En una carrera la media del tiempo empleado ha sido de 73 minutos y la desviación típica 7 minutos. Se elimina al 5 % de los corredores. A partir de qué tiempo queda eliminado un corredor.

    Solución: se eliminan los que tardan más de 84'48 minutos

16. Una máquina ha producido 1.000 varillas de en teoría 1 m de longitud, con una desviación típica de 0'8 mm. De ellas se necesitan 640. ¿Entre qué medidas habrá que tomar las varillas para quedarse con las más exactas?.

    Solución: N(1000,0'8)); 0'64 + 0'18=0'8200; $\left\{ \begin{array}{l} 0'8186 \\ 0'8212 \to z = 0'92 \end{array} \right.$ hay que tomarlas entre 999'27 y 1000'73

17. La media de las calificaciones obtenidas en una oposición fue $\mu = 5'5$ con una desviación típica $\sigma = 2$. Suponiendo que las calificaciones siguen una distribución normal, calcular:

    i) El porcentaje de alumnos que han sacado menos de 4.

    ii) Si hay 40 plazas y hay 2000 opositores, ¿cuál es la nota mínima para sacar plaza?

Tema 5

DISTRIBUCIÓN MUESTRAL. ESTIMACIÓN ESTADÍSTICA

5.1. Muestreo

Colectivo o población es el conjunto de elementos con alguna característica común.

Muestra es un subconjunto o parte representativa de un colectivo.

Muestreo es la operación de seleccionar los elementos de la población que van a constituir la muestra.

Puede ser **aleatorio** si se eligen al azar, **estratificado** si se divide la población en clases y en cada una se elige un número de elementos en la proporción conveniente para que la muestra reproduzca de forma adecuada los caracteres de la población.

Ejemplos

- Tres amigos hacen una quiniela poniendo respectivamente 3, 6 y 9 euros, les tocan 60.300 euros. Repartirlos proporcionalmente.

$$\left.\begin{array}{c}3\\6\\9\end{array}\right\} \quad 18; \quad \frac{60300}{18} = 3350 \quad \text{por cada euro, luego reciben} \quad \left\{\begin{array}{l}3350 \times 3 = 10050\\3350 \times 6 = 20100\\3350 \times 9 = 30150\end{array}\right.$$

- En un país, el porcentaje de declaraciones fiscales que son incorrectas es del 40 %, 60 % y 20 %, según se trate de industriales, profesionales liberales o asalariados. Se sabe que del total de declaraciones, el 10 % son de industriales, el 20 % de profesionales liberales y el resto de asalariados. Se van a realizar 1500 inspecciones:

  a) ¿Cuántos industriales, profesionales liberales y asalariados han de ser inspeccionados si se desea que la inspección sea proporcional a la probabilidad de declaración incorrecta en cada categoría profesional?

  b) Compara esta distribución de las 1500 inspecciones con la que se tendría en el caso de hacerla proporcional al número de declaraciones de cada categoría.

Sea I: industrial, L: liberal, A: asalariado, M: declaración incorrecta:

a)

declaración incorrecta	40 %	60 %	20 %
	I	L	A
total declaraciones	10 %	20 %	70 %
inspecciones	1500		

$$p(I \cap M) = 0'1 \cdot 0'4 = 0'04$$
$$p(L \cap M) = 0'2 \cdot 0'6 = 0'12$$
$$p(A \cap M) = 0'7 \cdot 0'2 = 0'14$$

Total: 0'30

$$\frac{1500}{0'30} = 5000 \quad \begin{cases} 5000\cdot = 200 \\ 5000 \cdot 0'12 = 600 \\ 5000 \cdot 0'14 = 700 \end{cases}$$

b)

$$I = 0'1 \qquad L = 0'2 \qquad A = 0'7$$

$$\frac{1500}{1} = 1500 \quad \begin{cases} 1500 \cdot 0'1 = 150 \\ 1500 \cdot 0'2 = 300 \\ 1500 \cdot 0'7 = 1050 \end{cases}$$

La **teoría de muestreo** es el estudio de las relaciones existentes entre una población y muestras extraídas de ella. Los parámetros (media, etc) de la población se suelen llamar frecuentemente parámetros, los parámetros de una muestra se suelen llamar estadísticos muestrales o simplemente estadísticos.

5.2. Distribución muestral de medias. Teorema Central del Límite.

Si consideramos todas las posibles muestras de tamaño n de una población de media μ y desviación típica σ y la media de cada muestra $\bar{x}$ obtenemos una variable aleatoria $\bar{X}$ que asigna a cada muestra su media, se llama **distribución muestral de medias** y tendrá una media y una desviación típica. .

Ejemplo Una población se compone de los cinco números 2,3,6,8,11. Considerar todas las muestras posibles de tamaño dos que pueden extraerse con reemplazamiento de esta población. Hallar: a) la media y la desviación típica de la población, b) las muestras de tamaño dos y sus medias, c) la media de la distribución muestral de medias y la desviación típica de la distribución muestral de medias.

a) $\mu = \frac{2+3+6+8+11}{5} = 6 \quad \sigma^2 = \frac{(2-6)^2+(6-3)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5} = \frac{234}{5} = 10'8; \quad \sigma = 3'29$

b) Hay $5^2 = 25$ muestras de tamaño 2

$(2,2)$	$(2,3)$	$(2,6)$	$(2,8)$	$(2,11)$
$(3,2)$	$(3,3)$	$(3,6)$	$(3,8)$	$(3,11)$
$(6,2)$	$(6,3)$	$(6,6)$	$(6,8)$	$(6,11)$
$(8,2)$	$(8,3)$	$(8,6)$	$(8,8)$	$(8,11)$
$(11,2)$	$(11,3)$	$(11,6)$	$(11,8)$	$(11,11)$

Las correspondientes medias muestrales son:

2	2'5	4	5	6'5
2'5	3	4'5	5'5	7
4	4'5	6	7	8'5
5	5'5	7	8	9'5
6'5	7	8'5	9'5	11

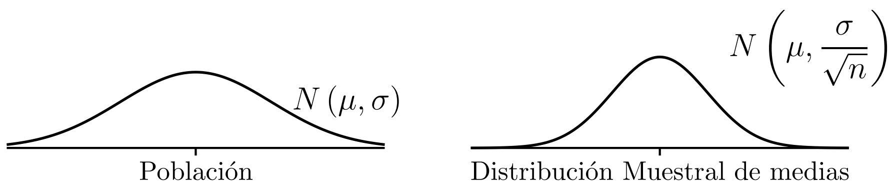
c) Introducidos estos números en la calculadora resulta:

La media de la distribución muestral de medias es 6.
La desviación típica de la distribución muestral de medias es $2'32$.

