LA ENSEÑANZA DE LAS MATEMÁTICAS A TRAVÉS DE LAS NUEVAS TECNOLOGÍAS

5. PROBABILIDAD E INFERENCIA ESTADÍSTICA CON PAQUETES ESTADÍSTICOS Y HOJAS DE CÁLCULO.

MAURICIO CONTRERAS

PROBABILÍDADES CON DERIVE Y EXCEL

1. Introducción

Podemos utilizar algunos programas de ordenador para calcular probabilidades, hacer recuentos y resolver problemas relacionados con frecuencias y probabilidades. De esta forma el ordenador se convierte en un instrumento más que facilita la asignación de probabilidades a sucesos aleatorios, especialmente cuando hay que recurrir a técnicas combinatorias para efectuar recuentos. En las siguientes actividades vamos a ver algunos ejemplos de cómo usar Excel y Derive para realizar este tipo de tareas.

2. Asignación de probabilidades con Derive

- **TÉCNICAS COMBINATORIAS CON DERIVE**

- El **factorial** de un número n es: n! = n·(n–1)·(n–2)·…·3·2·1.

- Inicia el programa Derive para Windows, seleccionando el comando **Inicio / Programas / Derive para Windows / Derive para Windows**.

- Para calcular el factorial de un número con Derive, haz clic en el botón **Editar expresión**, escribe dicho número seguido del signo **!** y haz clic en el botón **Sí**. Después haz clic en el botón **=Simplificar**.

Ejemplo 1.– Calcula 4!, 25!, 10! y 6!.

a) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **4!**. Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y observa el resultado.

b) Repite el mismo proceso para calcular 25!, 10! y 6!.

Resultados: 4!=24, 25!=15511210043330985984000000, 10!=3628800, 6!=720.

- Llamamos **variaciones de m elementos tomados de n en n** (n<m) a las agrupaciones que podemos formar con n elementos diferentes tomados de los m dados, de modo que dos agrupaciones son distintas si difieren en algún elemento o en el orden de colocación de los mismos. $V_{m,n} = m \cdot (m-1) \cdot (m-2) \cdots (m-n+1)$.

- Para calcular con Derive el número de variaciones de m elementos tomados de n en n, Derive dispone de la función PERM(m, n) que está definida de la siguiente forma: $\text{PERM}(m, n) = \frac{m!}{(m-n)!} = V_{m,n}$.

Ejemplo 2.– Halla el número de variaciones de 23 elementos tomados de 4 en 4.

a) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **PERM(23, 4)**. Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y observa el resultado.

b) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **23·22·21·20**. Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y comprueba que obtienes el mismo resultado: 212520.

- Llamamos **permutación de n elementos** a cada una de las posibles ordenaciones de estos n elementos: $P_n = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1 = n!$. Dos permutaciones son iguales si todos sus elementos están en el mismo orden.
- Para calcular las permutaciones de n elementos definimos la función: **P(n)=PERM(n, n)**. Para ello haz clic en el botón **Editar expresión** y escribe **P(n):=PERM(n, n)**. Haz clic en el botón **Sí**.

Ejemplo 3.– Calcula P_5, P_3 y P_9.

a) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **P(5)** Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y observa el resultado.

b) Repite el mismo procedimiento para obtener P_3 y P_9. Comprueba los siguientes resultados: $P(5) = 120$, $P(3) = 6$, $P(9) = 362880$.

- Llamamos **variaciones con repetición de m elementos tomados de n en n**, a las distintas agrupaciones que se pueden formar con n elementos tomados de los m dados, de modo que dos agrupaciones son distintas si difieren en algún elemento o en el orden de colocación de los mismos. $VR_{m,n} = m^n$.
- Llamamos **combinaciones de m elementos tomados de n en n**, a las diferentes agrupaciones que se pueden formar con n elementos distintos de los m dados. En las combinaciones no importa el orden de los elementos. Dos combinaciones son iguales si tienen los mismos elementos. $C_{m,n} = \dfrac{V_{m,n}}{P_n} = \dfrac{m!}{n! \cdot (m-n)!}$.
- Llamamos **número combinatorio "m sobre n"** a la expresión: $\binom{m}{n} = C_{m,n}$.
- Para calcular las combinaciones de m elementos tomados de n en n, Derive dispone de la función COMB(m, n), que está definida así: COMB(m, n)= $\binom{m}{n} = \dfrac{m!}{n! \cdot (m-n)!}$.

Ejemplo 4.– Calcula $C_{5,3}$ y $C_{76,2}$

a) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **COMB(5, 3)** Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y observa el resultado.

b) Haz clic en el botón **Editar expresión** y en la caja de texto escribe **5! / (3! · (5–3) !**. Haz clic en el botón **Sí**. Haz clic en el botón **= Simplificar** y comprueba que obtienes el resultado anterior.

c) De la misma forma, calcula **COMB(76, 2)** y comprueba que obtienes el mismo resultado que al efectuar la operación **76 ! / (2 ! · (76–2) !)**.

- **RECUENTO Y PROBABILIDAD**

Vamos a ver algunos ejemplos de cómo usar Derive para resolver algunos problemas sobre recuentos y cálculo de probabilidades mediante técnicas combinatorias.

Ejemplo 1.– Lanzamos una moneda cinco veces y anotamos los resultados obteniendo una secuencia de caras y cruces. ¿Cuántas secuencias podemos obtener?.

- Se trata de formar agrupaciones de 2 elementos (C = cara, X = cruz), tomados de 5 en 5. Evidentemente, los elementos pueden repetirse. Por tanto, se trata de variaciones con repetición de 2 elementos tomados de 5 en 5, $\mathbf{VR_{2,5} = 2^5}$.
- Haz clic en el botón **Editar expresión** y en la caja de texto escribe **2 ^ 5**. Haz clic en **Sí**. Haz clic en **= Simplificar**. Obtendrás como resultado 32.

Ejemplo 2.– Las apuestas en las carreras de galgos consisten en acertar la llegada de los tres primeros perros. Si en cada carrera participan seis perros, ¿es fácil acertar?.

- Se trata de formar agrupaciones de 6 elementos (los seis perros) tomados de 3 en 3. Además el orden de llegada ABC no es el mismo que el BAC, dado que en el primer caso gana A y en el segundo gana B. Además si un perro llega en primer lugar no puede volver a ocupar ninguno de los otros dos, es decir, los elementos no pueden repetirse. Por tanto, se trata de variaciones de 6 elementos tomadas de 3 en 3, **PERM(6, 3)**.
- Haz clic en el botón **Editar expresión** y en la caja de texto escribe **PERM(6, 3)**. Haz clic en **Sí**. Haz clic en **= Simplificar** y comprueba que el resultado es 120.
- Por tanto, la probabilidad de acertar un orden de llegada es $\mathbf{P = \frac{1}{120} = 0.0083333}$. Una probabilidad muy pequeña. No es fácil acertar.

Ejemplo 3.– Seis corredoras participan en una carrera de 200 metros. ¿De cuántas maneras puede quedar la clasificación?. ¿Qué probabilidad tenemos de acertar el orden de llegada?.

- Una clasificación es una agrupación de las seis corredoras. Dos clasificaciones solamente pueden ser distintas si cambiamos el orden de llegada de las atletas. Por tanto, se trata de permutaciones de 6 elementos, $\mathbf{P_6}$.
- Haz clic en el botón Editar expresión y en la caja de texto escribe P(6). Haz clic en Sí. Haz clic en = Simplificar. Obtendrás como resultado 720.
- Por tanto, la probabilidad de acertar el orden de llegada es $\mathbf{P = \frac{1}{720} = 0.001388}$. Es decir, acertaremos, aproximadamente, una de cada 1000 veces que apostemos.

Ejemplo 4.– Para rellenar un boleto de lotería primitiva hay que tachar 6 números de entre un total de 49. Estos 6 números constituyen tu apuesta. ¿Qué probabilidad tienes de acertar la combinación ganadora?. ¿Y si haces cinco apuestas?.

- Tienes que elegir 6 números de entre 49 para formar tu apuesta. Estos números puedes tacharlos en cualquier orden, ya que lo mismo da apostar por 2-5-12-14-16-27 que por 12-5-16-2-27-14, puesto que la combinación es la misma. Se trata pues, de hallar el número de combinaciones de 49 elementos tomados de 6 en 6, COMB(49, 6).
- Haz clic en el botón Editar expresión y en la caja de texto escribe COMB(49, 6). Haz clic en Sí. Haz clic en = Simplificar. Obtendrás como resultado 139838816.
- Por tanto, la probabilidad de acertar la combinación ganadora, con una única apuesta es: $\mathbf{P = \frac{1}{139838816} = 7.15112384\ E - 8}$. Si se hacen cinco apuestas, la probabilidad de acertar es: $\mathbf{5 \times 7.15112384\ E - 8 = 3.575561921\ E - 7}$. Estas probabilidades son prácticamente nulas. Es casi imposible acertar.

3. Asignación de probabilidades con Excel

Vamos a ver tres ejemplos de hojas de cálculo de Excel relacionadas con la asignación de probabilidades. En las dos primeras usaremos técnicas combinatorias y en la tercera descubriremos por medio de simulaciones la ley de los grandes números.

- **FACTORIALES CON EXCEL**

Vamos a crear una hoja de cálculo, de nombre **factorial.xls** que permite construir una tabla de factoriales. Para ello sigue los siguientes pasos:

- Inicia el programa **Excel**, haciendo clic en **Inicio / Programas / Microsoft Excel**.
- En las celdas **A1** y **B1** de la **Hoja1**, escribe respectivamente, "**NÚMERO**" y "**FACTORIAL**".
- En la celda **A2** escribe **0**. En la celda **A3** escribe **1**.
- Selecciona el rango **A2:A3**. Sitúa el cursor en la esquina inferior derecha. Haz clic y arrastra hacia abajo, hasta seleccionar el rango **A2:A102**. De esta forma hemos generado una lista de los 100 primeros números naturales.
- Sitúa el cursor en la celda **B2** y escribe un **1**. (La factorial de 0 es 1).
- Sitúa el cursor en la celda **B3** y escribe la fórmula **=A3*B2**.
- Selecciona la celda **B3**, sitúa el cursor en la esquina inferior derecha de dicha celda. Haz clic y arrastra hacia abajo hasta seleccionar el rango **B3:B102**. De esta forma copiamos a dicho rango la fórmula situada en la celda **B3**.

El resultado es una hoja como la de la siguiente figura, que muestra los factoriales de los 100 primeros números naturales. Observa que la factorial de 50 tiene 62 cifras. ¿Cuántas cifras tiene la factorial de 100?.

	A	B
1	NÚMERO	FACTORIAL
2	0	1
3	1	1
4	2	2
5	3	6
6	4	24
7	5	120
8	6	720
9	7	5040
10	8	40320
11	9	362880
12	10	3628800
13	11	39916800
14	12	479001600
15	13	6227020800
16	14	87178291200
17	15	1,30767E+12
18	16	2,09228E+13

- Selecciona el rango **B2:B102** y pulsa la tecla **Supr**. De esta forma borramos dicho rango.
- Sitúa el cursor en la celda **B2** y haz clic en el botón $\mathbf{f_x}$ **Asistente para funciones**. Selecciona la categoría **Matemáticas y trigonométricas**. En la columna de la derecha selecciona la función **FACT** y haz clic en **Aceptar**.
- Con el cursor en la caja de texto **Número**, mueve si es necesario la ventana y haz clic en la celda **B2** (o escribe **B2** en la caja **Número**). Haz clic en **Aceptar**.
- Selecciona la celda **B2**. Sitúa el cursor en la esquina inferior derecha, haz clic y arrastra hacia abajo hasta seleccionar el rango **B2:B102**. De esta forma hemos copiado a este rango la fórmula de la celda **B2**. Observa que obtienes el mismo resultado que antes. La función **FACT** de Excel permite obtener factoriales directamente.
- Guarda la hoja con el nombre **factorial.xls**, haciendo clic en el botón **Guardar como** e introduciendo su nombre en la caja de texto **Nombre**. Haz clic en **Aceptar**.

- **<u>TÉCNICAS COMBINATORIAS CON EXCEL</u>**

Vamos a crear una hoja de cálculo, como la siguiente, de nombre **combinatoria.xls** que permitirá efectuar cálculos combinatorios con los datos que se introduzcan en algunas celdas específicas.

	D	E	F
1	**COMBINATORIA**		
2			
3	**Número de elementos**	6	
4	**Tomados de… en …**	2	
5			
6	**Variaciones**	30	
7	**Variaciones con repetición**	36	
8	**Combinaciones**	15	
9	**Permutaciones**	720	

- Haz clic en el botón **Nuevo** para crear un nuevo libro de trabajo. En la celda **D1** de la **Hoja1** escribe "**COMBINATORIA**".
- En las celdas **D3** y **D4** escribe "**Número de elementos**" y "**Tomados de… en…**", respectivamente.
- En las celdas **D6**, **D7**, **D8** y **D9** escribe "**Variaciones**", "**Variaciones con repetición**", "**Combinaciones**" y "**Permutaciones**", respectivamente.
- Haz clic en la celda **E6** y escribe la fórmula **=FACT(E3) / FACT(E3−E4)**. Utiliza para ello el botón $\mathbf{f_x}$ **Asistente para funciones** seleccionando en la categoría **Matemáticas y trigonométricas** la función **FACT**.
- Selecciona la celda **E7**, haz clic en el botón $\mathbf{f_x}$ **Asistente para funciones** y en la categoría **Matemáticas y trigonométricas** selecciona la función **POTENCIA**.
- Con el cursor en la caja **Número**, haz clic en la celda **E3**. Sitúa el cursor en la caja **Potencia** y haz clic en la celda **E4**. Haz clic en **Aceptar**. De esta forma hemos introducido la fórmula **=POTENCIA(E3; E4)**.

