TALLER DE ESTADÍSTICA

6. MODELOS PROBABILÍSTICOS. DISTRIBUCIÓN BINOMIAL Y NORMAL.

MAURICIO CONTRERAS

MODELOS PROBABILÍSTICOS EN LA ESO

Introducción

A partir de los datos muestrales se asigna una probabilidad a cada uno de los datos, usando las frecuencias relativas. La forma de los histogramas correspondientes permite introducir el concepto de modelo probabilístico. El modelo que aparece con más frecuencia es el definido por la curva normal.

Algunos modelos de calculadora gráfica permiten obtener con facilidad áreas bajo la curva normal y valores de la distribución binomial, e incluso otros modelos probabilísticos, como la distribución ji–cuadrado, la T de Student, la distribución de Poisson, etc.

Y también podemos obtener gráficos, histogramas y cuartiles de las distribuciones probabilísticas más importantes con ayuda de la calculadora gráfica.

Existen en el mercado otros modelos físicos de distribuciones. Por ejemplo el binostato o aparato de Galton, que visualiza el perfil de la curva de Gauss.

En esta sesión se analizarán algunos ejemplos de uso en el aula de la calculadora gráfica y de otros materiales, relacionados con el cálculo de probabilidades utilizando las distribuciones estadísticas (especialmente la binominal y normal).

1.– Construcción experimental de modelos

Dados y monedas pueden ser usados para generar modelos funcionales y probabilísticos. Veamos algunos ejemplos.

- **DADOS**

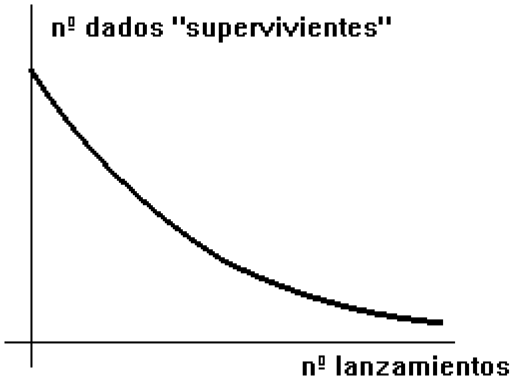
Lanzamos 100 dados y eliminamos todos aquellos que muestren un 6. Repetimos esta operación con los dados que quedan tantas veces como sea necesario para eliminarlos todos. ¿Cuántos dados quedarán después del primer lanzamiento?. ¿Y después del segundo?. ¿Y después del tercero?.

Este ejemplo es un modelo discreto y a escala reducida de la desintegración radiactiva, en la que la cantidad de material que se desintegra –o que queda– en cada momento es proporcional a la cantidad de material que había en el momento anterior.

En cada lanzamiento se puede esperar que desaparezca un sexto y que queden cinco sextos de los dados que había anteriormente (ya que la probabilidad de obtener 6 es 1/6).

La siguiente expresión $P_n = 100 \cdot (5/6)^n$ (que corresponde a una función exponencial) es un modelo que permite conocer aproximadamente el número de dados que quedará después de cada lanzamiento.

También, al hacer el histograma correspondiente a esta situación y ajustarle una curva, se obtiene una distribución estadística exponencial.



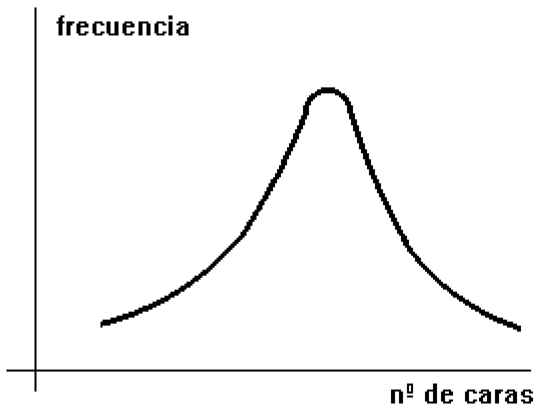
Las situaciones que siguen dan lugar a una distribución binomial

- **MONEDAS Y APARATO DE GALTON**

Lanzamos una moneda al aire 10 veces (o 10 monedas a la vez). ¿Qué número promedio de caras esperas obtener?.

Si cada alumno elabora una tabla con el número de caras que ha obtenido al realizar la experiencia en repetidas ocasiones, se puede llevar a cabo un recuento de los resultados de la clase.