En general se tiene:

Teorema Central del Límite .

> Para **población normal o muestra grande** ($n \geq 30$), si μ, σ son los parámetros de la población entonces:
>
> la **distribución muestral de medias** $\bar{X}$ es normal $\boxed{N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)}$



Ejemplo El peso de las naranjas de un campo se distribuye normalmente con media 180 gr y desviación típica 25 gr. Hallar:

a) La probabilidad de que al coger una naranja pese menos de 190 gr.

b) La probabilidad de que en una muestra de 16 naranjas la media de la muestra sea menor que 190 gr.

c) Si cogemos 100 naranjas ¿cuántas de ellas pesarán menos de 190 gr?

d) Si cogemos 100 muestras de 16 naranjas ¿en cuántas de ellas confiamos que la media sea menor que 190?

d) ¿Entre que valores alrededor de la media 180 gr estará el 95 % de las naranjas.?

f) ¿Entre que valores alrededor de la media 180 gr estará la media de una muestra de 16 naranjas con probabilidad 0'95.?

a) Es problema elemental de normal $N(180, 25)$

$p(X < 190) = \left\{ \text{tipificando } z = \dfrac{x-\mu}{\sigma} = \dfrac{190-180}{25} = 0'4 \right\} = p(Z < 0'4) = 0'6554,$

b) Es problema de muestreo. Como la distribución de partida es normal, aunque la muestra es de tamaño menor que 30, la distribución muestral de medias $\bar{X}$ es normal $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right) = N\left(180, \dfrac{25}{\sqrt{16}}\right) = N(180, 6'25)$

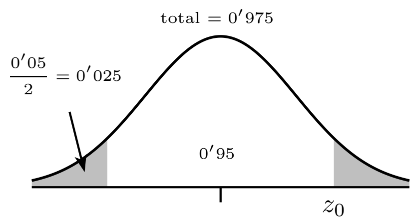
Entonces: $p(\bar{X} < 190) = \left\{ \text{tipificando } z = \dfrac{x-\mu}{\sigma} = \dfrac{190-180}{6'25} = 1'6 \right\} = p(Z < 1'6) = 0'9452$

c) Se relaciona con a):

número de naranjas con menos de 190 gr $= 100.p(X < 190) = 100 \cdot 0'6554 \approx 65$ naranjas.

d) Se relaciona con b):

número de muestras con media menor de 190 gr : $p(\bar{X} < 190),100 = 0'9452,100 = 94'52$, entre 94 y 95 de las cien de las muestras.



e) $p(180 - k < X < 180 + k) \leq 0'95$
Mirando las tablas: z_0 verificando $p(Z \leq z_0) = 0'95 + 0'05/2 = 0'975$, es $z_0 = 1'96$, destipificando $180 \pm 1'96 \cdot 25 = 180 \pm 49 = \begin{cases} = 131 \\ = 229 \end{cases}$
Por tanto el 95 % de las naranjas pesará entre 131 gr y 229 gr.

f) El cambio de variable para tipificar es $z = \dfrac{x - \mu}{\text{des. tip.}}$

En nuestro caso: $z = \dfrac{x - \mu}{\frac{\sigma}{\sqrt{n}}}$, despejando x queda

$x - \mu = z.\dfrac{\sigma}{\sqrt{n}}, \quad x = \mu + z.\dfrac{\sigma}{\sqrt{n}}$

Mirando las tablas: z_0 verificando $p(Z \leq z_0) = 0'95 + 0'05/2 = 0'975$, es $z_0 = 1'96$, destipificando $180 \pm 1'96\dfrac{25}{\sqrt{16}} = 180 \pm 12'25 = \begin{cases} = 167'75 \\ = 192'25 \end{cases}$

Por tanto: el 95 % de las medias de las muestras de 16 naranjas estará entre 167'75 gr y 192'25 gr.

5.3. Estimación estadística

En los apartados anteriores se vio como la teoría de muestreo podía emplearse para obtener información acerca de muestras extraídas al azar de una población conocida.

La estimación hace un proceso inverso, aproxima un parámetro de una población a partir de una muestra.

Si, por ejemplo, se estima la media de la población por la media de la muestra se ha hecho estimación puntual. Si lo que se da es un intervalo en el que cabe con cierta probabilidad que esté la media se ha hecho estimación por intervalo de confianza.

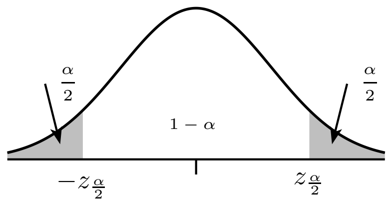
Por lo visto antes cabe afirmar: conocidos los parámetros poblacionales, que, por ejemplo, con un 95 % de confianza la media de una muestra está en un intervalo de la media poblacional. Recíprocamente conocida una muestra puedo afirmar, con un 95 % de confianza, que la media poblacional estará en un intervalo equivalente de la media de la muestra.

5.4. Estimas por intervalos de confianza

Supongamos que queremos estimar el valor de un parámetro poblacional por intervalo de confianza, se trata de encontrar un intervalo en el que esté el parámetro de la población con una probabilidad determinada $1 - \alpha$ que se llama **nivel de confianza**.

Al resto de probabilidad α se le llama **nivel de significación**.

Las distribuciones muestrales que usaremos serán normales. Al valor de la variable normal tipificada que nos da los extremos del intervalo de confianza $z_{\frac{\alpha}{2}}$ se le llama **valor crítico**.



[a]

nivel confianza $1-\alpha$	valor crítico $z_{\frac{\alpha}{2}}$
0'90	1'65
0'95	1'96
0'99	2'58

[a]Se puede pedir para otros porcentajes distintos de 90 %, 95 % 99 %

Intervalo de confianza para la media μ Los datos son: $\bar{x}, \sigma, n$.