- En la celda **E8** escribe la siguiente fórmula **=FACT(E3) / (FACT(E4)*FACT(E3–E4))**. Utiliza para ello el botón **f_x Asistente para funciones** y la función **FACT** de la categoría **Matemáticas y trigonométricas**.
- Sitúa el cursor en la celda **E9** y haz clic en el botón **f_x Asistente para funciones**. En la categoría **Matemáticas y trigonométricas** selecciona la función **FACT**.
- Con el cursor en la caja **Número** haz clic en la celda **E3**. Haz clic en **Aceptar**. Con ello hemos introducido la fórmula **=FACT(E3)** en la celda **E9**.
- Para probar el funcionamiento de la hoja, introduce en las celdas **E3** y **E4** los números **6** y **2**, respectivamente y comprueba que aparecen los resultados de la figura anterior.
- Guarda la hoja con el nombre **combinatoria.xls**, haciendo clic en el botón **Guardar como** e introduciendo su nombre en la caja de texto **Nombre**. Haz clic en **Aceptar**.

- **LEY DE LOS GRANDES NÚMEROS**

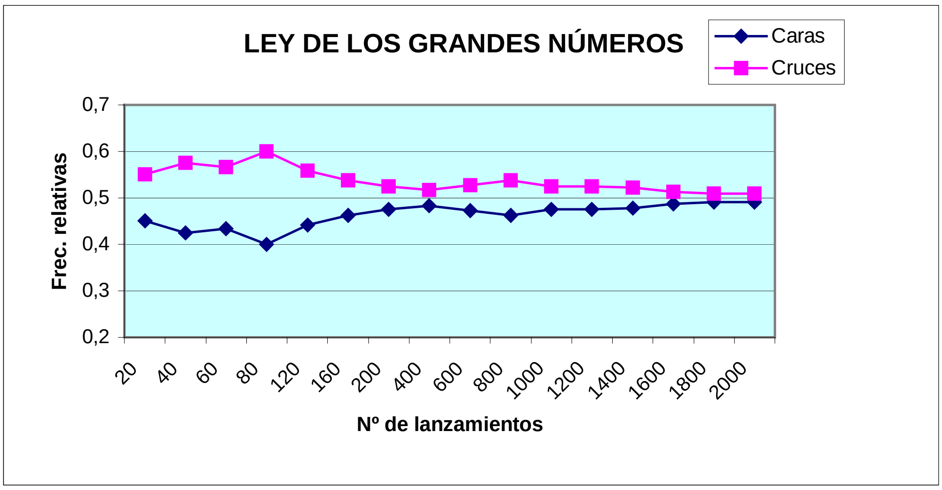
La ley de los grandes números, que permite averiguar empíricamente la probabilidad de un suceso, fue enunciada por Bernouilli. Éste observó que si se realiza un mismo experimento aleatorio un elevado número de veces, las frecuencias relativas de los sucesos tienden a estabilizarse alrededor de un número, que denominó probabilidad del suceso.

Vamos a diseñar una hoja de cálculo, de nombre **Bernouilli.xls** que permite simular el lanzamiento de una moneda un número elevado de veces (2000 lanzamientos) y sirve para comprobar experimentalmente la ley de los grandes números de Bernouilli.

- Haz clic en el botón **Nuevo** para crear un nuevo libro de trabajo.
- En la celda **A1** de la **Hoja1** escribe la fórmula **=SI(ALEATORIO( )<0,5; "Cara"; "Cruz")**, que permite simular el lanzamiento de una moneda; si el número aleatorio es menor que 0,5, el programa devolverá el valor de **Cara**, en caso contrario, devolverá el valor **Cruz**.
- Copia esta fórmula al rango **A1:T100**. Este rango contiene 2000 celdas, lo que permite simular el resultado de 2000 lanzamientos de la misma moneda.
- A continuación vamos a convertir en otro rango de celdas diferente, las caras en unos y las cruces en ceros; de este modo, al sumar el contenido de este nuevo rango, el número obtenido corresponderá a la cantidad de caras, mientras que el número de cruces se obtiene como diferencia entre el total de lanzamientos y el número de caras obtenidas.
- Introduce en la celda **A101** la función **=SI(A1="Cara"; 1; 0)** que permite obtener el valor **1** si el contenido de la celda **A1** es **Cara** y el valor **0** en caso contrario.
- Copia esta función en el rango **A101:T200** para realizar la misma operación con cada una de las celdas del rango **A1:T100**.
- Cuando se hayan simulado los lanzamientos de las monedas (rango **A1:T100**) y se haya preparado el rango de recuento (**A101:T200**), deberás diseñar el rango de resultados tal y como se muestra en la siguiente figura:

	W	X	Y
1		Frecuencias	
2	Lanzamientos	Caras	Cruces
3	=COLUMNAS($A$101:T101)*FILAS($A$101:T101)	=SUMA($A$101:T101)/W3	=1-X3
4	=COLUMNAS($A$101:T102)*FILAS($A$101:T102)	=SUMA($A$101:T101)/W4	=1-X4
5	=COLUMNAS($A$101:T103)*FILAS($A$101:T103)	=SUMA($A$101:T101)/W5	=1-X5
6	=COLUMNAS($A$101:T104)*FILAS($A$101:T104)	=SUMA($A$101:T101)/W6	=1-X6
7	=COLUMNAS($A$101:T106)*FILAS($A$101:T106)	=SUMA($A$101:T101)/W7	=1-X7
8	=COLUMNAS($A$101:T108)*FILAS($A$101:T108)	=SUMA($A$101:T101)/W8	=1-X8
9	=COLUMNAS($A$101:T110)*FILAS($A$101:T110)	=SUMA($A$101:T101)/W9	=1-X9
10	=COLUMNAS($A$101:T120)*FILAS($A$101:T120)	=SUMA($A$101:T101)/W10	=1-X10
11	=COLUMNAS($A$101:T130)*FILAS($A$101:T130)	=SUMA($A$101:T101)/W11	=1-X11
12	=COLUMNAS($A$101:T140)*FILAS($A$101:T140)	=SUMA($A$101:T101)/W12	=1-X12
13	=COLUMNAS($A$101:T150)*FILAS($A$101:T150)	=SUMA($A$101:T101)/W13	=1-X13
14	=COLUMNAS($A$101:T160)*FILAS($A$101:T160)	=SUMA($A$101:T101)/W14	=1-X14
15	=COLUMNAS($A$101:T170)*FILAS($A$101:T170)	=SUMA($A$101:T101)/W15	=1-X15
16	=COLUMNAS($A$101:T180)*FILAS($A$101:T180)	=SUMA($A$101:T101)/W16	=1-X16
17	=COLUMNAS($A$101:T190)*FILAS($A$101:T190)	=SUMA($A$101:T101)/W17	=1-X17
18	=COLUMNAS($A$101:T200)*FILAS($A$101:T200)	=SUMA($A$101:T101)/W18	=1-X18

- El número de lanzamientos se obtiene como el número de celdas de un rango determinado, que se calcula multiplicando el número de columnas por el número de filas de dicho rango.
- Las frecuencias relativas del suceso **Cara** se calculan dividiendo el número de caras (suma del rango en cuestión) entre el número de lanzamientos (calculado en la correspondiente celda de la columna **W**).
- Las frecuencias relativas del suceso **Cruz** se calculan como la diferencia entre 1 y las frecuencias relativas del suceso **Cara** (calculadas en las celdas de la columna **X**).
- Utilizando los botones **Aumentar decimales** y **Disminuir decimales**, modifica el formato numérico del rango **X3:Y18** a Fijo con cuatro decimales, para obtener todas las frecuencias relativas con las mismas cifras significativas.
- Vamos a comprobar la ley de los grandes números creando un gráfico o diagrama de líneas que muestre las frecuencias relativas de ambos sucesos (**Cara** y **Cruz**) frente al número de lanzamientos (rango **W3:Y18**). Modifica el formato del gráfico hasta que tenga el aspecto de la siguiente figura:



MODELOS PROBABILÍSTICOS CON EXCEL Y DERIVE

1. Introducción

También podemos utilizar programas de ordenador como Excel y Derive para representar gráficamente funciones de cuantía, densidad y distribución y estudiar sus propiedades. Al mismo tiempo podemos resolver problemas de probabilidad utilizando modelos probabilísticos sencillos, efectuar simulaciones y hacer predicciones. En las siguientes páginas veremos como utilizar estos programas para analizar procesos relativos a las ciencias sociales.

2. Modelos probabilísticos con Excel

Uno de los problemas fundamentales de la Estadística es la búsqueda de modelos probabilísticos que se puedan ajustar a datos empíricos. Las distribuciones de probabilidad pueden utilizarse para hacer predicciones acerca de los datos. El uso de una hoja de cálculo como Excel permite realizar muchas de las tareas relativas a las distribuciones de probabilidad más conocidas.

A continuación veremos cómo se crean y modifican distintos tipos de hojas de cálculo asociadas a problemas típicos sobre distribuciones de probabilidad. Analizaremos el uso de las posibilidades gráficas de Excel para describir modelos probabilísticos como el binomial y el normal.

- **DISTRIBUCIÓN BINOMIAL**

Dada una distribución binomial, **B(n, p)**, crea una hoja de cálculo que halle la probabilidad de obtener **k** éxitos al realizar una prueba **n** veces con una probabilidad de éxito en cada prueba igual a **p**.

Por ejemplo, queremos resolver un problema del tipo siguiente: *Halla la probabilidad de obtener una cara al lanzar una moneda 2 veces*.

- Abre el libro **Binomial y Normal** de la carpeta **Estadistica1** y comprueba que en la **Hoja1** están contenidos los siguientes datos:

	A	B	C	D
1	**k**	**n**	**p**	**Acumulado**
2	1	2	0’5	FALSO
3				
4				

- En la celda **B4**, introduce la fórmula **=DISTR.BINOM(A2; B2; C2; D2)**. Obtendrás como resultado 0’5.

- *Halla la probabilidad de que una familia con tres hijos tenga 2 chicos*. Para ello, en la casilla **A2** introduce un **2**, en la **B2** un **3**, en la **C2** escribe **0’5** y en la **D2** la palabra **FALSO**. Obtendrás como resultado **0’375**.

- *Halla la probabilidad de que una familia con tres hijos tenga a lo sumo 2 chicos*. Para ello, en la casilla **A2** introduce un **2**, en la **B2** un **3**, en la **C2** escribe **0’5** y en la **D2** la palabra **VERDADERO**. Obtendrás como resultado **0’875**.

- **TABLA Y GRÁFICA DE LA DISTRIBUCIÓN BINOMIAL**

Crea una hoja de cálculo que contenga una tabla para la distribución binomial, cuando n=5, y haz la representación gráfica de la función B(5; 0’35). Para ello sigue los siguientes pasos:

- Abre la **Hoja2** del libro **Binomial y Normal** y comprueba que contiene los siguientes datos:

	A	B	C	D	E	F	G	H	I	J	K
1	**Binomial**	**para n=5**									
2						**p**					
3	**k**	**0’05**	**0’1**	**0’15**	**0’2**	**0’25**	**0’3**	**0’35**	**0’4**	**0’45**	**0’5**
4	**0**										
5	**1**										
6	**2**										
7	**3**										
8	**4**										
9	**5**										

- En la celda **B4**, introduce la fórmula **=DISTR.BINOM($A4; 5; B$3; FALSO)**.
- Arrastra el controlador de relleno hasta el final de la fila (hasta la celda **K4**).
- Seleccionado el rango anterior, **B4: K4**, arrastra el controlador de relleno hasta el final de la columna (hasta la celda **K9**). Obtendrás como resultado la siguiente tabla:

k	**0’05**	**0’1**	**0’15**	**0’2**	**0’25**	**0’3**	**0’35**	**0’4**	**0’45**	**0’5**
0	0’7738	0’5905	0’4437	0’3277	0’2373	0’1681	0’116	0’0778	0’0503	0’0313
1	0’2036	0’3281	0’3915	0’4096	0’3955	0’3602	0’3124	0’2592	0’2059	0’1563
2	0’0214	0’0729	0’1382	0’2048	0’2637	0’3087	0’3364	0’3456	0’3369	0’3125
3	0’0011	0’0081	0’0244	0’0512	0’0879	0’1323	0’1811	0’2304	0’2757	0’3125
4	3E–05	0’0005	0’0022	0’0064	0’0146	0’0284	0’0488	0’0768	0’1128	0’1563
5	3E–07	1E–05	8E–05	0’0003	0’001	0’0024	0’0053	0’0102	0’0185	0’0313

- Para representar gráficamente la distribución B(5; 0’35), haz clic en el botón **Asistente para gráficos**.
- **Paso 1 de 4 – Tipo de gráfico**: Líneas. **Subtipo de gráfico**: Línea con marcadores en cada valor de datos.
- **Paso 2 de 4 – Datos de origen**: en la ficha **Rango de datos** introduce el rango **H4: H9**. Haz clic en la ficha **Serie** y en **Rótulos de la categoría (X)** introduce el rango **A4: A9**.
- **Paso 3 de 4 – Opciones de gráfico**: rellena la ficha **Títulos**, escribiendo como título del gráfico, **B(5;0’35)**, como título del eje OX, **Número de éxitos** y como título del eje OY, **p(A)=p**. Desactiva la ficha **Leyenda**.
- **Paso 4 de 4 – Ubicación del gráfico**: elegimos en la misma hoja.
- **DISTRIBUCIÓN NORMAL**

Dada una distribución normal, **N(μ, σ)**, crea una hoja de cálculo que halle la probabilidad **P(Z≤k)**. Por ejemplo, *halla P(Z≤1’12) en una distribución N(0, 1).* Para ello, sigue los siguientes pasos:

- Abre la **Hoja3** del libro **Binomial y Normal** y comprueba que contiene los siguientes datos:

	A	B	C	D
1	**k**	**Media**	**Desviación típica**	**Acumulado**
2	1’12	0	1	VERDADERO
3				
4		**P(Z≤K)**		

- En la celda **C4**, introduce la fórmula **=DISTR.NORM(A2; B2; C2; D2)**. Obtendrás como resultado 0’8686.
- *Halla P(Z≤6’4) en una distribución N(3, 5).* Para ello, en la casilla **A2** introduce **6’4**; en la **B2** un **3**, en la **C2** escribe **5** y en la **D2** la palabra **VERDADERO**. Obtendrás como resultado **0’7517**.

- **DISTRIBUCIÓN NORMAL INVERSA**

Dada una distribución normal, **N(μ, σ)**, crea una hoja de cálculo que halle **k** conociendo **P(Z≤k)**. Por ejemplo, *halla k sabiendo que P(Z≤k)=0’8686 en una distribución N(0, 1).* Para ello:

- Abre la **Hoja4** del libro **Binomial y Normal** y comprueba que contiene los siguientes datos:

	A	B	C
1	**P(Z≤k)**	**Media**	**Desviación típica**
2	0’8686	0	1
3			
4		**k =**	

- En la celda **C4**, introduce la fórmula **=DISTR.NORM.INV(A2; B2; C2)**. Obtendrás como resultado 1’12.