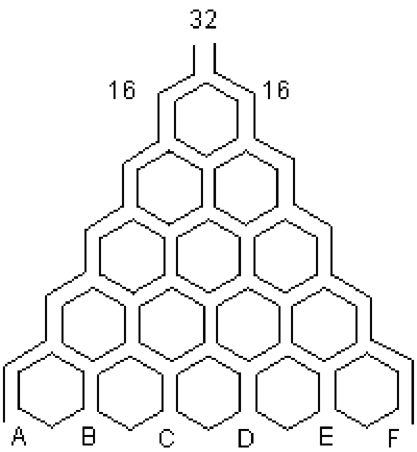
Si se hace un histograma con estos datos y se aproxima por una curva, ésta adoptará el aspecto de una distribución binomial.

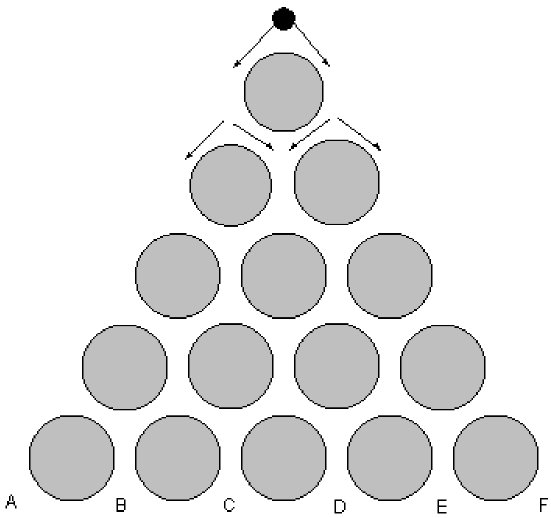


Si en la clase se ha realizado la experiencia un total de doscientas veces, por ejemplo, se puede obtener la misma distribución lanzando, por un aparato de Galton con diez filas de clavos, doscientas bolitas.

El trayecto seguido por cada bolita representa los diez resultados al lanzar la moneda diez veces. Las doscientas bolitas representan otras tantas realizaciones de la experiencia.

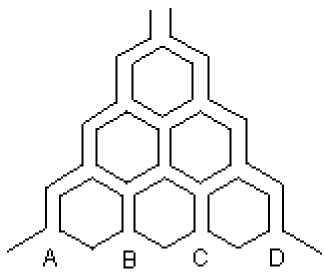
Al final, cada bolita irá a parar a un receptáculo que corresponde al número de caras que se ha obtenido al realizar esa experiencia, formándose con todas las bolitas un perfil que es una aproximación a la distribución binomial.





- **<u>APARATO DE GALTON</u>**

Un aparato de Galton está constituido por un conjunto de pisos con topes. En el primer piso hay 1 tope, en el segundo 2, en el tercero 3, etc. Al dejar caer bolitas desde el primer piso, en cada tope la bolita puede seguir uno cualquiera de los dos caminos posibles. Intenta descubrir cuántos caminos conducen a las posiciones A, B, C y D en un aparato de Galton de tres pisos.



- **<u>MONEDAS</u>**

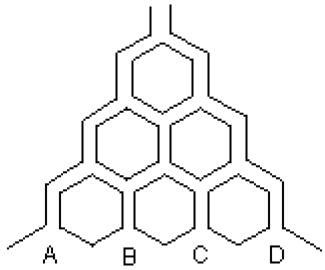
Se lanzan tres monedas al aire. Construye un diagrama de árbol que muestre los posibles resultados.

a) ¿Cuántos resultados conducen a obtener exactamente una cara?.

b) Relaciona el resultado del apartado anterior con los caminos que llevan a la posición C del aparato de Galton del problema anterior.

- **<u>NÚMEROS COMBINATORIOS</u>**

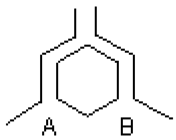
Representaremos el número de caminos que conducen a A, B, C o D por los siguientes números:



$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

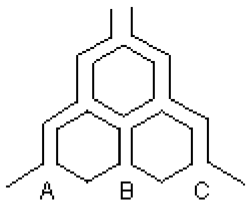
En donde el número de arriba indica el piso, y el de abajo, la posición horizontal, de izquierda a derecha. Por ejemplo, el número de caminos que conducen a C se representa por el número $\binom{3}{2}$, donde 3 indica el número de pisos y 2 la posición horizontal.

Análogamente, si hubiera un único piso:



$$\binom{1}{0} \quad \binom{1}{1}$$

Y si hubiera dos pisos:



$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$

Los números obtenidos se llaman números combinatorios.

Calcula cada uno de estos números, correspondientes a aparatos de Galton de 1, 2 y 3 pisos.