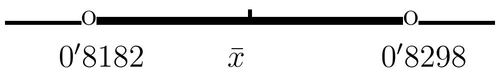
Entonces el intervalo de confianza tiene de extremos: $\bar{x} \pm z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$ con el valor crítico $z_{\frac{\alpha}{2}}$ correspondiente al nivel de confianza $1-\alpha$

Si en vez de σ lo que conocemos es s la desviación típica de la muestra, y n es grande (habitualmente se toma $n \geq 30$) en la expresión anterior se sustituye (estima) σ por s

Ejemplo Las medidas de los diámetros de una muestra al azar de 200 cojinetes de bolas hechos por una determinada máquina durante una semana dieron una media de 0'824 cm y una desviación típica de 0'042 cm. Hallar el intervalo de confianza del 95 % para el diámetro medio de todos los cojinetes.

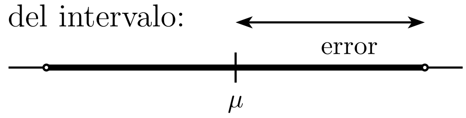
Los extremos del intervalo de confianza al(95 %)para la media μ son: $\bar{x} \pm 1'96 \dfrac{s}{\sqrt{n}} = 0'824 \pm 1'96 \dfrac{0'042}{\sqrt{200}} = 0'824 \pm 0'0058 = \begin{cases} = 0'8182\text{cm} \\ = 0'8298\text{cm} \end{cases}$ esto expresa que $p(0'8182 \leq \mu \leq 0'8298) = 0'95$

con probabilidad 95 % μ está en:



Error de la estima y tamaño muestral Error de estima o máximo o margen de error para un cierto nivel de confianza se define, como la semiamplitud del intervalo:

para las medias: **error** $= z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$



Ejemplo Al medir el tiempo de reacción, un psicólogo estima que la desviación típica del mismo es de 0'05 segundos. ¿Cuál será el número de medidas que deberá hacer para que sea del 99 % la confianza de que el error de su estima no excederá de 0'01 segundos?

El error de la estima viene dado para el nivel de confianza del 99 % por $2'58 \dfrac{\sigma}{\sqrt{n}}$, si se quiere sea menor de 0'01 entonces $2'58 \dfrac{0'05}{\sqrt{n}} \leq 0'01$

Despejamos n, $2'58 \dfrac{0'05}{\sqrt{n}} = 0'01, \quad \sqrt{n} = \dfrac{2'58 \cdot 0'05}{0'01} = 12'9 \quad n \geq 166'4.$

Así, pues, se tiene la confianza del 99 % de que el error de la estima será menor de 0'01 solamente si n es 167 o mayor.

5.5. Decisiones estadísticas. Hipótesis estadísticas

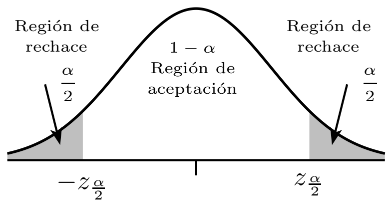
En la práctica es frecuente tener que tomar decisiones sobre una población a partir de la información suministrada por una muestra. Tales decisiones se llaman decisiones estadísticas.

Por ejemplo, se puede querer decidir a partir de los datos del muestreo, si un suero es realmente efectivo para la cura de una enfermedad, si un sistema educacional es mejor que otro, si una moneda determinada está o no cargada, etc.

Para ello se empieza formulando la hipótesis más razonable a la que se llama **hipótesis nula** y se denota H_0

Por ejemplo, si se quiere decidir si una moneda está cargada, se formula la hipótesis de que está bien, es decir: H_0 probabilidad de cara $p = 0'5$.

Una hipótesis que sea distinta de la H_0 se llama hipótesis alternativa y se denota por H_1. (En la práctica la nula es la que incluye el igual).
Lo que se va a hacer es ver con una muestra si la hipótesis nula se acepta o se rechaza. Esto se llama test de hipótesis. Se acepta si la media de la muestra cae dentro de la zona de aceptación prefijada de antemano en la distribución muestral, llamada **región de aceptación**, y se rechaza si cae fuera, o sea, en la región crítica.



Si se rechaza una hipótesis que debería ser aceptada se comete un error de Tipo I. La probabilidad máxima con la que en el test se puede cometer un error de tipo I se llama **nivel de significación** del test, se denota α.

A la situación contraria: aceptar una hipótesis que debería ser rechazada se le llama un error de Tipo II.

ERROR	
Tipo I	Rechazar H_0 siendo verdadera
Tipo II	Aceptar H_0 siendo falsa

Ejemplos

1. Se sabe que la longitud de las varillas producidas por una máquina sigue una distribución normal con desviación típica 0'2 cm. Si una muestra de 16 piezas dio una longitud media de 80'03 cm. ¿Se puede aceptar que la media de todas las varillas es 80 cm, con un nivel de significación del 10 %?.

   Planteamiento:

   Contrastamos $H_0 : \mu = 80$ cm frente a $H_1 : \mu \neq 80$ cm, es test bilateral.

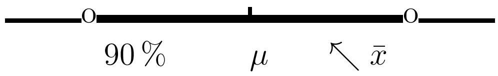
   $\sigma = 0'2$

   $n = 16$

   media muestral $\bar{x} = 80'03$

   nivel significación $\alpha = 10\,\%$ corresponde con $z_{\frac{\alpha}{2}} = 1'65$.

Resolución: El intervalo de aceptación es $\mu \pm z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}} = \mu \pm 1'65\dfrac{\sigma}{\sqrt{n}} = 80 \pm 1'65\dfrac{0'2}{\sqrt{16}} =$ $80 \pm 0'0825$ que da el intervalo $79'9175, 80'0825$. Como $\bar{x} = 80'03$ queda dentro del intervalo se acepta la hipótesis nula de que $\mu = 80$cm



Niveles de significación y valores críticos: Dependen del tipo de test:

nivel de significación α	valor crítico (bilateral) $z_{\frac{\alpha}{2}}$	valor crítico (unilateral) z_{α}
10 %	1’65	1’28
5 %	1’96	1’65
1 %	2’58	2’33

2. La duración media de una muestra de 100 tubos fluorescentes producidos por una compañía resulta ser 1.570 horas, con una desviación típica de 120 horas. Si μ es la duración media de todos los tubos producidos por la compañía, comprobar la hipótesis $\mu = 1600$ horas contra la hipótesis alternativa $\mu \neq 1600$ con un nivel de significación de 0’05.