3. Tres ejemplos comentados

Ya hemos visto que es posible obtener los datos de las tablas de probabilidad y de distribución a través de la hoja de cálculo Excel. Veremos a continuación tres ejemplos obteniendo los datos de las tablas con la hoja de cálculo.

Ejemplo 1.- En un determinado centro universitario, un estudio estadístico realizado a estudiantes de primer curso constata que el 80 por 100 cursan estudios universitarios por primera vez, mientras que el resto procede de otros centros universitarios. Se eligen 10 estudiantes al azar.

a) Calcula la probabilidad de exactamente cuatro estudiantes cursen estudios universitarios por primera vez.

b) Calcula la probabilidad de que no haya más de dos estudiantes que procedan de otros centros.

Para resolver el problema, llamamos X a la variable “número de estudiantes que cursan estudios universitarios por primera vez”. Entonces X sigue una distribución B(10, 0’8). Llamamos Y a la variable “número de estudiantes que proceden de otros centros universitarios de un total de 10”. Entonces Y sigue una distribución B(10, 0’2).

a) Nos piden la probabilidad de que X=4. Una vez iniciado Excel:

Pulsamos el botón $\mathbf{f_x}$ o elegimos el comando **Insertar Función**; aparecen las funciones; de ellas seleccionamos dentro de las estadísticas: **DISTR.BINOM**, apareciéndonos la ventana en la que nos pide los parámetros de la distribución: número de éxitos, número de ensayos, probabilidad de éxito, y un valor lógico, *verdadero* para la función de distribución y *falso* para la función de probabilidad. Haciendo clic en **Aceptar**, obtenemos:

p(X=4)=0,005505024

b) Nos piden la probabilidad de que Y sea menor o igual que 2.

Realizando en la hoja de cálculo los pasos que anteriormente hemos indicado, dando en el campo acumulado el valor de verdadero, **=DISTR.BINOM(2; 10; 0,2; verdadero)**, obtenemos:

$$p(Y \leq 2)=0{,}677799526$$

Ejemplo 2.- El tiempo empleado por un equipo de atletas para realizar una determinada prueba se distribuye normalmente con media 30 minutos y desviación típica 5.

a) ¿Cuál es la probabilidad de que uno de dichos atletas tarde menos de 28 minutos en realizar la prueba?.

b) ¿Qué tiempo utiliza como máximo el 80% de los atletas de ese equipo?.

Llamamos X a la variable "tiempo empleado en la prueba". Entonces X sigue una distribución normal de parámetros N(30, 5).

a) Nos piden la probabilidad de que X sea menor que 28.

Una vez seleccionada en la hoja de cálculo la función **DISTR.NORM** introducimos los valores de x (28), la media (30), la desviación típica (5) y el valor lógico *verdadero* para obtener la función de distribución. Al hacer clic en **Aceptar**, obtenemos:

$$p(X<28)=0{,}3445783$$

b) Nos piden calcular el valor de k tal que $p(X \leq k)=0{,}8$.

Una vez seleccionada en la hoja la función **DISTR.NORM.INV** introducimos el valor de la probabilidad y de los parámetros de la distribución y Excel nos devuelve el valor de x.

Probabilidad	0,8
Media	30
Desv_estándar	5

Por lo tanto, el valor de k pedido es k=34,2081069. El tiempo empleado por el 80% de los atletas es 34 minutos 12 segundos.

Ejemplo 3.- Los resultados de una encuesta sociológica sobre el nivel de aceptación de un determinado partido político han revelado que el 25 por 100 de la población es favorable a dicho partido, siendo desfavorable el resto. En una encuesta realizada telefónicamente sobre diez personas elegidas al azar, calcula la probabilidad de que:

a) Únicamente tres sean favorables al partido.

b) Al menos una sea favorable.

c) A lo sumo dos se muestren favorables.

Si definimos el suceso A="ser favorable al partido político", entonces p(A)=0,25 y la variable X="número de personas favorables al partido político", entonces X sigue una distribución binomial B(10; 0,25).

a) Nos piden la probabilidad de que X=3. Para ello elegimos el comando **Insertar Función** y elegimos la función estadística **DISTR.BINOM**. En la siguiente ventana introducimos los valores de la siguiente tabla:

Núm_éxito	3
Ensayos	10
Prob_éxito	0,25
Acumulado	falso

Al hacer clic en **Aceptar**, obtenemos: p(X=3)=0,25028229.

b) El suceso contrario es que no haya ninguna persona favorable, es decir que X=0. Calculemos, pues, p(X=0). Una vez seleccionada la función **DISTR.BINOM** introducimos los siguientes parámetros:

Núm_éxito	0
Ensayos	10
Prob_éxito	0,25
Acumulado	falso

Al hacer clic en **Aceptar**, obtenemos p(X=0)=0,05631351. Por tanto, la probabilidad de que al menos una persona sea favorable es p(X≥1)=1–p(X=0)=1–0,05631351=0,94368649.

c) Nos piden la probabilidad p(X≤2). Una vez seleccionada la función **DISTR.BINOM** introducimos los siguientes parámetros:

Núm_éxito	2
Ensayos	10
Prob_éxito	0,25
Acumulado	verdadero

Al hacer clic en **Aceptar**, obtenemos p(X≤2)=0,5255928.

4. Gráfica de la Distribución Normal

- Abre la **Hoja5** del libro **Binomial y Normal** y crea una tabla como la siguiente:

	A	B	C	D
1	**k**	**Media**	**Desviación típica**	**Acumulado**
2		7’0000	1’5	FALSO
3	**k**	**k + Media**	**P(Z≤k)**	

- En la columna **A** introducimos los números del –4 al 4 con un paso de 0’25. Para ello, en la celda **A4** escribimos –4 y en la celda **A5** ponemos –3’75; después seleccionamos las dos celdas **A4: A5** arrastrando el controlador de relleno hasta que lleguemos al número 4, es decir, hasta **A36**.

- En la columna **B** sumamos los datos de la columna **A** a la Media; para ello, en la celda **B4** introducimos la fórmula **=A4+$B$2** arrastrando el controlador de relleno hasta que lleguemos a la altura de los datos de la columna **A**, es decir, hasta **B36**.

- En la columna **C** ponemos los valores de la probabilidad; para ello, en la celda **C4** escribimos la fórmula **=DISTR.NORM(B4; $B$2; $C$2; $D$2)**, arrastrando el controlador de relleno hasta que lleguemos a la altura de los datos de las columnas **A** y **B**, es decir, hasta **C36**.

- Para obtener la gráfica de la N(7; 1'5) hacemos clic en el botón **Asistente para gráficos**.
- **Paso 1 de 4 – Tipo de gráfico**: Líneas. **Subtipo de gráfico**: línea. Presenta tendencias a lo largo del tiempo o entre categorías.
- **Paso 2 de 4 – Datos de origen**: en la ficha **Rango de datos**, introduce el rango **C4: C36**. En la ficha **Serie** en **Rótulos de la categoría (X)** introduce el rango **B4: B36**.
- **Paso 3 de 4 – Opciones de gráfico**: rellenamos la ficha **Títulos**, introduciendo como título del gráfico **B(7; 1'5)**, como título del eje X, **valores de k** y como título del eje Y, **P(Z≤k)**. Desactivamos la **Leyenda**.
- **Paso 4 de 4 – Ubicación del gráfico**: elegimos en la misma hoja.
- Por último, mejoramos la presentación del gráfico.

5. En resumen

- **Cálculo de probabilidades en una distribución binomial**

Si X es una variable aleatoria **B(n, p)**, la probabilidad de obtener k éxitos en n pruebas es:

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

La función que utilizamos en EXCEL es **=DISTR.BINOM(k; n; p; Acumulado)**.

Acumulado es un valor lógico. Si es FALSO, calcula P(X=k) y si es VERDADERO halla P(X≤k).

- **Cálculo de probabilidades en una distribución normal.**

Si X es una variable aleatoria **N(μ, σ)** :

a) Para calcular **P(Z≤k)** la función que utilizamos en EXCEL es **=DISTR.NORM(k; μ; σ; Acumulado)**.

   Acumulado es un valor lógico. Si es VERDADERO, calcula P(Z≤k) y si es FALSO, devuelve la función de densidad de probabilidad, P(Z=k).

b) Para calcular **P(Z≥k)**, la fórmula que utilizamos en EXCEL es **=1−DISTR.NORM(k; μ; σ; Acumulado)**

c) Para calcular **P(k_1≤Z≤k_2)** utilizamos en EXCEL la función **=DISTR.NORM(k_2; μ; σ; Acumulado) – DISTR.NORM(k_1; μ; σ; Acumulado)**.

- **Cálculo del valor de k conociendo h = P(Z≤k).**

Si X es una variable aleatoria **N(μ, σ)**, la función que utilizamos en EXCEL es
=DISTR.NORM.INV(h; μ; σ).

- **<u>Distribución normal estándar</u>**

La distribución normal estándar o normal tipificada que representamos por **N(0, 1)**, es una distribución normal de media cero, **μ=0**, y desviación típica uno, **σ=1**.

Utilizamos la misma fórmula que en los casos anteriores, aunque también podemos aplicar la fórmula **=DISTR.NORM.ESTAND(k)**.

6. Modelos probabilísticos con Derive

También podemos usar Derive para trabajar con modelos probabilísticos. Para ello hemos de definir previamente las funciones de densidad de probabilidad correspondientes.

- **DISTRIBUCIONES BINOMIALES**

- Inicia el programa **Derive para Windows**, haciendo clic en su icono del escritorio, o eligiendo **Inicio / Programas / Derive para Windows / Derive para Windows**.

- Edita las expresiones de las dos funciones siguientes que definen la función de probabilidad, **PB**, y la función de distribución, **PBD**, binomiales:

  PB(k, n, p) : = comb(n, k) p ^ k (1−p) ^ (n−k)
  PBD(k, n, p) : = sum(PB(i, n, p), i, 0, k)

- **En una distribución B(3; 0,2) calcula P(X=2).**

  Solución: Edita la expresión **PB(2, 3, 0,2)**. Haz clic en **Sí**. Con la expresión seleccionada haz clic en el botón ≈ **Aproximar**. El resultado es P(X=2)=0,096.

- **En una familia con 10 hijos, calcula la probabilidad de que tenga a lo sumo dos chicos.**

  Solución: En una distribución B(10; 0,5) nos piden P(X≤2). Para resolverlo, edita la expresión **PBD(2, 10, 0,5)** y haz clic en **Sí**. Con la expresión seleccionada, haz clic en el botón ≈ **Aproximar**. El resultado es: P(X≤2)=0,0546875.

- **Dibuja la gráfica correspondiente a la función de probabilidad de una B(3; 0,5).**

  Solución: Haz clic en el botón **Editar matriz** y en la siguiente ventana introduce 4 filas y 2 columnas. Completa la siguiente ventana de la siguiente forma:

	1	**2**
1	**0**	**pb(0, 3, 0.5)**
2	**1**	**pb(1, 3, 0.5)**
3	**2**	**pb(2, 3, 0.5)**
4	**3**	**pb(3, 3, 0.5)**

  Haz clic en **Simplificar**. Haz clic en **Representar** para abrir la ventana de **Gráficos 2D**. Selecciona en el menú **Ventana / Mosaico Vertical**. Estando en la ventana **Gráficos 2D** elige **Opciones/Rejilla** e introduce en **Horizontal: 12** y en **Vertical: 13**. Elige el comando **Opciones/Puntos** y señala la opción **Conectados**. Haz clic en **Sí**.

  Haz clic en el botón **Representar**. Sitúa el puntero del ratón aproximadamente en el punto (1,5; 0,2). Haz clic en el botón ∟**Centrar en cursor**. A continuación, haz zoom, una vez en el botón **Zoom hacia fuera** y tres veces en el botón ↕ **Zoom vertical hacia fuera**. Observa el resultado.

- **LA FUNCIÓN DE DISTRIBUCIÓN NORMAL**

DERIVE tiene la función de distribución normal. Está definida con dos tipos de parámetros:

Normal(x)	es la función de distribución normal estándar N(0, 1).
Normal(x, μ, σ)	es la función de distribución normal N(μ, σ).

Con estas funciones podemos hallar cualquier probabilidad de una distribución normal, sin necesidad de utilizar tablas ni tipificar la variable.

- **Calcula P(Z≤2) en una distribución N(3, 5).**

  Solución: Edita la expresión **normal(2, 3, 5)** y haz clic en **Sí**. Haz clic en el botón ≈ **Aproximar**. El resultado es P(Z≤2)=0,42074.

- **Calcula P(Z≤−1,83) en una distribución N(0, 1).**

  Solución: Haz clic en el botón **Editar expresión** de la ventana Álgebra. Edita la expresión **normal(−1.83)** y pulsa **Sí**. Selecciona el botón ≈ **Aproximar**. El resultado es P(Z≤−1,83)=0,0336247.

- **Calcula P(1,4 ≤ Z ≤ 2,1) en una distribución N(3, 5).**

  Solución: Haz clic en el botón **Editar expresión** de la ventana Álgebra. Edita la expresión **normal(2.1, 3, 5)−normal(1.4, 3, 5)** y pulsa **Sí**. Selecciona el botón ≈ **Aproximar**. El resultado es P(1,4≤Z≤2.1)=0,0540920.

- **Dibuja la función de densidad de la variable N(0, 1).**

  Solución: La función que tenemos que dibujar es: $f(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$. En la ventana Álgebra editamos la expresión $1/\sqrt{(2\pi)}\,\hat{e}\,\text{^}(-x\,\text{^}\,2/2)$. Haz clic en el botón **Representar** para abrir la ventana **Gráficos 2D**. Selecciona el menú **Ventana / Mosaico Vertical**. Estando en la ventana **Gráficos 2D**, elige el menú **Opciones / Rejilla** e introduce en **Horizontal: 12** y en **Vertical: 13**. Haz clic en **Representar**.