- **<u>TRIÁNGULO DE PASCAL</u>**

Si agrupamos los resultados de la actividad anterior, obtenemos los siguientes valores:

$$\text{1 piso}\rightarrow \quad \binom{1}{0}\ \binom{1}{1} \qquad\qquad 1 \quad 1$$

$$\text{2 pisos}\rightarrow \quad \binom{2}{0}\ \binom{2}{1}\ \binom{2}{2} \qquad\qquad 1 \quad 2 \quad 1$$

$$\text{3 pisos}\rightarrow \quad \binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3} \qquad\qquad 1 \quad 3 \quad 3 \quad 1$$

Esta figura se llama triángulo de Tartaglia o de Pascal.

a) Completa cuatro filas más.

b) Observa detenidamente cada una de las filas y escribe todas las propiedades y regularidades que observes.

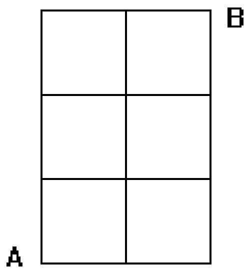
- **<u>LENGUAJE MÁQUINA</u>**

Como sabes, los ordenadores funcionan en lenguaje máquina en base 2, con ceros y unos (cero, apagado; uno, encendido). A esta posibilidad de 0 y 1 se le llama bit, y las combinaciones de éstos, nos dan los distintos caracteres.

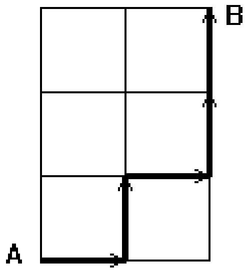
Si combinas 5 bits, ¿cuántos caracteres puedes formar que tengan como el número 10010, 3 ceros y 2 unos?.

- **<u>CAMINOS</u>**

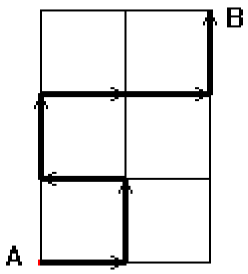
Intentamos ir desde el punto A hasta el B a través de las líneas del dibujo.



Los caminos que recorremos serán lógicos como el siguiente, en el que siempre avanzamos hacia la derecha o hacia arriba.



Y no como el siguiente, que nos haría retroceder.



Indica cuántos caminos lógicos son posibles para ir desde A hasta B.

- **TRIÁNGULO DE TARTAGLIA**

a) Construye el triángulo de Tartaglia hasta el piso octavo.

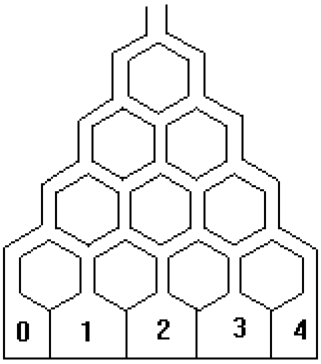
b) Observando este triángulo dí de cuántas formas pueden salir 5 caras y 3 cruces en el lanzamiento de 8 monedas.

c) Si realizas una apuesta con un amigo consistente en lanzar 7 monedas, si tú apuestas 1 euro porque salgan todas caras y él por 4 caras y 3 cruces, ¿cuánto dinero debe apostar tu amigo para que el juego sea justo?.

ACTIVIDADES

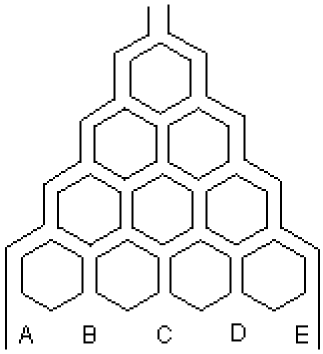
- **RECORRIDOS**

Observa este aparato de Galton con sólo 4 filas de topes. ¿Qué relación hay entre él, el triángulo de Tartaglia de 4 filas y el número de caras obtenidas en el lanzamiento de 4 monedas?. ¿Cuántos recorridos distintos llevan a la casilla 3?. ¿Y a la 4?. Si lanzamos 4 monedas distintas, en cuántos resultados habrá 0, 1, 2, 3 y 4 caras, respectivamente?.



- **BOLITAS**

En un aparato de Galton con 4 filas de topes dejamos caer 400 bolitas. ¿Cuántas, aproximadamente, llegarán a cada casillero?.