   Planteamiento: Estimamos la desviación típica de la población por la desviación típica de la muestra.

   Contrastamos $H_0 : \mu = 1600$ cm frente a $H_1 : \mu \neq 1600$ cm, es test bilateral.

   Desv. tip. de la muestra = 120, estimamos $\sigma = 120$

   $n = 100$

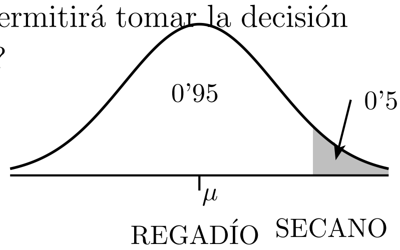
   media muestral $\bar{X} = 1570$ horas

   nivel significación $\alpha = 0'05$ corresponde con $z_{\frac{\alpha}{2}} = 1'96$.

   Resolución: El intervalo de aceptación es $\mu \pm z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}} = \mu \pm 1'96\dfrac{\sigma}{\sqrt{n}} = 1600 \pm 1'96\dfrac{120}{\sqrt{100}} =$ $1600 \pm 23'52$ que da el intervalo $(1576'48, 1623'52)$.

   Como $\bar{x} = 1570$ queda fuera del intervalo se rechaza la hipótesis nula de que $\mu = 1600$cm

3. Se quiere contrastar el contenido de azúcar de distintos cargamentos de remolacha. Se sabe que el contenido medio de azúcar en remolacha de regadío es 18 % y en cambio la media para la de secano es superior, en ambos casos la desviación típica es del 6 %. Se coge una muestra de 20 cargamentos. ¿Qué valor de la media permitirá tomar la decisión de si es de secano o de regadío al nivel de significación del 5 %?

   Planteamiento:

   Contrastamos $H_0 : \mu \leq 18\,\%$ frente a $H_1 : \mu > 18\,\%$, es test unilateral.
   Desv. tip. $\sigma = 6\,\%$
   $n = 20$
   nivel significación $\alpha = 0'05$ corresponde con $z_{\alpha} = 1'65$.

Resolución: El extremo de la región de aceptación es $\mu + z_\alpha \frac{\sigma}{\sqrt{n}} = \mu + 1'65 \frac{\sigma}{\sqrt{n}} = 18 + 1'65 \frac{6}{\sqrt{20}} = 18 + 2'21 = 20'21$.

Luego la regla para decidir es:

si la media de la muestra es menor o igual que 20'21, se acepta al nivel de significación del 5 % que el cargamento es de remolacha de regadío.

5.6. Distribución muestral de proporciones

Ejemplo Un dado de quiniela tiene como reultados 1,X,2.

a) Hallar la proporción p de resultado numérico, es decir, salir 1 o 2 al tirar el dado.

Se consideran todas las muestras posibles de tamaño 3 que se pueden formar. Hallar:

b) Las posibles muestras de tamaño 3 y sus proporciones $\hat{p}$ de resultado numérico.

c) La media de la distribución muestral de proporciones y la desviación típica de la distribución muestral.

a) al tirar el dado los tres resultados tienen igual probabilidad $p = \frac{2}{3}$

b)

muestras			$\hat{p}$
1	1	1	1
1	1	X	2/3
1	1	2	1
1	X	1	2/3
1	X	X	1/3
1	X	2	2/3
1	2	1	1
1	2	X	2/3
1	2	2	1

muestras			$\hat{p}$
2	1	1	1
2	1	X	2/3
2	1	2	1
2	X	1	2/3
2	X	X	1/3
2	X	2	2/3
2	2	1	1
2	2	X	2/3
2	2	2	1

muestras			$\hat{p}$
X	1	1	2/3
X	1	X	1/3
X	1	2	2/3
X	X	1	1/3
X	X	X	0
X	X	2	1/3
X	2	1	2/3
X	2	X	1/3
X	2	2	2/3

$\hat{p}$	n^0 de veces
0	1
1/3	6
2/3	12
1	8

c) Operando obtenemos: Media = 2/3, Desviación típica: 0'27216

Que cumple:

Media de la distribución muestral de proporciones $= p = 2/3$

Desviación típica de la distribución muestral de proporciones $= \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{2}{3}\frac{1}{3}}{3}} = \sqrt{\frac{2}{27}} = 0'27216$

pero no es normal por ser muestra pequeña.

Distribución muestral de proporciones Supongamos que tenemos una población en la que una proporción p (por ejemplo 1/2, 87 %) de esa población cumple cierta característica (por ejemplo ser aficionado a los toros). Consideremos las muestras de tamaño n y para cada una de ellas la proporción $\hat{p}$ que tiene esa característica, se tiene entonces la v. a. $\hat{P}$ que a cada muestra le asigna su proporción, es la distribución muestral de proporciones que tiene de media $= p$ y desviación típica $= \sqrt{\frac{p(1-p)}{n}}$

Para las **muestras grandes** ($np > 5, \quad n(1-p) > 5$), donde p es la proporción de la población se tiene que:

la **distribución de las proporciones de las muestras** $\hat{P}$ es normal $\boxed{N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)}$

Ejemplo Los resultados de una elección demostraron que un cierto candidato obtuvo el 46 % de los votos.

a) Determinar la probabilidad de que de 200 individuos elegidos al azar de entre la población votante se hubiese obtenido al menos un 50 % de votos para dicho candidato.

b) Si se hicieran 98 muestras de 200 individuos ¿en cuántas de ellas cabe esperar que saque mayoría el candidato?