  Sitúa el puntero del ratón aproximadamente en el punto (0; 0,2). Haz clic en el botón ∟**Centrar en cursor** y haz zoom una vez en el botón **Zoom hacia dentro** y dos veces en el botón ↕ **Zoom vertical hacia fuera**. Observa el resultado.

  La siguiente actividad sólo tiene sentido plantearse en 2º de Bachillerato:

- **Dibuja la función de densidad de la variable N(2; 1,5).**

a) Haz clic en el botón **Editar expresión**. Edita la expresión **normal(x, 2, 1.5)** y haz clic en **Sí**. Haz clic en el botón δ **Calcular derivadas** y haz clic en **Simplificar**.

b) Haz clic en el botón **Gráficos 2D** y elige el comando **Ventana / Mosaico Vertical**. Selecciona **Opciones / Rejilla** e introduce en **Horizontal: 12** y en **Vertical: 14**. Haz clic en el botón **Sí**.

c) Haz clic en el botón **Representar** y sitúa el puntero del ratón aproximadamente en el punto (2; 0,1). Centra con el botón ∟**Centrar en cursor** y haz zoom cuatro veces en el botón **Ampliación vertical**. Observa el resultado.

ACTIVIDADES

- **PROBABILIDADES BINOMIALES**

Utilizando el programa Derive, calcula las siguientes probabilidades:

a) P(X=5) en una distribución binomial B(8; 0,4)

b) P(X=3) en una binomial B(6; 0,3)

c) P(X≤9) en una distribución binomial B(25; 0,6)

d) P(X≤5) en una binomial B(8; 0,2)

MODELOS PROBABILÍSTICOS CON STATGRAPHICS PARA WINDOWS

1. Introducción

En las siguientes actividades estudiaremos algunas de las distribuciones de probabilidad más conocidas, como son la binomial y la normal, además de analizar algunos conceptos abstractos como variable aleatoria, probabilidad, función de probabilidad, etc. Especialmente se pretende ver la utilidad de STATGRAPHICS para resolver problemas relacionados con variables aleatorias.

2. Distribución binomial con Statgraphics

- **REPRESENTACIONES GRÁFICAS**

Desde el punto de vista del tratamiento con STATGRAPHICS, una primera aplicación puede ser comparar las gráficas de las funciones de probabilidad de distintas distribuciones binomiales, por ejemplo B(9, 0.2), B(9, 0.5) y B(9, 0.9), para detectar el efecto que tiene sobre ellas el valor del parámetro p. Para ello sigue los siguientes pasos:

- Inicia el programa STATGRAPHICS, si no lo está ya.
- Selecciona **Gráficos / Distribuciones de Probabilidad**.

Como puedes observar en el siguiente cuadro de diálogo, STATGRAPHICS permite trabajar con 24 distribuciones de probabilidad, las seis primeras de variables aleatorias discretas y las restantes de variables aleatorias continuas.

- Activa la casilla **Binomial** y haz clic sobre el botón **Aceptar**.
- Sitúa el cursor en cualquier punto de la nueva pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** en el menú de contexto.

Aparecerá una pantalla que permite el estudio simultáneo de hasta cinco distribuciones del tipo seleccionado, con diferentes valores de los parámetros.

- Introduce los datos de las tres distribuciones; en la caja **Probabilidad de Evento** escribe las probabilidades de éxito en cada prueba (parámetro p) y en el campo **Ensayos** escribe el número de pruebas (n).
- Haz clic en el botón **Aceptar**. Observa que en el panel se muestran las tres distribuciones definidas.
- Haz clic en el botón **Opciones Gráficas** de la barra de herramientas de análisis.
- En el siguiente cuadro de diálogo selecciona **Densidad / Función de Masa** y haz clic en **Aceptar**.
- Haz doble clic sobre el panel gráfico para maximizarlo.

Observa que la gráfica es simétrica para p=0,5; para valores de p menores de 0,5 presenta asimetría a la derecha y para los mayores de 0,5, asimetría a la izquierda.

- Haz clic sobre el botón **Opciones Gráficas** de la barra de herramientas de análisis.
- En el siguiente cuadro de diálogo selecciona la opción **CDF** y haz clic en **Aceptar**.

- Haz doble clic sobre el panel para maximizarlo.

Observa en el panel gráfico las gráficas de las funciones de distribución de las tres variables binomiales definidas. ¿Qué semejanzas y diferencias encuentras entre dichas gráficas?.

- Se desea saber cómo son las gráficas de las funciones de probabilidad, cuando se mantiene constante el valor del parámetro p. Utiliza el procedimiento anterior con las distribuciones B(4, 0.3), B(10, 0.3) y B(50, 0.3). ¿Qué relación existe entre dichas gráficas?.

- Compara también las gráficas de las funciones de distribución de las variables aleatorias binomiales distribuciones B(4, 0.3), B(10, 0.3) y B(50, 0.3).

- **CÁLCULO DE PROBABILIDADES**

Con STATGRAPHICS podemos calcular probabilidades con relación a un valor crítico determinado. Supón, por ejemplo, que una variable aleatoria X sigue una distribución B(12, 0.4). Para hallar las probabilidades $P(X=7)$, $P(X>3)$, $P(X\leq 8)$, $P(X<5)$ procedemos de la siguiente forma:

- Selecciona **Gráficos / Distribuciones de Probabilidad**, activa la casilla **Binomial** y haz clic en **Aceptar**.

- Haz clic sobre el botón **Opciones Tabulares**.

El cuadro de diálogo que aparece presenta las siguientes opciones:

- **Resumen del Análisis**, que el programa activa por defecto, y que muestra los valores actuales de los parámetros de la distribución.

- **Distribución Acumulada**, que proporciona las probabilidades de que la variable aleatoria tome valores menores, iguales (en el caso de variables discretas) y mayores que un valor dado, denominado valor crítico.

- **CDF Inverso**, que permite obtener el valor crítico a de la variable X para una probabilidad concreta p, de tal manera que $P(X<a)=p$.

- **Números Aleatorios**, que permite obtener una muestra de observaciones, extraídas de la distribución seleccionada.

- Activa la casilla **Distribución Acumulada** y haz clic sobre **Aceptar**.

- Sitúa el cursor en cualquier punto de la pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** en el menú de contexto.

- Introduce los valores de los parámetros de la distribución (en este caso 0.4 en **Probabilidad de Evento** y 12 en **Ensayos**) y haz clic sobre el botón **Aceptar**.

- Sitúa el cursor en cualquier punto del panel de texto **Distribución Acumulada**, pulsa el botón derecho del ratón y selecciona **Opciones de Ventana** en el menú contextual.

- Introduce los valores críticos 7, 3, 8 y 5 en el cuadro de diálogo y haz clic sobre el botón **Aceptar**.

En la nueva pantalla se obtiene:

P(X=7)=0,100902
P(X>3)=0,774663
P(X≤8)=P(X<8)+P(X=8)=0,94269+0,0420427=0,9847327
P(X<5)=0,438178

- **CÁLCULO DE VALORES CRÍTICOS**

También podemos calcular el valor crítico a de la variable X para una probabilidad concreta p, de tal modo que P(X<a)=p. Si la variable X tiene, por ejemplo, una distribución B(8, 0.75), para hallar el valor crítico a, tal que P(X<a)=0,5642 utilizaremos el siguiente procedimiento:

- Sitúa el cursor en cualquier punto de la pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** del menú de contexto.
- Introduce los valores de los parámetros de la distribución (0,75 en la casilla **Probabilidad de Evento** y 8 en **Ensayos**) y haz clic en **Aceptar**.
- Haz clic sobre el botón **Opciones Tabulares**, activa la casilla **CDF Inverso** y haz clic sobre **Aceptar**.
- Sitúa el cursor en cualquier punto del panel de texto **CDF Inverso**, pulsa el botón derecho del ratón y selecciona **Opciones de Ventana** en el menú contextual.
- En el siguiente cuadro de diálogo, introduce el valor de la probabilidad, 0.5642 y haz clic sobre **Aceptar**. En el nuevo panel se observa que el valor crítico buscado es 6.

- **UN EJEMPLO PRÁCTICO**

Un viajero de metro llega todas las mañanas a la misma hora a un andén donde encuentra el tren estacionado el 18% de las ocasiones. Suponiendo que "cada mañana" es una prueba independiente:

a) En siete días consecutivos, ¿cuál es la probabilidad de que encuentre el tren estacionado uno sólo de esos días?.

b) En quince días consecutivos, ¿cuál es la probabilidad de que encuentre el tren estacionado tres días, como máximo?.

c) En dieciocho días consecutivos, ¿cuál es la probabilidad de que encuentre el tren estacionado más de cinco días?.

SOLUCIÓN:

La variable aleatoria X="número de días, de los considerados de forma consecutiva, en los que el viajero encuentra el tren en el andén" tiene una distribución binomial, con probabilidad de éxito p=0,18. Si calculas las probabilidades pedidas con STATGRAPHICS obtendrás:

a) X~B(7, 0.18), P(X=1)=0,383048.

b) X~B(15, 0.18), P(X≤3)=P(X<3)+P(X=3)=0,476563+0,245242=0,721805

c) X~B(18, 0.18), P(X>5)=0,0889352

3. Distribución normal con Statgraphics

- **REPRESENTACIONES GRÁFICAS**

La gráfica de la función de densidad es una curva simétrica, con eje de simetría x = μ, generalmente conocida como campana de Gauss. Las variaciones en el valor de la media hacen que la gráfica se desplace a la izquierda o a la derecha; las variaciones en la desviación típica se manifiestan en el mayor o menor apuntamiento de la curva.

Para comparar las gráficas de las funciones de densidad de distintas distribuciones normales que tienen la misma desviación típica, por ejemplo N(12, 3), N(16, 3) y N(21, 3), procederemos de la siguiente forma:

- Selecciona **Gráficos / Distribuciones de Probabilidad**. En el cuadro de diálogo aparece activada por defecto la distribución normal.
- Haz clic sobre el botón **Aceptar**.
- Sitúa el cursor en cualquier punto de la nueva pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** en el menú de contexto.
- En el siguiente cuadro de diálogo, introduce los valores de los parámetros (la media en la caja **Media** y la desviación típica en el campo **Desv. Típ**.
- Haz clic sobre el botón **Aceptar**.
- Haz clic en el botón **Opciones Gráficas**.
- Selecciona la casilla **Densidad / Función de Masa** y haz clic en **Aceptar**.
- Haz doble clic sobre el panel gráfico para maximizarlo. Comenta las semejanzas y diferencias entre las gráficas.

También podemos comparar las gráficas de las funciones de distribución. Para ello:

- Haz clic sobre el botón **Opciones Gráficas**.
- Selecciona **CDF** y haz clic en **Aceptar**.
- Haz doble clic sobre el panel gráfico para maximizarlo. Comenta las gráficas obtenidas.

Para visualizar las modificaciones que experimenta la gráfica de la distribución normal para distintos valores de la desviación típica, cuando se mantiene fija la media, se pueden representar las gráficas de las funciones de densidad de las distribuciones N(13, 0.4), N(13, 0.7) y N(13, 1.2). Para ello, sigue los siguientes pasos, suponiendo que está abierto el análisis anterior:

- Selecciona **Gráficos / Distribuciones de Probabilidad** y, en el cuadro de diálogo selecciona **Normal** y haz clic sobre **Aceptar**.
- Pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** en el menú de contexto.
- Introduce los valores de los parámetros (las medias en la caja **Media** y las desviaciones típicas en el campo **Desv. Típ.**).
- Haz clic sobre el botón **Aceptar**.
- Haz clic sobre el botón **Opciones Gráficas**.

- Selecciona **Densidad / Función de Masa** y haz clic en **Aceptar**.
- Haz doble clic sobre el panel gráfico para maximizarlo. Comenta las semejanzas y diferencias entre las gráficas.

También podemos comparar las gráficas de las funciones de distribución. Para ello:

- Haz clic sobre el botón **Opciones Gráficas**.
- Selecciona **CDF** y haz clic en **Aceptar**.
- Haz doble clic sobre el panel gráfico para maximizarlo. Comenta las gráficas obtenidas.

- **CÁLCULO DE PROBABILIDADES**

Teniendo en cuenta la definición de función de distribución, el cálculo de cualquier probabilidad se convierte en el cálculo de un área bajo la gráfica de la función de densidad. En la práctica se han tabulado estas áreas para la N(0, 1), conocida como **distribución normal estándar** y que es la que se utiliza en el cálculo clásico de probabilidades. Llamando Z a la variable aleatoria con distribución normal estándar, para calcular $P(Z<k)$ sólo tenemos que buscar las unidades y las décimas de los valores de k en la columna de la izquierda, y las centésimas en la fila superior. Por ejemplo, $P(Z<0.36)=0.6406$.

Igualmente se puede calcular por medio de la tabla $P(Z\geq 0,79)$ y $P(0.05<Z<0.62)$:

$$P(Z\geq 0,79)=1-P(Z<0,79)=1-0,7852=0,2148$$

$$P(0,05<Z<0,62)=P(Z<0,62)-P(Z<0,05)=0,7324-0,5199=0,2125$$

Cualquier variable X con distribución N(μ, σ) puede transformarse en otra Z~N(0, 1), por medio del cambio de variable $Z=\dfrac{X-\mu}{\sigma}$. Esta transformación se conoce con el nombre de **tipificación de la variable** y permite utilizar la tabla de la N(0, 1) para calcular probabilidades en cualquier distribución normal N(μ, σ) ya que

$$P(X\leq x)=P\left(\frac{X-\mu}{\sigma}\leq\frac{x-\mu}{\sigma}\right)=P(Z\leq z)$$

Por ejemplo, si X es una variable N(12, 2.5) y queremos calcular $P(X\leq 13.8)$, procedemos así:

$$P(X\leq 13,8)=P\left(\frac{X-\mu}{\sigma}\leq\frac{13,8-12}{2,5}\right)=P(Z\leq 0,72)=0,7642$$

STATGRAPHICS permite calcular las probabilidades correspondientes a un valor de la variable de forma prácticamente inmediata. Por ejemplo, si la variable X tiene una distribución N(8, 2.6) y queremos calcular $P(X>11.3)$, $P(X<7.9)$, $P(-1<X<4)$ y $P(X\geq 18)$, procedemos así:

- Selecciona **Gráficos / Distribuciones de Probabilidad.** En la nueva ventana elige **Normal** y haz clic en **Aceptar**.
- Haz clic sobre el botón **Opciones Tabulares** de la barra de herramientas de análisis.
- Activa la casilla **Distribución Acumulada** y haz clic sobre **Aceptar**.
- Sitúa el cursor en cualquier punto de la nueva pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** del menú contextual.