- **¿CUÁNTAS FILAS?**

En un aparato de Galton hemos dejado caer un montón de bolitas y hemos contado las que se han depositado en cada casillero:

1	**9**	**46**	**121**	**209**	**251**	**211**	**119**	**45**	**11**	**1**

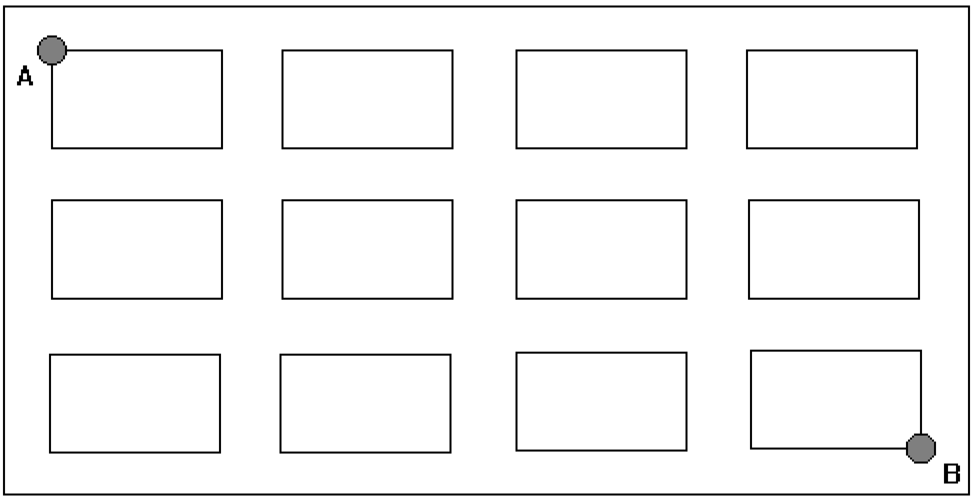
¿Cuántas filas de topes crees que tenía, en este caso, el aparato?.

- **APUESTAS**

Lanzamos 6 monedas y hemos de apostar por el número de caras que van a salir. ¿Por qué número apostarías?. Apostar cuesta 1 euro y, si ganas, te dan 4 euros. Al cabo de 64 partidas, ¿cómo irían tus finanzas, aproximadamente?.

- **CAMINOS**

La siguiente figura muestra un plano de las calles de un barrio. Se supone que las únicas direcciones permitidas son hacia el Este (E) y hacia el Sur (S). Para ir de A a B, ¿cuántos caminos diferentes podemos tomar?.



- **DADOS**

Se tiran 7 dados. ¿Qué crees que es más difícil, no sacar ningún 6 o sacar exactamente un 6?.

- **DARDOS**

En una partida se lanzan 7 dardos. ¿Cuál es el número de sucesos en los que 3 de ellos dan en el blanco?. ¿Cuál es la probabilidad de dar 3 veces en el blanco?.

- **JARRONES**

Una fábrica de jarrones tiene 7 modelos diferentes. ¿Cuántos muestrarios distintos de 3 jarrones podrá hacer?.

- **UNA MONEDA**

¿Cuál es la probabilidad de que salga 3 veces cara en 7 lanzamientos de una moneda?. ¿Y 4 veces cara en 9 lanzamientos?.