Es $\hat{P}$ normal $N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right) = N\left(0'46, \sqrt{\dfrac{0'46 \cdot 0'54}{200}} = 0'0352\right)$

a) $p(\hat{P} \geq 0'5) = p(Z \geq \dfrac{0'50 - 0'46}{0'0352} \approx 1'13) = 1 - 0'8708 = 0'129$

b) Hemos visto que la probabilidad de que saque mayoría en una muestra de 200 es $0'129$. Entre las 98 muestras se puede esperar que en $98 \cdot 0'129 = 12'6 \approx 12$ muestras saque mayoría el candidato.

Intervalo de confianza para la proporción Los datos son: $\hat{p}, n$. Entonces los extremos del intervalo de confianza son: $\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ con el $z_{\frac{\alpha}{2}}$ correspondiente al nivel de confianza $1 - \alpha$

nota: si no dan el valor de la proporción se supone 0'5.

Ejemplo Se selecciona una muestra de 400 habitantes de nuestra ciudad y se les pregunta si son del Madrid, responden afirmativamente 180. Calcular el intervalo de confianza al 90 % para la proporción de ciudadanos partidarios del Madrid.

Tenemos $\hat{p} = \dfrac{180}{400} = 0'45$ luego:

Los extremos del intervalo de confianza al(90 %)para la proporción p son: $\hat{p} \pm 1'65\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} =$

$$0'45 \pm 1'65\sqrt{\frac{0'45 \cdot 0'55}{400}} = 0'45 \pm 1'65 \cdot 0'0248 = 0'45 \pm 0'041 = \begin{cases} = 0'408 \\ = 0'491 \end{cases}$$

Error de la estima y tamaño muestral Error de estima o máximo o margen de error para un cierto nivel de confianza se define:

para las proporciones: **error** $= z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Ejemplo Se va a realizar una encuesta entre la población de nuestra comunidad autónoma mayor de edad. Si se admite un margen de error del 3 %, ¿a cuantas personas habrá que preguntar para un nivel de confianza del 99 %?

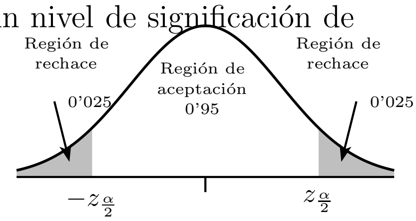
nota: cuando no se dice nada de la proporción se supone que es $0'5$

$$2'58\sqrt{\frac{0'5\cdot 0'5}{n}} \leq 0'03; \quad 2'58\cdot\frac{0'5}{\sqrt{n}} \leq 0'03; \quad ;2'58\cdot\frac{0'5}{0'03} \leq \sqrt{n}; n \geq 1849$$

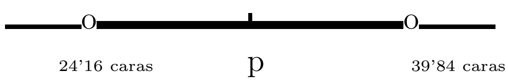
Test de contraste de hipótesis para la proporción.

Ejemplos

1. Diseñar una regla de decisión para ensayar la hipótesis de que una moneda está bien hecha si en una muestra de 64 lanzamientos de la moneda se toma un nivel de significación de 0'05.

   El nivel de significación expresa que el área de los extremos es 0'05, que corresponde con $-z_{\frac{\alpha}{2}} = -1'96, \quad z_{\frac{\alpha}{2}} = 1'96$.
   Así, pues, una regla de decisión es:
   (1) Aceptar la hipótesis de que la moneda está bien hecha si la proporción de caras en la muestra de 64 tiradas está dentro del intervalo de aceptación
   (2) Rechazar la hipótesis en cualquier otro caso.

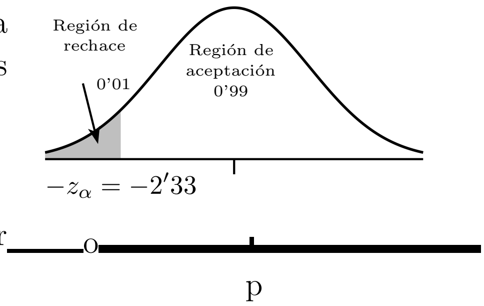
   Intervalo de aceptación:

   $$p\pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} = p\pm 1'96\sqrt{\frac{p(1-p)}{n}} = 0'5\pm\frac{0'5}{8} = 0'5\pm 0'1225 = \begin{cases} = 0'3775 \rightarrow 0'3775\cdot 64 = 24'16 \\ = 0'6225 \rightarrow 0'6225\cdot 64 = 39'84 \end{cases}$$

   (1) se acepta la hipótesis de que la moneda está bien si se obtienen entre 25 y 39 caras ambos inclusive.

   (2) se rechaza la hipótesis en caso contrario.

2. El fabricante de una patente médica afirma que la misma tiene un 90 % de efectividad en el alivio de una alergia, por un periodo de 8 horas. En una muestra de 200 individuos que tenían alergia la medicina suministrada alivió a 160 personas. Determinar si la aseveración del fabricante es cierta con un nivel de significación del $0'01$.



Denótese por p la probabilidad de obtener alivio de la alergia utilizando la medicina. Entonces se debe decidir entre las dos hipótesis:
$H_0 : p = 0'9$ y la aseveración es correcta.
$H_1 : p < 0'9$ y la aseveración es falsa.
Se elige un **ensayo por un lado**, puesto que se trata de saber si la proporción de aliviados es baja.
Para el nivel de significación $0'01$, ese área a la izquierda bajo la normal corresponde con $z_\alpha = -2'33$.
La región de aceptación tiene como extremo
$p - z_\alpha\sqrt{\dfrac{p(1-p)}{n}} = 0'9 - 2'33\sqrt{\dfrac{0'9 \cdot 0'1}{200}} = 0'85$
Luego la región de aceptación es el intervalo $(0,85,\infty)$.

Como la proporción de la muestra $\hat{p} = \dfrac{160}{200} = 0'8$ está fuera del intervalo de aceptación se rechaza H_0

Luego los resultados muestrales llevan a rechazar la afirmación del fabricante.

5.7. Problemas

1. Tres amigos invierten respectivamente 7, 3 y 5 euros en una quiniela. Aciertan y ganan 2000 euros. Repartir el premio proporcionalmente.