- En el siguiente cuadro de diálogo introduce los valores de los parámetros (8 en **Media** y 2,6 en **Desv**.**Típ**.) y haz clic sobre el botón **Aceptar**.
- Sitúa el cursor en cualquier punto del panel de texto **Distribución Acumulada**, pulsa el botón derecho del ratón y selecciona **Opciones de Ventana** del menú contextual.
- En el siguiente cuadro de diálogo, introduce los valores críticos (11.3, 7.9, –1, 4 y 18).
- Haz clic sobre **Aceptar**. De la nueva pantalla deducimos:

P(X>11.3)=0,102179
P(X<7.9)=0,484657
P(–1<X<4)=P(X<4)–P(X<–1)=0,0619677–0,000268595=0,061699105
P(X≥18)=P(X>18)=0,000060015

En ocasiones interesa conocer cuáles son las probabilidades de que el valor de la variable X~N(μ, σ) se encuentre a una, dos o tres desviaciones típicas de la media, o lo que es lo mismo, contenido en intervalos de la forma $(\mu - k\sigma,\ \mu + k\sigma)$, con k=1, 2, 3. Para k=1, tipificando la variable resulta:

$$P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < Z < \frac{\mu + \sigma - \mu}{\sigma}\right) =$$
$$= P(-1 < Z < 1) = 0{,}841345 - 0{,}158655 = 0{,}68269$$

Análogamente se obtiene: P(μ–2σ<X<μ+2σ)=0,9545 P(μ–3σ<X<μ+3σ)=0,9973

Estos resultados demuestran que en una distribución normal hay una gran concentración de valores alrededor de la media.

- **CÁLCULO DE VALORES CRÍTICOS**

Para obtener valores críticos se utiliza el mismo procedimiento que el empleado para variables discretas. El programa calcula el valor crítico a de la variable X para una probabilidad concreta p, tal que P(X<a)=p. Si suponemos que la variable aleatoria X sigue una distribución normal N(5, 4), los puntos críticos a y b que verifican P(X<a)=0,9332 y P(X>b)=0,7736 se calculan utilizando en siguiente procedimiento:

- Selecciona **Gráficos / Distribuciones de Probabilidad** y en la nueva ventana elige **Normal** y haz clic sobre el botón **Aceptar**.
- Haz clic sobre el botón **Opciones Tabulares** de la barra de herramientas. Activa la casilla **CDF Inverso** y haz clic sobre **Aceptar**.
- Sitúa el cursor en cualquier punto de la nueva pantalla, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** del menú contextual.
- Introduce los valores de los parámetros (**Media**=5, **Desv.Típ.**=4) y haz clic sobre **Aceptar**.
- Sitúa el cursor en cualquier punto del panel de texto **CDF Inverso**, pulsa el botón derecho del ratón y selecciona **Opciones de Ventana** en el menú de contexto.
- Como queremos que P(X>b)=0,7736 debe ser 1–P(X≤b)=0,7736. Luego: P(X≤b)=1–0,7736=0,2264. Por tanto, la probabilidad a introducir en el nuevo cuadro de diálogo es 0,2264.
- Introduce las probabilidades (0,9332 y 0,2264) en el cuadro de diálogo. Haz clic en **Aceptar**. La nueva pantalla proporciona los resultados: a=11,0002 y b=1,99697.

- **UN EJEMPLO PRÁCTICO**

La resistencia a la compresión de las probetas de un determinado tipo de hormigón es una variable aleatoria que se distribuye normalmente con media 368 kg/cm^2 y desviación típica σ = 14 kg/cm^2.

a) Representa gráficamente la función de densidad y la función de distribución de dicha variable.

b) ¿Cuál es la probabilidad de que la resistencia de una probeta sea inferior a 370 kg/cm^2 ?.

c) ¿Cuál es la probabilidad de que la resistencia de una probeta cualquiera esté comprendida entre 355 y 380 kg/cm^2 ?.

d) ¿Cuál es el menor valor de la resistencia que no es superado por el 25% de las probetas?.

e) ¿Cuál es el mayor valor de la resistencia que es superado por el 10% de las probetas?.

f) Si se rechazan todas las probetas con resistencia menor de 342 o mayor de 410 kg/cm^2, ¿cuál será el porcentaje de probetas rechazadas?.

SOLUCIÓN:

a) Las gráficas pedidas se obtienen por el procedimiento visto en ejemplos anteriores.

b) $P(X<370)=0{,}556801$

c) $P(355<X<380)=0{,}804318-0{,}176555=0{,}627763$

d) $P(X<a)=0{,}25 \rightarrow a=358{,}557$ kg/cm^2.

e) $P(X>b)=0{,}10=1-P(X\leq b) \rightarrow P(X\leq b)=1-0{,}10=0{,}9 \rightarrow b=385{,}942$ kg/cm^2.

f) $P(X<342)+P(X>410)=0{,}0316453+0{,}00134996=0{,}03299526$. Por lo tanto, alrededor del 3,3% de las probetas se rechaza.

ACTIVIDADES

- **EFECTOS SECUNDARIOS**

Un laboratorio farmacéutico ha comprobado que el 40% de las personas que toman un antidepresivo sufren efectos secundarios. Si lo toman 15 personas, ¿qué probabilidad hay de que sufran efectos secundarios:

a) Más de 4 personas.

b) Más de 10 personas?.

- **ALOPECIA**

El 90% de los hombres padecen algún grado de alopecia. Se toman 100 hombres al azar. ¿Qué probabilidad hay de que más de la mitad padezcan alopecia?.

- **TRIBUNAL CONSTITUCIONAL**

Según DIARIO 16, de un total de 2847 asuntos presentados en el Tribunal Constitucional en el período 1991−1992 sólo se han podido resolver 2147, lo que equivale aproximadamente al 75% de los presentados. ¿Cuál es la probabilidad de que de diez asuntos elegidos al azar todos hayan sido resueltos?.

- **ENCUESTA**

Los resultados de una estadística sociológica sobre el nivel de aceptación de un determinado partido político han revelado que el 25 por 100 de la población es favorable a dicho partido, siendo desfavorable el resto. En una encuesta realizada telefónicamente sobre diez personas elegidas al azar, se desea saber la probabilidad de que:

a) Únicamente tres sean favorables al partido.

b) Al menos una sea favorable.

c) A lo sumo dos se muestren favorables.

- **MEDICAMENTO**

Se sabe que un determinado medicamento produce mejoría de cierta enfermedad a dos de cada tres pacientes. Se le administra a siete enfermos.

a) Calcula la probabilidad de que mejoren cuatro personas.

b) Calcula la probabilidad de que mejoren al menos cuatro personas.

- **APROXIMACIONES**

Considera tres distribuciones binomiales B(10; 0,1), B(200; 0,1) y B(1000; 0,1). Explica cuál de ellas se puede aproximar mejor y cuál peor por una distribución normal.

- **EMBARAZO**

La distribución de la duración de un embarazo en mujeres es aproximadamente normal, con media 266 días y desviación típica 16 días. Calcula:

a) La proporción de embarazos con una duración máxima de 244 días.

b) Los percentiles del 25%, del 50% y del 75% de la distribución considerada y comenta su significado.

- **CUARTILES**

Halla el primer y el tercer cuartil en una distribución normal N(25, 3).

- **PROFESORADO**

En una encuesta sobre la distribución de las edades del profesorado de Educación Primaria se ha observado que se distribuyen normalmente con media 38 años y desviación típica 6. De un total de 500 profesores:

a) ¿Cuántos profesores hay con edades menores o iguales a 35 años?.

b) ¿Cuántos mayores de 55 años?.

- **NORMAL ESTÁNDAR**

Calcula las siguientes áreas para una distribución normal estándar N(0, 1):

a) $P(Z<-1{,}5)$; b) $P(Z>2)$; c) $P(3<Z<4)$

Calcula los siguientes puntos críticos: $Z_{0,5}$ $Z_{0,025}$ $Z_{0,975}$ $Z_{0,98}$ $Z_{0,95}$ $Z_{0,995}$

MUESTREO E INFERENCIA ESTADÍSTICA CON EXCEL

1.– Introducción

La hoja de cálculo Excel no es, propiamente, un programa de Estadística, pero dispone de un conjunto de funciones que permiten trabajar conceptos estadísticos. En particular, Excel contiene comandos que permiten generar números aleatorios y hacer simulaciones.

Por otra parte, esta conocida hoja de cálculo dispone de un conjunto de algoritmos rápidos que se pueden ejecutar con el comando Herramientas / Análisis de datos. Algunos de estos procedimientos son especialmente útiles en Inferencia Estadística. Así ocurre con Muestra, Generación de números aleatorios, y otros de mayor nivel como Análisis de la varianza, Prueba F para dos muestras, Prueba t para dos muestras, etc, que sobrepasan los contenidos propios de Bachillerato.

En las siguientes páginas analizaremos cómo se puede usar Excel para abordar algunas cuestiones propias de Inferencia Estadística con estudiantes de la modalidad de Humanidades y Ciencias Sociales. En particular, estudiaremos las posibilidades del programa para realizar simulaciones, obtener intervalos de confianza y contrastar hipótesis sencillas.

2.– Simulación

Actualmente se dispone de software apropiado y gran parte de los paquetes integrados disponen de funciones específicas para la generación de números aleatorios, así como de funciones específicas para el tratamiento de datos estadísticos. También la hoja de cálculo Excel dispone de funciones específicas que permiten generar números aleatorios y simular experiencias aleatorias. Se trata de las funciones ENTERO y ALEATORIO que vamos a estudiar a continuación con algún ejemplo.

Ejemplo.– **Simula con la ayuda del ordenador el lanzamiento de un dado 100, 300 y 1000 veces y construye las tablas de frecuencias relativas de los resultados, comparándolos con las distribuciones teóricas.**

Si utilizamos la hoja de cálculo EXCEL, bastará generar 100, 300, 1000 números aleatorios entre 1 y 6. Para ello utilizaremos las funciones ENTERO() que devuelve un número sin su parte decimal, ALEATORIO(), que devuelve un número aleatorio mayor que cero y menor que 1. Entonces, para generar números aleatorios entre 1 y 6 utilizamos la expresión 1+ENTERO(6∗ALEATORIO()). Para que el proceso sea rápido basta con escribir dicha función en una celda y copiarla al resto de celdas que deseemos. En nuestro caso, desde A1 hasta B50; C1 hasta H50; I1 hasta AB50. Para contar el número de ceros 1, 2, 3, 4, 5 y 6 se utilizará la función CONTAR.SI. La siguiente tabla muestra los resultados obtenidos:

Valor del dado	**1**	**2**	**3**	**4**	**5**	**6**
100 veces	10	15	17	21	21	16
300 veces	44	56	41	51	48	60
1000 veces	155	161	171	181	165	167

Como vemos en la tabla, si calculamos las frecuencias relativas, a medida que el número de lanzamientos se hace más grande se van aproximado más a las posibilidades teóricas $1/6$.

La tabla de frecuencias relativas sería la siguiente:

Valor del dado	**1**	**2**	**3**	**4**	**5**	**6**
100 veces	0,10	0,15	0,17	0,21	0,21	0,16
300 veces	0,14666667	0,18666667	0,13666667	0,17	0,16	0,2
1000 veces	0,155	0,161	0,171	0,181	0,165	0,167
Prob. teórica	0,1666667	0,1666667	0,1666667	0,1666667	0,1666667	0,1666667

3.– Procedimientos rápidos con Excel

Excel permite utilizar algoritmos rápidos para generar números aleatorios y para extraer una muestra de un conjunto de datos. Estos procedimientos se encuentran en el menú Herramientas y en el comando Análisis de datos. Veamos a continuación algunos ejemplos de su funcionamiento.

- **Generación de números aleatorios**

Ejemplo.– Se supone que las estaturas de los individuos de una población siguen una distribución normal de media 174 cm y desviación típica 6 cm. Simula con el ordenador una muestra de 200 personas de esa población. A continuación halla la media y la desviación típica y comprueba si coinciden con los de la población de partida.

- Selecciona el comando **Herramientas / Análisis de datos**. En la siguiente ventana selecciona **Generación de números aleatorios** y haz clic en **Aceptar**. Aparece la ventana *Generación de números aleatorios* en la que hemos de indicar los valores adecuados de los parámetros.

- En la casilla **Número de variables** introduce **1**, en la caja **Cantidad de números aleatorios** teclea **200**. En la lista desplegable **Distribución**, selecciona **Normal** para indicar a Excel que queremos simular una variable aleatoria normal. En la caja **Parámetros** introduce la **media =174** y la **desviación típica = 6**.

- Haz clic en la casilla **Rango de salida** y, con el cursor en dicha caja, teclea el rango **$A$1: $A$200**. Finalmente, haz clic en **Aceptar**. Para que se muestren solamente resultados enteros, haz clic sucesivas veces en el botón **Disminuir decimales**, con el rango **A1: A200** seleccionado. Observa el resultado.

Vamos a calcular ahora la media y la desviación típica de la muestra obtenida.

- Haz clic en una celda vacía, por ejemplo, la **C5**. Haz clic en el botón f_x **Pegar función** y en la categoría **Estadísticas** selecciona **PROMEDIO**. En la casilla **Número1** introduce el rango **A1: A200** y haz clic en **Aceptar**. Observa el resultado.

- Sitúa el cursor en una celda vacía, por ejemplo la **C6**. Haz clic en el botón f_x **Pegar función** y en la categoría **Estadísticas** selecciona **DESVESTP**. En la casilla **Número1** introduce el rango **A1:A200** y haz clic en **Aceptar**. Observa el resultado.

Compara los valores obtenidos con los poblacionales y extrae conclusiones.

Ejemplo.– Un jugador de baloncesto tiene una probabilidad de 0,4 de encestar una canasta de tres puntos. Si efectúa 10 lanzamientos de tres puntos:

a) ¿Cuál es la probabilidad de obtener 5 canastas?.

b) ¿Cuál es la probabilidad de obtener al menos 4 canastas?.