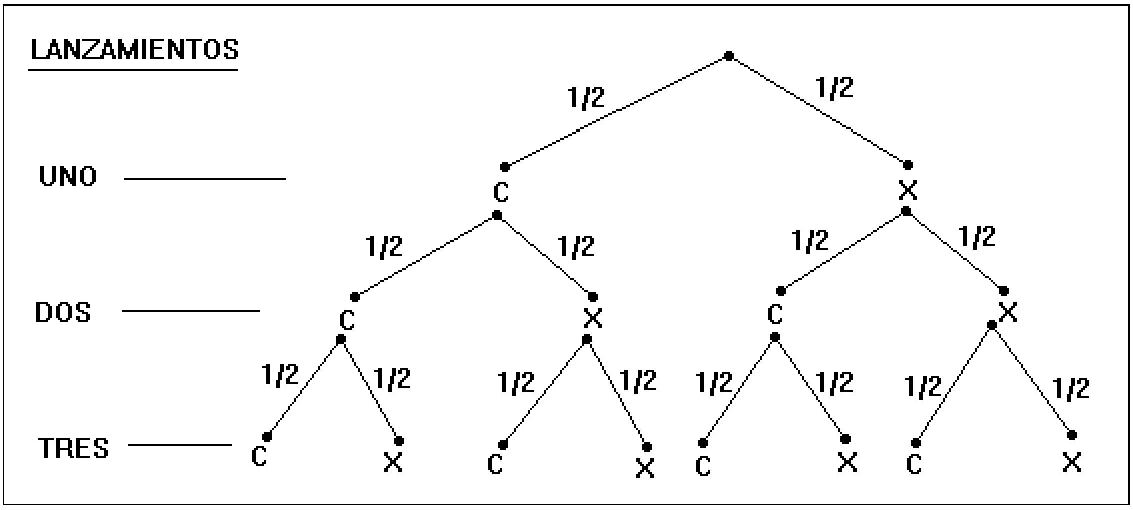
MODELOS PROBABILÍSTICOS EN BACHILLERATO

1.– Construcción del modelo binomial

- **MONEDAS**

Se supone por experiencias previas, que cierta moneda no está sesgada (es decir que ambas caras tienen la misma probabilidad de salir en cada lanzamiento). Si apuestas por la obtención de una cara (que llamamos "éxito"), tendrás pues la misma posibilidad de ganar que tu contrincante. ¿Tendrás la misma probabilidad que tu contrincante si apuestas por un "éxito" al menos en dos lanzamientos?. ¿Cómo tendrían que ser las apuestas en este caso?. ¿Y si apuestas por un "éxito" al menos en tres lanzamientos?. ¿Y en ocho lanzamientos?.

El diagrama que sigue te ayuda a responder a todas las cuestiones planteadas. Cada fracción indica la probabilidad de que ocurra lo que se indica en el final del trazo.



Ahora puedes construir las distribuciones de probabilidad que corresponden a la variable aleatoria "número de éxitos" en uno, dos y tres lanzamientos. Así, en dos lanzamientos, los sucesos cara–cruz y cruz–cara dan ambos 1 "éxito", luego $p(1)=\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2=2\cdot\left(\frac{1}{2}\right)^2$*. Aquí tienes, pues, las distribuciones citadas:*

Un lanzamiento		
Nº de éxitos	*0*	*1*
Probabilidad	$1/2$	$1/2$

Dos lanzamientos			
Nº de éxitos	*0*	*1*	*2*
Probabilidad	$(1/2)^2$	$2\cdot(1/2)^2$	$(1/2)^2$

Tres lanzamientos				
Nº de éxitos	*0*	*1*	*2*	*3*
Probabilidad	$(1/2)^3$	$3\cdot(1/2)^3$	$3\cdot(1/2)^3$	$(1/2)^3$

Por tanto, la probabilidad de ganar en dos lanzamientos, es decir, la probabilidad de obtener al menos una cara es: $p(1)+p(2)=2\cdot(1/2)^2+(1/2)^2=3/4$*, mientras que nuestro contrincante tiene como probabilidad* $1/4$*. Las apuestas deben, pues, realizarse en la proporción* $3/1$.

La probabilidad de ganar en tres lanzamientos es p(1) + p(2) + p(3) y la de perder p(0). Esta última es más sencilla de calcular que la suma anterior, $p(0)=(1/2)^3=1/8$ *y, por tanto,* $p(1)+p(2)+p(3)=1-p(0)=1-1/8=7/8$*. Las apuestas deben en este caso estar en la proporción* $7/1$.

Para responder al caso de ocho lanzamientos, no es necesario construir la tabla correspondiente, ya que basta calcular p(0), es decir, la probabilidad de nuestro contrincante, $p(0)=(1/2)^8=1/256$*, y la probabilidad de ganar será para ti:* $1-p(0)=255/256$*. Las apuestas deben en este caso estar en la proporción* $255/1$.

Si deseamos construir la tabla de probabilidades para el caso de ocho lanzamientos, o muchos más, no hay que pensar en construir diagramas de árbol como el anterior por razones de espacio y tiempo. ¿Habrá algún procedimiento para calcular la probabilidad, p(k), de k éxitos en n lanzamientos?. Si observas las tres tablas anteriores, y dispones en filas los coeficientes que multiplican a las potencias de $1/2$ *, tendrás:*

		1		1			*para un lanzamiento*
	1		2		1		*para dos lanzamientos*
1		3		3		1	*para tres lanzamientos*

Puedes verificar que obtienes el triángulo de Pascal, cuyos elementos son los números combinatorios:

$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

Por tanto, la probabilidad de obtener k éxitos en n pruebas es: $p(k)=\binom{n}{k}\cdot(1/2)^n$ *, siendo*

$$\binom{n}{k}=\frac{n\,(n-1)(n-2)\ldots(n-k+1)}{k\,(k-1)(k-2)\ldots 1}$$

- **DADOS**

Disponemos de un dado cúbico no sesgado (todas las caras tienen la misma probabilidad de aparecer en cada lanzamiento). Apuestas por el seis ("éxito"). ¿Qué probabilidad tienes de ganar en un lanzamiento?. ¿Cómo deben ser las apuestas?. ¿Tendrás la misma probabilidad que tu contrincante si apuestas por un éxito al menos en dos lanzamientos?. ¿A cuántos lanzamientos debes apostar por un éxito al menos, para tener más posibilidades de ganar que tu contrincante?.