   Solución: $\frac{2000}{7+3+5} = 133'3; 933'1, 399'9, 666'5$

2. En un barrio se quiere hacer un estudio para conocer mejor el tipo de actividades de ocio que gustan más a sus habitantes. Para ello, van a ser encuestados 100 individuos elegidos al azar.

   a) Explica qué procedimiento de selección sería más adecuado utilizar: muestreo con o sin reposición. ¿Por qué?

   b) Como los gustos cambian con la edad y se sabe que en el barrio viven 2500 niños, 7000 adultos y 500 ancianos, más tarde se decide elegir la muestra anterior utilizando muestreo estratificado. Define los estratos y determina el tamaño muestral correspondiente a cada estrato.

   Solución: a) Sin reemplazamiento

   b) $\dfrac{A}{2500} = \dfrac{B}{7000} = \dfrac{C}{500} = \dfrac{100}{10000}$ $A = 25, B = 70, C = 5$

3. Se sabe que el cociente intelectual de los alumnos de una universidad se distribuye según una normal de media 100 y varianza 729.

   a) Hallar la probabilidad de que una muestra de 81 alumnos tenga un cociente intelectual medio inferior a 109.

   b) Hallar la probabilidad de que una muestra de 36 alumnos tenga un cociente intelectual medio superior a 109.

   c) ¿Entre qué valores alrededor de la media 100 de coeficiente intelectual estará la media de una muestra de 25 alumnos con probabilidad 0'93?

   Solución: es de muestreo,a) 99'87 %, b) 2'28 % , c) $100 \pm 9'774$

4. Se supone que los ingresos diarios en una empresa siguen una distribución normal con media 400 euros y desviación típica 250 euros.

   1. ¿Cómo se distribuye la media muestral, para muestras de tamaño n?.

   2. Se dispone de una muestra aleatoria de 25 observaciones. Calcular la probabilidad de que el promedio de ingresos esté entre 350 y 450 euros.

   Solución: $0'6826$

5. El cociente intelectual (CI) de los alumnos de un centro se distribuye $N(110, 15)$. Nos proponemos extraer una muestra aleatoria de tamaño $n = 25$.

   a. ¿Cuál es la distribución de las medias de las muestras que pueden extraerse?

   b. ¿Cuál es la probabilidad de que la media del CI de los 25 alumnos de una muestra sea superior a 115?

   c. Dar el intervalo característico de las medias muestrales correspondientes a una probabilidad del 93 % ?.

   d) ¿Cuál es el tamaño mínimo de la muestra para que el error de estimación de la media poblacional no supere a 3 con un nivel de confianza del 87 %?

   Solución:

   a) $\bar{X}$ es normal $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right) = N(110, 3)$

   b) $p(\bar{X} \geq 115) = 0'0485$

   c) $(104'564, 115'435)$

   d) $n > 57, 31$

6. Se sabe que la desviación típica del peso de los individuos de una población es 6 kg. Calcula el tamaño de la muestra que se ha de considerar para, con un nivel de confianza del 95 %, estimar el peso medio

de los individuos de la población con un error inferior a 1 kg.

Solución: error $n \geq 138'29$

7. Una máquina produce clavos de longitud media 80 mm con una desviación típica de 3 mm.

   a) ¿Cual es la probabilidad de que la longitud media de una muestra de 100 clavos sea superior a 81 mm?

   b) Si se toman 50 cajas de 100 clavos, ¿en cuántas cabe esperar que la longitud media esté comprendida entre 79 mm y 81 mm.

   Solución: es de distribución muestral, a) $p(\bar{X} > 81) = 0'0004$, b) $p(79 < \bar{X} < 81) = 0'9992$, habrá $0'9992{,}50 = 49'96 \approx 50$

8. En cierta población humana, la media muestral $\bar{X}$ de una característica se distribuye mediante una distribución normal. La probabilidad de que $\bar{X}$ sea menor o igual a 75 es 0,58 y la de que $\bar{X}$ sea mayor que 80 es 0,04. Hallar la media y la desviación típica de la población. (Tamaño muestral n = 100).

   Solución: nivel de confianza: $\mu = 74'35, \sigma = 32'25$

9. Un fabricante de bombillas sabe que la desviación típica de la duración de las bombillas es 100 horas. Calcula el tamaño de la muestra que se debe someter a prueba para tener una confianza del 95 % de que el error de la duración media que se calcula sea menor que 10 horas.

   Solución: error $n \geq 384'16$

10. El tiempo de reacción de una alarma electrónica ante un fallo del sistema es una variable aleatoria normal con desviaci ón típica 1 segundo. A partir de una muestra de 100 alarmas se ha estimado la media poblacional del tiempo de reacción, mediante un intervalo de confianza, con un error máximo de estimación igual a 0.2 segundos. ¿Con qué nivel de confianza se ha realizado la estimación?.

    Solución: $95'44\,\%$

11. Las estaturas de una muestra aleatoria de 50 estudiantes tienen una media de 174'5 cm; se conoce que la desviación típica de la variable estatura es 6'9 cm. Calcúlese un intervalo de confianza del 95 % para la estatura media de todos los estudiantes.

    Solución:

    $\mathrm{IC}(95\,\%){:}\mu \in \bar{x} \pm 1'96\dfrac{s}{\sqrt{N}} = 174'5 \pm 1'96\dfrac{6'9}{\sqrt{50}} = 174'5 \pm 1'91, \quad (172'59, 176'41)$ cm

12. Una muestra aleatoria de 100 alumnos que se presentan a las pruebas de selectividad revela que la media de edad es 18'1 años. Halla un intervalo de confianza del 90 % para la edad media de todos los estudiantes que se presentan a las pruebas, sabiendo que la desviación típica de la población es 0'4.

    Solución:

    Busquemos en $N(0,1)$ el valor de z_c correspondiente al 90 %: $p(z \leq z_c) = 0'95 = 1'65$,

    $\mathrm{IC}(90\,\%){:}\mu \in \bar{x} \pm 1'65\dfrac{\sigma}{\sqrt{N}} = 18'1 \pm 1'65\dfrac{0'4}{\sqrt{100}} = 18'1 \pm 0'066$

13. Se tiene una población $N(\mu, 2)$ y una muestra formada por 16 datos de media 2'5.

    a) Obtener el intervalo de confianza al 90 % para la media μ de la población.

    b) ¿Qué tamaño ha de tomar la muestra que permita estimar con un nivel de confianza del 95 % la media con un error de 0'2?