Evidentemente el número de canastas de tres puntos conseguidas sigue una distribución binomial de parámetros p=0,4 y n=10. Para resolver el problema podemos utilizar un método de simulación, consistente en generar una muestra suficientemente grande, por ejemplo de tamaño 100, de una variable aleatoria binomial de parámetros p=0,4 y n=10 y contar, primero, el número de veces que se obtienen 5 canastas y, después, el número de veces donde se obtienen 4 o más de 4 canastas.

- Elige el comando **Herramientas / Análisis de datos**, seleccionamos **Generación de números aleatorios** y hacemos clic en **Aceptar**.

- En la casilla **Número de variables** introduce **1**, en **Cantidad de números aleatorios** teclea **100**. En la lista desplegable **Distribución**, selecciona **Binomial**. En la caja **Parámetros** introduce **Probabilidad=0,4** y **Número de muestras=10**. Haz clic en **Rango de salida** e introduce en dicha caja el rango **A1: A100**. Haz clic en **Aceptar**.
- A continuación cuenta el número de 5 que aparecen en el rango **A1: A100**. Sea n dicho número. Entonces la probabilidad de obtener 5 canastas será el cociente: $\frac{n}{100}$.
- A continuación cuenta el número de resultados 0, 1, 2 y 3 (inferiores a 4 canastas). Sea m dicho número. Entonces la probabilidad de obtener al menos 4 canastas es: $\frac{1-m}{100}$.

Evidentemente estos resultados dependen del número de simulaciones. Serán más fiables cuanto mayor sea el número de simulaciones (en virtud de la ley de los grandes números).

- **Selección de muestras aleatorias**

Ejemplo.– En una ciudad se ha estimado que la temperatura máxima que se alcanza en el día durante el mes de junio sigue una distribución normal de media 25ºC y desviación típica de 5ºC. Utilizando el ordenador simula las temperaturas máximas de los 30 días del mes de junio en esa ciudad. A continuación extrae una muestra de 10 días y averigua en cuántos de ellos la temperatura máxima está comprendida entre 22ºC y 28ºC.

En este caso, generaremos en primer lugar las temperaturas de los 30 días del mes, y posteriormente seleccionaremos una muestra de 10 días.

- Selecciona el comando **Herramientas / Análisis de datos**. En la siguiente ventana selecciona **Generación de números aleatorios** y haz clic en **Aceptar**.
- En la casilla **Número de variables** introduce **1**, en la caja **Cantidad de números aleatorios** teclea **30**. En la lista desplegable **Distribución**, selecciona **Normal** para indicar a Excel que queremos simular una variable aleatoria normal. En la caja **Parámetros** introduce la **media =25** y la **desviación típica = 5**.
- Haz clic en la casilla **Rango de salida** y, con el cursor en dicha caja, teclea el rango **$A$1: $A$30**. Finalmente, haz clic en **Aceptar**. Para que se muestren solamente resultados enteros, haz clic sucesivas veces en el botón **Disminuir decimales**, con el rango **A1: A30** seleccionado. Observa el resultado.

Vamos ahora a seleccionar una muestra de tamaño 10 de todas las temperaturas del mes.

- Selecciona el comando **Herramientas / Análisis de datos**. En la siguiente ventana selecciona **Muestra** y haz clic en **Aceptar**. Aparece la ventana *Muestra* en la que hemos de introducir los valores adecuados de los parámetros.
- En la caja **Rango de entrada** introduce **$A$1: $A$30**. En la caja **Método de muestreo** selecciona **Aleatorio**. En la caja **Número de muestras** introduce el tamaño de la muestra, es decir, teclea **10**. Haz clic en **Rango de salida** e introduce en la casilla correspondiente el rango **C1:C10**. Haz clic en **Aceptar** y observa el resultado. Si es preciso, utiliza el botón **Disminuir decimales** hasta conseguir que se muestren los resultados como números enteros.

En la muestra de tamaño 10 obtenida averigua cuantos días hay con temperatura máxima comprendida entre 22ºC y 28ºC. El resultado dependerá, como es lógico, de la muestra extraída. Pero, si se extrajeran muchas muestras de tamaño 10 de este conjunto de 30 temperaturas, los resultados se acercarían a la probabilidad teórica, en virtud de la ley de los grandes números.

4.– Intervalos de confianza y test de hipótesis

- **Hipótesis bilaterales**

Cuando se realiza un contraste de hipótesis bilateral y se rechaza la hipótesis nula (H_0: parámetro=k_0) no sabemos cuál es el valor del parámetro de la población que estamos considerando, únicamente qué valor no es, con un determinado nivel de confianza.

A veces, en lugar de realizar un contraste de hipótesis podemos construir el intervalo de confianza para dicho parámetro y con ello podemos rechazar todas aquellas hipótesis nulas de la forma H_0: parámetro = k_0 siempre que el valor no se encuentre dentro del intervalo de confianza construido.

- **Hipótesis unilaterales**

a) H_0: parámetro $\leq k_0$ frente a H_1: parámetro $> k_0$ a un nivel de significación α, se construye un intervalo de confianza para el parámetro poblacional desconocido a un nivel de confianza $1-2\cdot\alpha$. Si el valor k_0 es mayor que el extremo superior de dicho intervalo, debemos rechazar la hipótesis H_0. En general, se rechazará cualquier hipótesis nula de la forma H_0: parámetro = un valor mayor que el extremo superior del intervalo de confianza.

b) H_0: parámetro $\geq k_0$ frente a H_1: parámetro $< k_0$ a un nivel de significación α, se construye un intervalo de confianza para el parámetro poblacional desconocido a un nivel de confianza $1-2\cdot\alpha$. Si el valor k_0 es menor que el extremo inferior de dicho intervalo debemos rechazar la hipótesis H_0. Se rechazan todas las hipótesis nulas de la forma H_0: parámetro = un valor menor que el extremo inferior del intervalo de confianza.

Ejemplo.– Dos máquinas expenden cajetillas de tabaco de forma aleatoria con igual desviación típica y distribución normal. Se observa durante ciertos períodos de tiempo iguales el número de cajetillas expendidas: Máquina I: 42, 43, 41, 40; Máquina II: 38, 40, 30

Utilizando un nivel de significación del 0,05, responde las siguientes cuestiones:

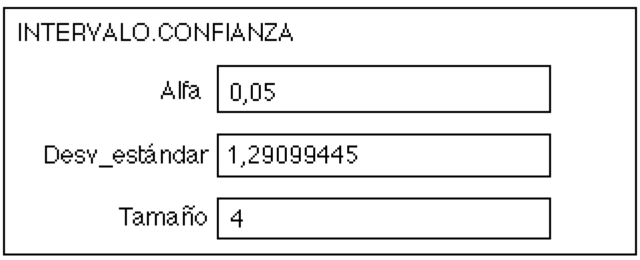
a) Halla un intervalo de confianza para el número medio de cajetillas expedido por cada máquina.

b) ¿Podemos aceptar la hipótesis de que el número medio de cajetillas expedido en la máquina I es 42?.

c) ¿Podemos aceptar la hipótesis de que el número medio de cajetillas expedido en la máquina II es mayor que 38?.

Una vez introducidos los datos en la hoja de cálculo Excel, bastará utilizar las funciones estadísticas para calcular $\bar{x}$ (PROMEDIO) y s (DESVEST) de cada población.

a) Podemos utilizar la función INTERVALO.CONFIANZA, que nos devuelve el intervalo de confianza para una población normal con desviación típica conocida. Para ello se introduce el nivel de confianza en el campo Alfa, la desviación típica en el campo Desv_estándar y el tamaño de la muestra en el campo Tamaño, como indica la figura. El resultado obtenido es: 1.26514944.



Dicha función nos devuelve la constante que debemos sumar y restar a la media muestral para obtener el intervalo de confianza de la media, luego en nuestro caso, el intervalo de confianza sería: $(41.5 - 1.265,\ 41.5 + 1.265) = (40.235,\ \ 42.765)$.

Repitiendo el procedimiento anterior para la segunda máquina, tendríamos un intervalo de confianza igual a $(36 - 5.98778001,\ \ 36 + 5.98778001) = (30.0122,\ \ 41.987)$.

b) Debemos contrastar si la media poblacional es igual a 42 en la máquina I. Como hemos construido el intervalo de confianza en el apartado (a) y vemos que contiene a 42, no rechazamos dicha hipótesis al nivel de confianza del 95% con estos datos muestrales.

c) Establecemos la hipótesis nula: H_0: media poblacional menor o igual que 38 frente a la hipótesis alternativa H_1: media poblacional mayor que 38. Para contrastar esta hipótesis, podemos construir un intervalo de confianza del (100–2·5)=90% y ver si 38 pertenece o no a dicho intervalo. Utilizando la función INTERVALO.CONFIANZA e introduciendo los valores: Alfa=0.9, Desv_estándar=5.2915026221, Tamaño=3. Obtenemos como resultado: 5.025. Este valor hemos de restarlo y sumarlo a la media 36 para obtener los extremos del intervalo de confianza. Así obtenemos como intervalo de confianza: $(36 - 5.025,\ \ 36 + 5.025) = (30.975,\ \ 41.025)$. Como 38 pertenece a este intervalo de confianza, no podemos rechazar la hipótesis nula, es decir, no podemos rechazar la hipótesis de que el número medio de cajetillas expedido en la máquina II sea menor o igual que 38.

En la siguiente figura se muestran los cálculos desarrollados y los resultados del problema:

A	B	C	D
42	38		
43	40		
41	30		
40		Intervalo de confianza de la máquina 1	1,26514944
		Intervalo de confianza de la máquina 2	5,98778001
Media	Media		
41,5	36		
Cuasivarianza	Cuasivarianza		
1,29099445	5,29150262		

ACTIVIDADES

- **PRECIOS**

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) Halla el intervalo de confianza, al 95%, para la media poblacional.

- **PESOS**

El peso medio de una muestra aleatoria de 81 personas de una determinada población es de 63,5 kg. Se sabe que la desviación típica poblacional es de 6 kg. Con un nivel de significación del 0,05, ¿hay suficientes evidencias para rechazar la afirmación de que el peso medio poblacional es de 65 kg?.

INFERENCIA ESTADÍSTICA CON STATGRAPHICS

1. Introducción

En este capítulo estudiaremos algunas cuestiones básicas de la Inferencia Estadística, como la estimación de parámetros y los contrastes de hipótesis. Podemos estimar los parámetros poblacionales partiendo de los datos muestrales mediante la estimación puntual y mediante intervalos de confianza. Por otra parte, un contraste de hipótesis es un algoritmo que permite aceptar o rechazar alguna suposición previa que se hace sobre cierto parámetro de la población, con base en la información que proporciona la muestra. Los intervalos de confianza están muy relacionados con los tests de hipótesis. STATGRAPHICS proporciona técnicas muy sencillas para resolver los problemas clásicos de Inferencia Estadística.

2. Intervalos de confianza

- **UN EJEMPLO CON STATGRAPHICS**

Consideremos el siguiente problema que resolveremos “manualmente” y con el programa:

> **El peso de los perros adultos de una determinada raza se distribuye de forma aproximadamente normal con una desviación típica de 0,65 kg. Una muestra de 32 ejemplares ha dado un peso medio de 6,5 kg y se pretende obtener un intervalo de confianza del 95% para el peso medio de estos perros.**

RESOLUCIÓN MANUAL:

Como $z_{\alpha/2} = z_{0,025} = 1{,}96$ se cumple: $\left(6{,}5 - 1{,}96 \cdot \frac{0{,}65}{\sqrt{32}};\ 6{,}5 + 1{,}96 \cdot \frac{0{,}65}{\sqrt{32}}\right) = (6{,}27;\ 6{,}73)$, es decir: con una confianza del 95%, el peso medio está entre 6,27 y 6,73 kg.

RESOLUCIÓN CON STATGRAPHICS:

Para resolver el problema seguiremos los siguientes pasos:

- Selecciona **Descripción / Contraste de Hipótesis**. Aparece un cuadro de diálogo en la que está activada por defecto la casilla **Media Normal**, que es la que se necesita para el intervalo de confianza de la media.
- Introduce el valor de la media muestral, 6.5, en **Media de la Muestra** y la desviación típica, 0.65, en **Desv. Típica de la Muestra**. No introduzcas ningún dato en la casilla **Hipótesis Nula**.
- Introduce el tamaño muestral, 32, en **Tamaño de la Muestra**.
- Haz clic sobre **Aceptar**. Si maximizas el panel de texto se visualiza la pantalla **Contraste de Hipótesis** con el intervalo de confianza buscado (6,26565; 6,73435)

STATGRAPHICS calcula los intervalos de confianza tomando por defecto el valor $1-\alpha=0{,}95$ (95%) como nivel de confianza. Para modificarlo, sólo hay que pulsar el botón derecho del ratón sobre cualquier punto de la pantalla **Contraste de Hipótesis**, consignar el nuevo valor de α en la casilla **Alpha** y hacer clic en el botón **Aceptar**.

- Utilizando el procedimiento anterior, halla un intervalo de confianza del 99% y otro con un nivel de confianza del 90%. ¿Qué ocurre al aumentar el nivel de confianza?.

- **PROCEDIMIENTO CUANDO SE DISPONE DE TODOS LOS DATOS MUESTRALES**

Cuando se conocen todos los valores muestrales, STATGRAPHICS dispone de otro procedimiento para obtener el intervalo de confianza. Para aplicarlo vamos a considerar el siguiente ejemplo:

La estatura de los individuos de cierta comarca es una variable aleatoria con distribución de probabilidad normal. Se extrae una muestra aleatoria simple con los siguientes resultados:

170	*164*	*165*	*169*	*165*	*168*	*165*	*162*	*166*

Se desea calcular un intervalo de confianza del 99% para la estatura media de la población.

Para resolver el problema con STATGRAPHICS sigue los siguientes pasos:

- Crea un archivo de datos con la variable estatura, que contenga los datos de la muestra.
- Selecciona **Descripción / Datos Numéricos / Análisis Unidimensional**.
- Haz clic sobre la variable estatura y sobre el botón **Datos**.
- Haz clic sobre el botón **Aceptar**.
- Haz clic sobre el botón **Opciones Tabulares** de la barra de herramientas de análisis.
- Activa la casilla de **Intervalos de Confianza** y pulsa **Aceptar**.
- Pulsa el botón derecho del ratón en cualquier punto del panel **Intervalos de Confianza para estatura** y selecciona **Opciones de Ventana**.
- En el cuadro de diálogo que aparece introduce el nivel de confianza elegido, 99%.
- Haz clic sobre el botón **Aceptar**. En la pantalla que aparece se visualiza el intervalo de confianza para la estatura media de la población: (163,148; 168,852).