Comprueba que, en este caso, la probabilidad de obtener k éxitos en n lanzamientos es: $p(k)=\binom{n}{k}\cdot(1/6)^k\cdot(5/6)^{n-k}$

En los dos ejemplos anteriores estamos frente a una distribución de probabilidad binomial, caracterizada por el hecho de que puede ser interpretada como la repetición de una prueba en la que solo pueden presentarse dos sucesos contrarios, A y B. Si p es la probabilidad de "éxito" en una prueba, y q la de no tenerlo, la probabilidad de tener k éxitos en n pruebas repetidas es:

$$p(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \qquad con \quad p + q = 1$$

La importancia de esta distribución radica en su frecuente aparición en multitud de problemas.

- **DOS CARAS**

Disponemos de dos monedas no sesgadas que van a ser lanzadas simultáneamente. Apuestas por la aparición de dos caras ("éxito") y tu contrincante por la no aparición. ¿Qué probabilidad de ganar tienes en un lanzamiento?. ¿Y en dos, si apuestas por la aparición de la doble cara en uno al menos de los lanzamientos?. ¿A cuántos lanzamientos debes apostar por el suceso anterior, para tener más posibilidades de ganar que tu contrincante?.

- **DOBLE SEIS**

Dispones de dos dados cúbicos no sesgados que van a ser lanzados simultáneamente. Apuestas por la aparición de un doble seis y tu contrincante por la no aparición de dicho doble. ¿Qué probabilidad tienes de ganar en un lanzamiento?. ¿Y en dos lanzamientos, si apuestas por la aparición de un doble seis ("éxito") en uno al menos de los lanzamientos?. ¿A cuántos lanzamientos debes realizar la apuesta para tener más posibilidades de ganar que tu contrincante?. ¿Cómo deben ser las apuestas en cada caso?.

- **NACIMIENTOS**

Repetidas estadísticas realizadas en una clínica maternal han dado origen a la siguiente distribución de probabilidad del sexo de un recién nacido:

Sexo	0	1
Probabilidad	0'485	0'515

en la que 0 indica que el recién nacido es niña y 1 que es niño.

Calcula la distribución de probabilidad, según el sexo, de los próximos 10 nacimientos en dicha clínica. Dibuja el histograma correspondiente. ¿Cuál es la probabilidad de que el número de niñas esté comprendido entre 3 y 7?. ¿Cuántos niños esperan que nazcan?. ¿Cuál es la desviación típica?.

- **DETERGENTES**

El porcentaje de hogares que utilizan una determinada marca de detergente se ha estimado en un 26%. En una muestra de 12 hogares, ¿cuál es la probabilidad de que encontremos un número de usuarios de la marca en cuestión comprendido entre 6 y 9?. Dibuja el histograma de la distribución de probabilidad que encuentres. Calcula la media y la desviación típica del número de hogares.

En las dos últimas actividades puedes comprobar que se cumplen las siguientes propiedades:

NACIMIENTOS	$\sigma = 1'58$ *niños*	$\sqrt{n \cdot p \cdot q} = 1'58$

DETERGENTES	$\sigma = 1'94$ *hogares*	$\sqrt{n \cdot p \cdot q} = 1'94$

En general, se cumple que:

- *la varianza de una variable aleatoria binomial es* $V = n \cdot p \cdot q = n \cdot p \cdot (1-p)$
- *La media de una variable aleatoria binomial de parámetros n y p es* $\sigma = \sqrt{n \cdot p \cdot q}$.

ACTIVIDADES

- **UNA MONEDA**

¿Cuál es la probabilidad de obtener 5 caras lanzando 11 veces una moneda?. ¿Cuántas caras se obtienen por término medio?. ¿Cuál es la desviación típica?.

- **UN DADO**

Se lanza un dado 6 veces. ¿Cuál es la probabilidad de obtener 3 cincos?. Calcula el número medio de cincos obtenidos y la desviación típica.

- **UN TEST**

En un test hay 100 preguntas con cuatro opciones de respuesta, de las que hay que seleccionar una. Si se responde totalmente al azar, ¿cuál es el número medio esperado de respuestas correctas?. ¿Cuál es la desviación típica?.