    Solución:

    a) Busquemos en $N(0,1)$ el valor de z_c correspondiente al 90 %: $p(z \leq z_c) = 0'95 = 1'65$, IC(90 %):$\mu \in \bar{x} \pm 1'65\dfrac{\sigma}{\sqrt{N}} = 2'5 \pm 1'65\dfrac{2}{\sqrt{16}} = 2'5 \pm 0'825$

b) para el nivel de confianza del 95 %: el error es: $1'96\dfrac{\sigma}{\sqrt{N}}$, entonces $1'96\dfrac{2}{\sqrt{N}} \leq 0'2, \quad N \geq 384'16$

14. El diámetro de unos ejes sigue una distribución normal de media desconocida y desviación típica 2 mm. Se toma una muestra de tamaño 25 y se obtiene un diámetro medio de 36 mm. ¿Se puede afirmar con un nivel de significación de 0'01 que la media de la población es de 40 mm?

    Solución:

    $H_0 : \mu = 40$, valor crítico 2'58, se rechaza pues 36 queda fuera de $(38'968, 41'032)$

15. Un equipo de psicólogos ha comprobado que en cierta población infantil el tiempo (en minutos) empleado en realizar determinada actividad manual sigue un modelo normal de probabilidad. Un grupo de 36 niños, seleccionados aleatoriamente en dicha población, realizaron esa actividad manual en un tiempo medio de 6'5 minutos con una desviación típica muestral de 1'5 minutos. A partir de esta información:

    Para un nivel de significación del 1 % ¿podríamos rechazar la hipótesis de que el tiempo medio en la población es de 7 minutos? Justifica las respuestas.

    Solución:

    $H_0 : \mu = 7$, valor crítico 2'58

    $7 \pm 0'645; (6'355, 7'645)$, Se acepta H_0

16. La capacidad de absorción de agua de las esponjas producidas por un fabricante tiene una media de 1800 ml y una desviación típica de 100 ml. mediante una nueva técnica en el proceso de fabricación se aspira a que esa capacidad pueda ser incrementada. Para contrastar esa posibilidad, se ensaya una muestra de 50 esponjas y se encuentra que su capacidad media de absorción es de 1850 ml. ¿Es admisible plantearse que, en efecto, hay un aumento de absorción al nivel de significación del 0'01?

    Solución: $H_0 : \mu = 1800, H_1 : \mu > 1800$, ensayo unilateral por la derecha, $\mu + z_\alpha\dfrac{\sigma}{\sqrt{n}} = 1800 + 32'95 = 1832'95$ Se rechaza H_0, la aspiración de mejora debe ser admitida

17. Una empresa comercializa bebidas refrescantes en un envase en cuya etiqueta se puede leer çontenido 250 cm^3. El Departamento de Consumo toma aleatoriamente 36 envases y estudia el contenido, obteniendo una media de 234 cm^3 y una desviación típica muestral de 18 cm^3. ¿Puede afirmarse con un 5 % de significación que se está estafando al público? (Consideramos estafa que el contenido sea menor que el expresado en la etiqueta.)

    Solución:

    $H_0 : \mu \geq 250, H_1 : \mu < 250$ ensayo unilateral por la izquierda

    $\mu - z_\alpha\dfrac{\sigma}{\sqrt{n}} = 250 - 1'65\dfrac{18}{\sqrt{36}} = 245'05$, la media muestral 234 queda fuera de $(245'05, \infty)$, Se rechaza H_0, los envases contienen menos de lo que dicen.

18. Se ha tomado una muestra de los precios de un mismo producto alimenticio en 16 comercios elegidos al azar en un barrio de una ciudad, y se han encontrado los siguientes precios:

    95, 108, 97, 112, 99, 106, 105, 100, 99, 98, 104, 110, 107, 111, 103, 110

    Suponiendo que los precios de este producto se distribuyen según una ley normal de varianza 25 y media desconocida:

    a) ¿Cuál es la distribución de la media muestral?

b) Determine el intervalo de confianza, al 95 %, para la media poblacional.

Solución: a) N(104; 1'25) b) (101'55; 106'45)

19. Se supone que la estatura de los chicos de 18 años de cierta población sigue una distribución normal de media 162 cm y desviación típica 12cm. Se toma una muestra al azar de 100 de estos chicos encuestados y se calcula la media.

    ¿Cuál es la probabilidad de que esta media esté entre 159 y 165 cm?

    Solución: 0'9876

20. Un fabricante de electrodomésticos sabe que la vida media de éstos sigue una distribución normal con media m = 100 meses y desviación típica s = 12 meses.

    Determínese el mínimo tamaño muestral que garantiza, con una probabilidad de 0'98, que la vida media de los electrodomésticos en dicha muestra se encuentra entre 90 y 110 meses.

    Solución: al menos 8 electrodomésticos

21. En las últimas elecciones sindicales, el 53 % de los trabajadores estaba a favor de su representante sindical. Transcurrido un año se hace una encuesta a 360 personas elegidas al azar y resultó que 176 de ellas estaban a favor de ese representante sindical. Con estos datos, ¿podemos afirmar con un nivel de confianza del 90 % que el actual representante sindical mantiene su popularidad?

    Solución: $H_0 : p = 0'53, H_1 : p \neq 0'53$ $\hat{p} = 176$ de 360 a favor $\hat{p} = \dfrac{176}{360} = 0'488$

    $p \pm 1'65\sqrt{\frac{p(1-p)}{n}} = 0'53 \pm 1'65\sqrt{\frac{0'53(1-0'53)}{360}} = 0'53 \pm 0'043; (0'487, 0'573)$ el valor $0'488$ está dentro

22. Antes de tirar 100 veces una moneda perfecta queremos saber entre qué dos valores estará el número de caras que saldrán con una probabilidad de 95 %.

    Solución: es de muestreo, entre 40 y 60 caras.