- **INTERVALO DE CONFIANZA PARA LA PROPORCIÓN**

Si $\hat{p}$ es la proporción de individuos de una muestra aleatoria de tamaño grande, que presentan una característica concreta, el estadístico que permite construir el intervalo de confianza es $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot (1-p)}{n}}}$ y el intervalo de confianza para la proporción p de la población es:

$$\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \right)$$

Veamos con un ejemplo, como se puede utilizar STATGRAPHICS para obtener intervalos de confianza para la proporción poblacional:

Se recibe un gran lote de inodoros provenientes de un fabricante que asegura que el porcentaje de elementos defectuosos en la producción es del 1,5%. Al seleccionar una muestra de 200 artículos e inspeccionarlos se descubren 8 defectuosos.

a) Halla el intervalo de confianza del 99% para la proporción de inodoros defectuosos en el proceso de fabricación.

b) ¿Qué se puede concluir acerca de la afirmación del fabricante?.

Para aplicar el procedimiento con STATGRAPHICS sigue los siguientes pasos:

- Selecciona **Descripción / Contraste de Hipótesis**.
- Activa, en el cuadro de diálogo que aparece la casilla, **Proporción Binomial**, que es la que se necesita para el intervalo de confianza de la proporción.
- Introduce el valor de la proporción de la muestra, $8/200 = 0{,}04$, en **Proporción de la Muestra**.
- Introduce el tamaño muestral, 200, en **Tamaño de la Muestra**.
- Haz clic en **Aceptar**.
- Sitúa el cursor en cualquier punto de la pantalla **Contraste de Hipótesis**, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** del menú contextual.
- En el cuadro de diálogo que aparece introduce el valor de α correspondiente al nivel de confianza elegido (en nuestro caso, α=1%) y haz clic sobre **Aceptar**.
- Maximiza el panel de texto y visualizarás la pantalla **Contraste de Hipótesis** con el intervalo de confianza buscado (0,0130004; 0,0904546).

Como la proporción de inodoros defectuosos en la población toma valores entre el 1,30004 % y el 9,04546 %, y el 1,5% que asegura el fabricante es un valor contenido en el intervalo, de los datos muestrales concluimos, con una confianza del 99%, que no existe evidencia para dudar de la afirmación del fabricante.

3. Tamaño de una muestra

- **PRIMER CASO: ESTIMACIÓN DE LA MEDIA**

Nos planteamos ahora el problema de determinar el tamaño n de la muestra, de forma que si estimamos la media poblacional por medio de la media muestral, la distancia $\left|\overline{X}-\mu\right|$ no supere una cierta cantidad ε. Recordemos que $P\left(\overline{X}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$, por lo que se deduce que $\varepsilon\ge z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$.

De donde obtenemos: $n\ge\frac{z_{\alpha/2}^2\cdot\sigma^2}{\varepsilon^2}$. Veamos un ejemplo:

Supongamos que la duración de cierto tipo de bombillas tiene una distribución normal, con una desviación típica de 50 horas. Una muestra aleatoria de 25 bombillas ha dado una duración media de 1290 horas. Estamos interesados en conocer el tamaño mínimo de la muestra par que, con una confianza del 95%, el error en la estimación de la duración media de este tipo de bombillas no supere las 6 horas.

Para aplicar el procedimiento con STATGRAPHICS, sigue los siguientes pasos:

- Selecciona **Descripción / Determinación del Tamaño de Muestra**.
- Comprueba que en el campo **Parámetros** está activada la opción **Media Normal**.
- Introduce los datos en el cuadro de diálogo (la media 1290 en **Media Supuesta**, y la desviación típica 50 en **Desv. Típica Supuesta**) y pulsa **Aceptar**.
- En el siguiente cuadro de diálogo introduce el error absoluto (6) en la casilla E**rror Absoluto** y el nivel de confianza (95%) en la casilla **Nivel de Confianza**.
- Haz clic en **Aceptar**.

En el panel **Determinación del Tamaño de la Muestra** aparece el tamaño requerido para la muestra; en este caso n=270.

- **SEGUNDO CASO: ESTIMACIÓN DE LA PROPORCIÓN**

El problema consiste en determinar el tamaño de la muestra cuando queremos estimar la proporción poblacional a partir de los datos muestrales, de forma que la distancia entre la proporción muestral y la poblacional no supere una determinada cota de error.

Veamos un ejemplo de cómo utilizar STATGRAPHICS para resolver este problema.

Se ha observado en una muestra que la proporción de ratas que sucumben a un determinado raticida es de 0,8. Queremos conocer el tamaño muestral que con un nivel de confianza del 98% nos garantiza que la proporción muestral no va a diferir de la poblacional en más del 5%.

Para resolver este problema con STATGRAPHICS sigue los siguientes pasos:

- Selecciona **Descripción / Determinación del Tamaño de Muestra**.
- Activa la casilla **Proporción Binomial**, introduce en **Proporción Supuesta** la proporción muestral 0,8 y pulsa **Aceptar**.
- En el cuadro de diálogo que aparece, activa la casilla **Error Relativo**, inserta en ella el valor 5, introduce el nivel de confianza, 98%, y pulsa **Aceptar**.

Comprueba que la muestra debe tener un tamaño mínimo n = 484.

4. Contrastes de hipótesis

Un **contraste de hipótesis** es un procedimiento que permite aceptar o rechazar alguna suposición previa que se hace sobre cierto parámetro de la población, teniendo en cuenta la información que proporciona la muestra.

En un contraste de hipótesis se establece una hipótesis estadística (alguna afirmación) acerca de uno o varios parámetros de una o más poblaciones, se extrae una muestra aleatoria y si los resultados que ella proporciona concuerdan con la hipótesis establecida se acepta ésta, todo ello con un grado de confianza fijado de antemano.

- **CONTRASTE DE HIPÓTESIS PARA LA MEDIA**

En este caso se quiere contrastar, por ejemplo, que la media de la variable X, que se distribuye normalmente, toma el valor μ_0 frente a que toma un valor mayor.

Las hipótesis del contraste son: $H_0: \mu = \mu_0.$
$H_1: \mu > \mu_0.$

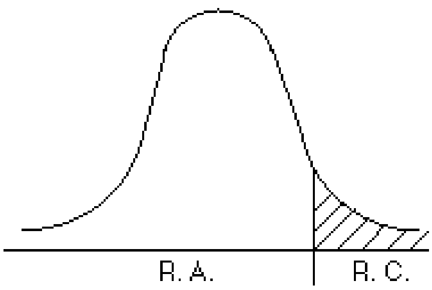
Parece lógico utilizar la media muestral $\overline{X}$ como estimador de la media poblacional μ. La región crítica estará delimitada por aquellos valores de $\overline{X}$ que apoyan la veracidad de la hipótesis alternativa, al ser significativamente mayores que μ_0. Para un nivel de significación α, $P(\overline{X} > k \mid \mu = \mu_0) = \alpha$.

Si $H_0: \mu = \mu_0$ es cierta, el estadístico $\dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ tiene una distribución T de Student con n–1 grados de libertad, por lo que $P\left(\dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} > \dfrac{k - \mu_0}{s/\sqrt{n}}\right) = P\left(t_{n-1} > \dfrac{k - \mu_0}{s/\sqrt{n}}\right) = \alpha$.

Si $t_{\alpha, n-1}$ es el valor que deja a su derecha un área de probabilidad α, se cumple $t_{\alpha, n-1} = \dfrac{k - \mu_0}{s/\sqrt{n}}$.

Despejando: $K = \mu_0 + t_{\alpha, n-1} \cdot \dfrac{s}{\sqrt{n}}$

La región crítica queda definida por $\overline{X} > \mu_0 + t_{\alpha, n-1} \cdot \dfrac{s}{\sqrt{n}}$



La región de rechazo puede expresarse también en términos del estadístico de contraste $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$. Así, H_0 debe rechazarse si $t > t_{\alpha, n-1}$ y aceptarse en caso contrario. Para el contraste unilateral izquierdo la región crítica sería $t < -t_{\alpha, n-1}$ y en el caso de contraste bilateral se rechazaría H_0 si $|t| > t_{\alpha/2, n-1}$.

En el caso en el que la varianza σ^2 de la población con distribución normal, sea conocida, lo que no es frecuente en la práctica, el estadístico de contraste a utilizar es $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ que si $H_0 : \mu = \mu_0$ es cierta, sigue una distribución normal estándar N(0, 1).

- **UN EJEMPLO CON STATGRAPHICS**

Supongamos que se desea resolver el siguiente problema:

Una máquina–herramienta produce taladros cuyo diámetro es una magnitud que se distribuye normalmente. Para contrastar si el diámetro medio de esos taladros es μ_0 = 12 mm, frente a la hipótesis de que es mayor, se toma una muestra de tamaño n = 10, que ha dado los siguientes valores:

11,9	*12,4*	*12,1*	*12,2*	*11,8*	*12,4*	*12,4*	*12,3*	*11,9*	*12,0*

Con un nivel de confianza del 95%, ¿podemos afirmar que el diámetro medio de los taladros es superior a 12 milímetros?.

RESOLUCIÓN MANUAL:

1) Las hipótesis del contraste son: $H_0: \mu = 12$
$H_1: \mu > 12$

2) El estadístico de contraste es $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$.

3) El nivel de significación es α=0,05

4) Como $t_{\alpha,\, n-1} = t_{0.05,\, 9} = 1{,}8331$, rechazamos H_0 si t >1,8331.

5) La media muestral es $\overline{X} = 12{,}14$ y la desviación típica muestral s = 0,2319. El valor del estadístico de contraste es entonces: $t = \dfrac{12{,}14 - 12}{0{,}2319/\sqrt{10}} = 1{,}9091$

6) Conclusión: como t = 1,9091 > 1,8331, rechazamos H_0 y concluimos que el diámetro medio de los taladros que produce esa máquina es mayor de 12 mm.

Si se expresa la región crítica del contraste en términos de la media muestral:

$$K = \mu_0 + t_{\alpha,\, n-1} \cdot \frac{s}{\sqrt{n}} = 12 + 1{,}8331 \cdot \frac{0{,}2319}{\sqrt{10}} = 12{,}13$$

Igualmente se rechazaría la hipótesis nula H_0, ya que $\overline{X} = 12{,}14 > 12{,}13$.

Si se opta por realizar el contraste con el enfoque del p–valor: $P(t_9 > 1{,}9091) = 0{,}0443$

H_0 debe ser rechazada con cualquier nivel de significación α >0,0443, como ocurre en este caso.

RESOLUCIÓN CON STATGRAPHICS:

Para resolver el problema con STATGRAPHICS seguiremos los siguientes pasos:

- Crea un archivo de datos con la variable taladro, que recoja los datos de la muestra.
- Selecciona **Descripción / Datos Numéricos / Análisis Unidimensional**.

- Haz clic sobre la variable taladro y sobre el botón **Datos**.
- Haz clic sobre el botón **Aceptar**.
- Haz clic sobre el botón **Opciones Tabulares** de la barra de herramientas de análisis.
- Activa la casilla de **Contraste de Hipótesis** y pulsa **Aceptar**.
- Sitúa el cursor en cualquier punto del panel **Contraste de Hipótesis para…,** pulsa el botón derecho del ratón y selecciona **Opciones de Ventana** del menú contextual.
- En el cuadro de diálogo que aparece introduce en **Media** el valor asignado a la media en la hipótesis nula (en este caso, 12) y en **Alpha** el nivel de significación elegido (en este caso, 5%).
- Selecciona el tipo de hipótesis alternativa **Mayor que**.
- Haz clic en el botón **Aceptar**.

En el panel resultante se visualizan las hipótesis del contraste, el valor que toma el estadístico, el p–valor del contraste y la conclusión, al nivel de significación elegido.

- **PROCEDIMIENTO ALTERNATIVO SI SE CONOCEN $\overline{X}$, S Y N**

Cuando no están disponibles los valores de todas las observaciones de la muestra, pero se conocen los parámetros muestrales media, desviación típica y tamaño muestral, STATGRAPHICS dispone de otro procedimiento para realizar el contraste de la media.Veamos el siguiente ejemplo:

Una empresa automovilística afirma que el consumo a los cien kilómetros de ciertos vehículos que fabrica, se distribuye normalmente con media μ_0 = 8 litros. Hay una asociación de consumidores interesada en comprobar la veracidad de esa afirmación y para ello se realizan pruebas sobre una muestra de 35 unidades del modelo, que dan como resultado un consumo medio $\overline{x}$=8,3 litros, con una desviación típica de 0,56 litros. ¿Se puede admitir como válida la afirmación de la empresa?.

Como queremos contrastar si el consumo medio corresponde o no al valor que se indica, nos interesa plantear el contraste bilateral. Para ello seguiremos los siguientes pasos:

- Selecciona **Descripción / Contraste de Hipótesis**. Aparece un cuadro de diálogo en el que está activada por defecto la casilla **Media Normal**, que es la que se necesita para el contraste de la media.
- Introduce el valor asignado a la media en la hipótesis nula (8) en la casilla **Hipótesis Nula**.
- Introduce la media muestral, 8.3, en **Media de la Muestra** y la desviación típica muestral, 0.56, en la caja **Desv. Típica de la Muestra**.
- Introduce el tamaño muestral, 35, en la caja **Tamaño de la Muestra**. Haz clic en **Aceptar**.
- Maximiza el nuevo panel de texto y visualizarás las hipótesis del contraste, el valor que toma el estadístico, el p–valor del contraste y la conclusión.

El programa toma por defecto el valor α=0,05 para el nivel de significación y muestra los resultados del contraste bilateral. Estos resultados indican que debemos rechazar la hipótesis nula, con un nivel de confianza del 95%. Es decir, la afirmación de la empresa no es cierta. Evidentemente, puedes introducir variaciones en el análisis. Por ejemplo, supongamos que queremos contrastar H_0: $\mu = 8$ frente a H_1: $\mu > 8$ para un nivel de significación $\alpha = 0,01$. Para resolverlo, sigue los siguientes pasos:

- Pulsa el botón derecho del ratón sobre el panel y selecciona **Opciones de Análisis** del menú contextual.
- En el siguiente cuadro de diálogo, activa la casilla **Mayor que**, e introduce en el campo **Alpha** el nuevo nivel de significación 1%.
- Haz clic en **Aceptar**.

Observa en el panel de texto que de esta forma se rechaza la hipótesis nula y se concluye que el consumo medio es mayor de 8 litros, con un nivel de significación del 1%.

- Utiliza este procedimiento para resolver el problema de la página 18 (taladros). Compara los resultados con los obtenidos entonces.

- **<u>CONTRASTE DE HIPÓTESIS PARA LA PROPORCIÓN</u>**

Como estimador de la proporción p de individuos de una población que presentan o no una característica concreta, se toma la proporción de la muestra $\hat{p}$.

El estadístico de contraste es $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 \cdot (1 - p_0)}{n}}}$ que, si H_0: $p = p_0$ es cierta, y el tamaño de la muestra suficientemente grande, se distribuye como la normal estándar N(0, 1).

En el contraste bilateral se rechaza H_0: $p = p_0$ si $|z| > z_{\alpha/2}$, en el contraste unilateral derecho si $z > z_\alpha$ y en el contraste unilateral izquierdo si $z < -z_\alpha$.

Para aplicar el procedimiento con STATGRAPHICS consideremos el siguiente ejemplo:

En una ciudad, una muestra de 150 motoristas dio como resultado que 17 de ellos circulaban sin el casco reglamentario. Las autoridades locales consideran que debe iniciarse una campaña de concienciación sobre su uso cuando más del 10% de los motoristas no lo utilice. ¿Hay evidencia suficiente, al nivel de significación $\alpha = 0{,}05$, para poner en marcha la campaña?.

El contraste a realizar es: H_0: $p = p_0$
H_1: $p > p_0$

Con STATGRAPHICS procedemos de la siguiente forma:

- Selecciona **Descripción / Contraste de Hipótesis**.
- Selecciona **Proporción Binomial**.
- Introduce en **Hipótesis Nula** el valor que en la hipótesis nula se asigna a la proporción, en este caso, 0,10.
- Introduce el valor de la proporción muestral en **Proporción de la Muestra**. En nuestro caso, $17/150 = 0{,}113333$.
- Indica el dato del tamaño de la muestra, 150, en el campo **Tamaño de la Muestra**.
- Haz clic en **Aceptar**.
- Maximiza el panel de texto.

- Pulsa el botón derecho del ratón sobre el panel y selecciona **Opciones de Análisis** en el menú de contexto.
- Selecciona el tipo de hipótesis alternativa **Mayor que** y pulsa **Aceptar**.

En la siguiente pantalla se visualizan los resultados del contraste. Podrás comprobar que no hay evidencia para iniciar la campaña al nivel de significación elegido.

- **PRUEBAS DE BONDAD DE AJUSTE**

Las **pruebas de bondad de ajuste** se emplean para determinar si unos datos muestrales pueden considerarse pertenecientes a una población con una distribución de probabilidad determinada.

CONTRASTE χ^2

Para realizar el contraste de la χ^2 se extrae una muestra de tamaño n de la población con la supuesta distribución de probabilidad y se agrupan las n observaciones en k intervalos de clase. Si O_i es la frecuencia de datos muestrales observada en el intervalo de clase i–ésimo y E_i es la frecuencia esperada del intervalo de clase i–ésimo según la distribución de probabilidad supuesta, cuando $n \geq 30$ y $E_i \geq 5$, el estadístico $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ tiene una distribución χ^2 con $m = n-k-p$ grados de libertad, siendo p el número de parámetros que se han estimado a partir de los datos.

Las hipótesis del contraste son: H_0: la distribución es del tipo ……
H_1: no lo es

La región crítica, al nivel de significación α, se obtiene teniendo en cuenta que si $P(\chi^2_m > \chi^2_{m,\alpha}) = \alpha$, se rechaza H_0 cuando $\chi^2 > \chi^2_{m,\alpha}$. En caso contrario, se acepta.

Este contraste se emplea para distribuciones discretas y continuas. Estudiemos el siguiente ejemplo:

En una reserva natural se realiza un estudio para medir la esperanza de vida de una cierta especie de aves. Para ello se analiza una muestra, obteniéndose los siguientes resultados:

Tiempo de vida (años)	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Frecuencia	*13*	*27*	*29*	*16*	*8*	*5*	*1*	*1*

Se piensa que la variable "Tiempo de vida" sigue una distribución normal. Contrasta la validez de esta suposición, con un nivel de significación del 5%.

Para resolver el problema con STATGRAPHICS, sigue los siguientes pasos:

- Crea un archivo de datos con la variable Tiempo, que recoja los datos de la tabla (en total hay que introducir 100 datos).
- Selecciona **Descripción / Distribuciones / Ajuste de Distribuciones (Datos no Censurados)**.
- En el siguiente cuadro de diálogo, haz clic en la variable Tiempo y haz clic en el botón **Datos**. Haz clic en **Aceptar**.
- Sitúa el cursor en cualquier punto del panel de texto **Resumen del Análisis**, pulsa el botón derecho del ratón y selecciona **Opciones de Análisis** del menú contextual.

- En la siguiente ventana se especifica el tipo de distribución de la hipótesis nula. En ella aparece seleccionada por defecto la distribución normal, que es la que se pide. Haz clic en **Aceptar**.
- El panel de texto **Resumen de Análisis** recoge la media, 2.03, y la desviación típica, 1.46649 de la distribución normal que se contrasta.
- Haz clic sobre el botón **Opciones Tabulares** de la barra de herramientas de análisis.
- Activa la casilla **Tests para la Normalidad** y haz clic en **Aceptar**.

En la siguiente ventana se muestra el resultado del test ji–cuadrado. Observa que el p–valor es 0, por tanto, hemos de rechazar la hipótesis nula, es decir, la variable Tiempo no sigue una distribución normal.

- Para confirmarlo, haz clic en el botón **Opciones Gráficas** de la barra de herramientas de análisis.
- En el siguiente cuadro de diálogo, selecciona las casillas **Gráfico de Normalidad** y **Histograma de Frecuencias** y haz clic en **Aceptar**.
- Maximiza cada uno de los paneles gráficos.

El Gráfico de Normalidad muestra que hay puntos alejados de la recta, luego la distribución no es normal. El histograma de frecuencias indica que hay una cola a la derecha, luego la variable no sigue una distribución normal.

CONTRASTE DE KOLMOGOROV–SMIRNOV PARA UNA MUESTRA

Si X es una variable aleatoria continua, pretendemos contrastar la hipótesis de que la variable sigue una determinada función de distribución F(x). Se basa este contraste en la comparación de la función de distribución teórica con la empírica. Para ello se ordenan los valores muestrales $x_1 < x_2 < \ldots < x_n$ y se define la función empírica $F_n(x)$ de la siguiente forma:

$$F_n(x) = \begin{cases} 0, & \text{si } x < x_i \\ \dfrac{i}{n}, & \text{si } x_i \le x < x_{i+1} \\ 1, & \text{si } x \ge x_n \end{cases}$$

Se calcula la discrepancia máxima entre la función de distribución empírica y la teórica por medio del estadístico de Kolmogorov–Smirnov: $D_n = \text{máx}\left| F_n(x) - F(x) \right|$, cuya distribución, cuando H_0 es cierta, está tabulada y permite obtener, fijado el nivel de significación α, el valor $d_{n,\alpha}$.

La región crítica o de rechazo del contraste viene dada por $D_n > d_{n,\alpha}$. Se acepta H_0 si $D_n < d_{n,\alpha}$. Este contraste sólo es de aplicación a distribuciones continuas. Veamos un ejemplo:

Se han medido los diámetros interiores de doce tubos con los resultados siguientes:

18,9	*29,3*	*29,0*	*28,5*	*30,2*	*31,2*	*30,8*	*30,5*	*29,7*	*30,9*	*30,8*	*31,3*

¿Es razonable la hipótesis de normalidad para la población de los diámetros interiores de los tubos, al nivel de significación del 5%?.

Para resolverlo con STATGRAPHICS sigue los siguientes pasos:

- Crea un archivo de datos con la variable diaminter, que recoja los datos anteriores.
- Selecciona **Descripción / Distribuciones / Ajuste de Distribuciones (datos no Censurados)**.
- En el cuadro de diálogo que aparece, introduce la variable diaminter en **Datos** y haz clic en **Aceptar**. Los datos que aparecen en el panel de texto **Resumen de Análisis** corresponden, por defecto, a la distribución normal, que es la que nos interesa.
- Haz clic en el botón **Opciones Tabulares** de la barra de herramientas.
- Activa la casilla **Tests de Bondad de Ajuste** y pulsa **Aceptar**.

En el panel de texto Tests de Bondad de Ajuste para … se muestran las cantidades máximas en que la función de distribución empírica excede o está por debajo de la función de distribución teórica, DMAS y DMENOS respectivamente, el estadístico de Kolmogorov, DN, que es el mayor de los dos valores anteriores y el p–valor del contraste. En este caso, el p–valor es 0,151274 > 0,05. Por tanto, no se puede rechazar la hipótesis nula de normalidad.

- **PRUEBAS DE NORMALIDAD**

Para ver si los datos se ajustan a una distribución normal podemos utilizar diferentes tipos de contrastes:

- Contraste de Shapiro y Wilks.
- Contraste de asimetría.
- Contraste de curtosis.

Estos constrastes se basan en comparar la asimetría y la curtosis de la distribución empírica con las de la distribución normal. Veamos con un ejemplo, como se puede utilizar STATGRAPHICS para aplicar estos contrastes.

A quince alumnos de un centro de primaria se les sometió a un test de nivel de vocabulario. Las puntuaciones obtenidas fueron las siguientes:

5,1	*5,2*	*5,1*	*6,7*	*3,7*	*7,0*	*6,2*	*7,5*	*6,3*	*5,3*	*5,8*	*4,1*	*6,7*	*5,0*	*5,7*

Se quiere comprobar si la distribución normal es un modelo de probabilidad razonable para las puntuaciones.

Para resolver este problema con STATGRAPHICS sigue los siguientes pasos:

- Crea un archivo de datos con la variable vocabulario, formada por los datos muestrales.
- Selecciona **Descripción / Distribuciones / Ajuste de Distribuciones (datos no Censurados)**.
- En el cuadro de diálogo que aparece introduce la variable vocabulario en **Datos** y pulsa **Aceptar**.
- Haz clic en el botón **Opciones Tabulares**, activa la casilla **Tests para la Normalidad** y pulsa **Aceptar**.

En el panel de texto **Tests para la Normalidad para …** quedan reflejados los resultados. El p–valor del contraste de Shapiro–Wilks es 0,874182 > 0,05, por lo que se acepta la hipótesis de normalidad. Lo mismo indican los p–valores de los contrastes de asimetría y curtosis.

- Selecciona **Gráficos / Gráficos Exploratorios / Gráfico Probabilístico** para comprobar gráficamente el ajuste a la recta de probabilidad normal.
- En el cuadro de diálogo introduce la variable vocabulario en **Datos** y haz clic en **Aceptar**. La gráfica que se obtiene indica un buen ajuste a la recta, lo que confirma que la variable es normal.

ACTIVIDADES

- **COEFICIENTE DE INTELIGENCIA**

Para investigar el coeficiente de inteligencia medio de cierta población estudiantil, se propuso un test a 400 estudiantes. La media y desviación típica obtenidos fueron 86 y 10,2 respectivamente. Con los datos de esta muestra, determina un intervalo de confianza para μ, con un nivel de confianza del 98%.

- **VARIABLE NORMAL**

Se han tomado 12 valores de una variable física X, que se supone normal, obteniendo los siguientes resultados:

30,2	30,8	29,3	29	30,9	30,8	29,7	28,9	30,5	31,2	31,3	28,5

Construye un intervalo de confianza para la media de la población al 95% de confianza.

- **RECIÉN NACIDOS**

Los pesos de 50 recién nacidos en un hospital maternal, tomados durante un determinado período de tiempo son los siguientes:

3,74	3,09	3,23	2,88	3,01	3,45	3,58	3,40	3,29	3,09
3,63	3,58	3,55	3,61	2,65	3,11	3,88	2,40	2,60	2,88
2,52	3,32	4,03	2,77	2,08	3,22	3,07	3,79	3,19	3,69
3,05	3,87	3,49	3,03	2,93	3,56	3,50	4,70	3,62	2,88
3,23	3,29	3,27	3,11	2,56	3,10	3,42	2,47	3,29	3,58

a) Crea un archivo de datos, con la variable peso, que recoja los datos anteriores.

b) Halla un intervalo de confianza para la media, con un nivel de significación del 5%.

c) Contrasta la hipótesis de que el peso medio de los recién nacidos es 3 kg con un nivel de significación de $\alpha = 0,05$.

- **ATLETISMO**

Se han medido los tiempos (en segundos) que invierten doce atletas en recorrer la prueba de los 100 metros lisos. Los resultados son los siguientes:

12	10	21	13	19	23	14	20	16	13	10	18

a) Construye un archivo de datos, con la variable tiempo, que recoja los datos anteriores.

b) Analiza si es aceptable la hipótesis de normalidad de estos datos.

- **POR CORRESPONDENCIA**

En una prueba de acceso a la Universidad, el porcentaje de admitidos ha sido del 40%. De los que prepararon el examen por correspondencia se escogió una muestra de 200, de los que fueron admitidos 70. ¿Qué conclusión se puede sacar con un nivel de confianza del 95%?.

- **COEFICIENTE INTELECTUAL**

En un estudio realizado sobre la influencia de la educación preescolar sobre el CI se aseguraba que los niños que habían seguido la educación preescolar en un centro público tenían efectos beneficiosos sobre el CI, puesto que se obtuvo como media 115 (a los seis años), siendo la media de la población 105. Dicho estudio ha sido criticado por el sector de la enseñanza privada aludiendo que la media de la población no es 105. Para contrastar dicha hipótesis a un nivel de significación del 5%, se seleccionan 60 niños de seis años y se obtiene una media de 110 y una desviación típica de 5. Se supone que el CI se distribuye normalmente. ¿Es aceptable la afirmación de que la media del CI es 105 ?.