23. Se desea estimar la proporción p de individuos daltónicos de una población a través del porcentaje observado en una muestra aleatoria de individuos de tamaño n.

    a) Si el porcentaje de individuos daltónicos en la muestra es igual al 30 %, calcula el valor de n para que, con un nivel de confianza dde 0,95, el error cometido en la estimación sea inferior al 3,1 %.

    b) Si el tamaño de la muestra es de 64 individuos y el porcentaje de individuos daltónicos en la muestra es del 35 %, determina, usando un nivel de significación del 1 %, el correspondiente intervalo de confianza para la proporción de daltónicos de la población.

    Solución: 840, $(0'196, 0'504)$

24. En una determinada población se toma una muestra al azar de 256 personas. De esta muestra, el 20 % de las personas lleva gafas graduadas y el resto no. Calcula el intervalo de confianza aproximado para la proporción poblacional de las personas que llevan gafas graduadas para un nivel de confianza del 95 %.

    Solución: el intervalo de confianza para la proporción poblacional de personas con gafas es $(0'151, 0'249)$

25. El Ministerio de Educación, Política Social y Deporte desea conocer el interés de los padres por la introducción de la primera Lengua Extranjera en el primer curso de Primaria. Encuestados 1024 padres elegidos al azar, el 80 % está a favor. ¿Cuál es el intervalo de confianza para el porcentaje de los padres que están a favor

de esta medida, con un nivel de confianza del 0,99?

(0,768; 0,832)

26. Si al lanzar 80 veces una moneda se obtienen 45 caras, ¿se puede aceptar que la moneda está trucada, con un nivel de significación del 5 %?

(0'391; 0'609). Como $\hat{p} = 0'5625$ cae dentro del intervalo hallado, no puede aceptarse que la moneda está trucada.

27. Se selecciona aleatoriamente una muestra de 600 personas en una ciudad y se les pregunta si consideran que el tráfico en la misma es aceptablemente fluido. Responden afirmativamente 250 personas. ¿Cuál es el intervalo de confianza para la proporción de ciudadanos que en esa ciudad consideran aceptable la fluidez del tráfico, con un nivel de confianza del 90 %?

(0,3836; 0,4498).

28. En una encuesta realizada a 800 personas elegidas al azar del censo electoral, 240 declararon su intención de votar al partido A.

a) Estima con un nivel de confianza del 95'45 % entre que valores se encuentra la intención de voto a dicho partido en todo el censo.

b) Discute razonadamente el efecto que tendría sobre el intervalo de confianza el aumento o la disminución del nivel de confianza.

a) (0,268; 0,332)

b) Si se quiere aumentar el nivel de confianza, la amplitud del intervalo se hace mayor.

29. Para estimar la proporción de habitantes de una ciudad que poseen ordenador personal se toma una muestra de tamaño n. Calcula el valor mínimo de n para garantizar, con un nivel de confianza del 95 %, que el error de estimación no supera el 2 %. (Como se desconoce la proporción, se ha de partir del caso mas desfavorable, que será 0,5.)

El tamaño muestral debe ser de mas de 2401 habitantes.

30. Para estimar la proporción de familias de una determinada ciudad que poseen microondas, se quiere utilizar una muestra aleatoria de medida n. Calcula el valor mínimo de n para garantizar que, a un nivel de confianza del 95 %, el error en la estimación sea menor que 0'05. (Como se desconoce la proporción, se ha de tomar el caso mas desfavorable, que será 0'5.)

El tamaño muestral sera: n = 385 familias.

31. Tomada al azar una muestra de 60 alumnos de la universidad, se encontró que un tercio hablaban el idioma inglés.

a) Halla, con un nivel de confianza del 90 %, un intervalo para estimar la proporción de alumnos que hablan el idioma inglés entre los alumnos de la universidad.

b) A la vista del resultado anterior se pretende repetir la experiencia para conseguir una cota de error del 0,01 con el mismo nivel de confianza del 90 %. ¿Cuántos individuos ha de tener la muestra?

a) (0'23; 0'43)

b) El tamaño muestral ha de ser al menos de 6014 alumnos.

32. En el juzgado de cierta ciudad se presentaron en el año 2005 un total de 5500 denuncias. Se seleccionó una muestra aleatoria de un 5 % de ellas. Entre las denuncias seleccionadas se determinó que 55 habían sido producidas por violencia doméstica. Determina, justificando la respuesta:

a) La estimación puntual que podríamos dar por el porcentaje de denuncias por violencia doméstica en esa ciudad en el año 2005.

b) El error máximo que cometeríamos con dicha estimación puntual con un nivel de confianza del 99 %.

a) 20 %.

b) error= 6’2 %.

33. En los últimos meses, una cadena comercial ha intentado potenciar con precios mas atractivos y publicidad la venta de productos con la marca genérica de la cadena, frente a los de otras marcas más conocidas por los consumidores. Antes, un 15 % de los productos que vendía eran de la marca de la cadena. Recientemente, en una muestra de 200 productos vendidos, 36 eran de dicha marca. Plantea un test para contrastar que las medidas no han surtido efecto frente a que si lo han hecho, como parecen indicar los datos .A qué conclusion se llega con una significación del 10 %? Contraste bilateral para la proporción

    $(0'1083, 0'1916)$, $\hat{p} = 36/200 = 0'18$ dentro, se acepta la hipótesis nula.

34. Un experto, basado en los anteriores comicios, sostiene que si se celebrasen elecciones generales en este momento, tan solo acudiría a votar el 48 %. Preguntadas 1500 personas; 800 tienen intención de votar. ¿Supone esto, con un nivel de confianza del 99 %, que el experto se equivoca y que la participación sería mayor?

    Contraste unilateral para la proporción, $(-\infty, 0, 510)$

    Como $\hat{p} = 0'5333$ está fuera de $(-\infty, 0'51)$, se rechaza la hipótesis nula. Se deduce que la intención de voto es mayor del 48 %, por lo que se equivoca el experto.

35. De una muestra aleatoria de 225 habitantes de una población hay 18 que hablan alemán. A un nivel de significación de 0,05, .hay suficiente evidencia para refutar la afirmación de que al menos el 10 % de los habitantes de la población hablan alemán?

    Contraste unilateral para la proporción

    Como $0'08$ dentro de $(0'067; \infty)$, se acepta la hipótesis nula. Por tanto, no existe suficiente evidencia para refutar la afirmación de que al menos el 10 % de los habitantes de la población hablan alemán.