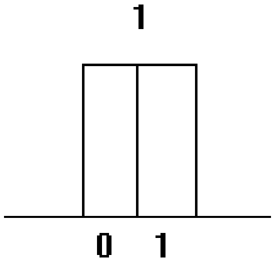
- **MONEDA TRUCADA**

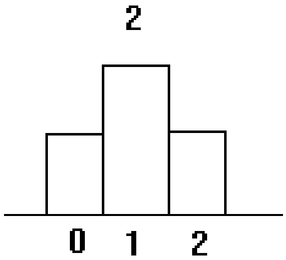
La probabilidad de obtener cara con una moneda trucada es de 0'3 y la lanzamos 100 veces. ¿Cuál es el número esperado de caras?. ¿Y la desviación típica?.

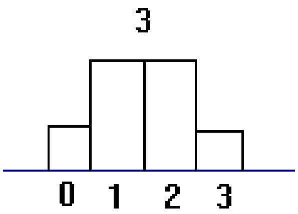
2.–De la binomial a la normal

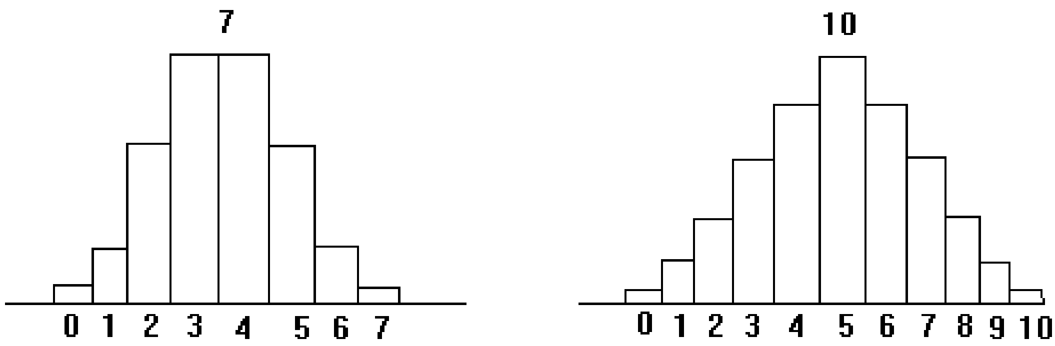
- **APROXIMACIÓN DE HISTOGRAMAS**

Aquí tienes los histogramas correspondientes al lanzamiento de una moneda 1, 2, 3, 4, 7 y 10 veces, para los que se tiene p=q= $1/2$



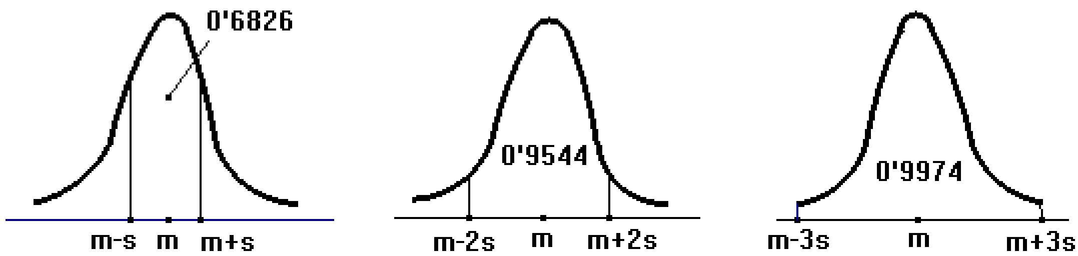




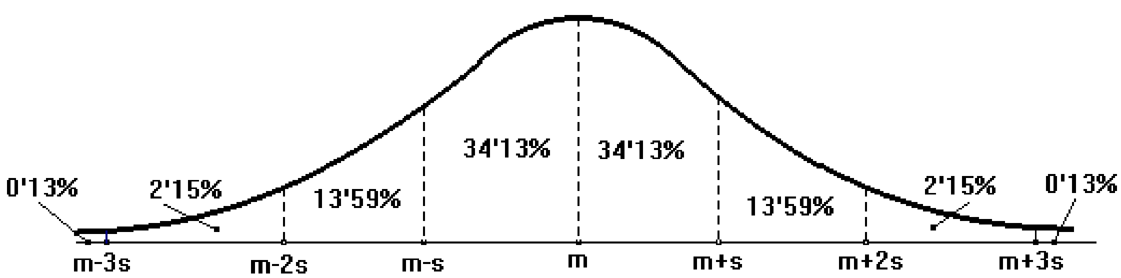


Estos histogramas se pueden aproximar por una curva, llamada CURVA NORMAL, con la condición de que el área bajo la curva coincida con el área del histograma. Podemos considerar dicha curva como un modelo probabilístico continuo. La función correspondiente a esta curva, $x \to f(x)$, es una función de densidad de probabilidad que cumple las siguientes propiedades:

1) $f(x) \geq 0$ *para todo x del dominio de la función.*
2) $\int_{-\infty}^{\infty} f(x) = 1$ *(El área bajo la curva normal es igual a 1).*
3) *El área bajo la curva normal se distribuye del siguiente modo:*



4) *Afinando todavía más, podemos decir que el área bajo la curva normal se distribuye de la siguiente manera:*



a) Las estaturas de 1400 mujeres se distribuyen según una curva normal de media 160'8 y desviación típica 6'4. Calcula los valores m–3s, m–2s, m–s, m+s, m+2s y m+3s. Reparte a las 1400 mujeres, aproximadamente, en esos intervalos.

b) En un estanque de una piscifactoria se ha tomado una muestra de 3000 truchas y se ha medido, en cm, la longitud de las mismas, resultado que se distribuyen según una curva normal de media 26 y desviación típica 7. Calcula los valores m–3s, m–2s, m–s, m+s, m+2s y m+3s, establece los intervalos de longitud correspondientes y reparte las 3000 truchas de la muestra en esos intervalos.

- **FABRICACIÓN**

En el proceso de fabricación de unas piezas intervienen dos máquinas: la máquina A produce un taladro cilíndrico y la máquina B secciona las piezas con un grosor determinado. Ambos procesos son independientes. El diámetro del taladro producido por A, en mm, sigue una curva normal de media 23 y desviación típica 0'5. El grosor producido por B, en mm, viene dado por una curva normal de media 11'5 y desviación típica 0'4.

a) Calcula qué porcentaje de piezas tienen un taladro comprendido entre 22 y 24 mm.

b) Halla el porcentaje de piezas que tienen un grosor comprendido entre 10'3 y 12'7 mm.

- **ESTATURAS**

En una muestra de 1000 personas, la altura media fue de 170 cm, con una desviación típica de 10 cm. Suponiendo que sus alturas siguen una distribución normal, calcula cuántas personas de la muestra tienen:

a) Más de 190 cm.

b) Entre 160 y 190 cm.

c) Menos de 160 cm.

MODELOS PROBABILÍSTICOS CON LA CALCULADORA GRÁFICA

1.– Variables aleatorias discretas

Una **variable aleatoria** es una función $X: E \rightarrow R$, que a cada suceso elemental del espacio muestral E le asocia un número real. Las variables aleatorias pueden ser discretas y continuas.

Una variable aleatoria X es **discreta** si solamente toma una cantidad finita (o infinita numerable) de valores.

Una variable aleatoria X es **continua** si puede tomar todos los valores de un intervalo de números reales.

Ejemplos : 1) Lanzamiento de dos dados. Sea X = número de seises obtenidos en cada lanzamiento. X es una variable discreta.

2) Sea X = peso (o talla) de los jugadores de un equipo de baloncesto. X es una variable continua.

- **Distribución de una variable aleatoria. Función de probabilidad.**

Llamamos **distribución** de una variable aleatoria X a una tabla del tipo

X	Valores	x_1	x_2	...	x_n
P	Probabilidades	$p(X=x_1)$	$p(X=x_2)$	...	$p(X=x_n)$

También se le llama **función de probabilidad** o **función de cuantía**.

La función de probabilidad es la función que a cada valor de la variable le asocia su correspondiente probabilidad.

La gráfica de una función de probabilidad viene dada por un **diagrama de barras** o por un **diagrama de rectángulos.**

- **Función de distribución.**

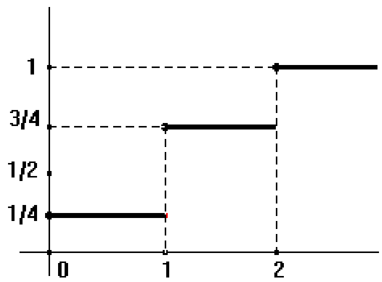
Sea X una variable aleatoria discreta cuyos valores están ordenados de menor a mayor. Llamamos **función de distribución** de la variable X a la función $F(a) = p(X \leq a)$. Es decir, la función de distribución asocia a cada valor de la variable aleatoria la probabilidad acumulada hasta ese valor.

Ejemplo: En el lanzamiento de dos monedas, sea X = nº de caras obtenidas. La función de distribución es:

X	0	1	2
p	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

$$F(x)=\begin{cases} 0, & \text{si } x < 0 \\ 1/4, & \text{si } 0 \leq x < 1 \\ 3/4, & \text{si } 1 \leq x < 2 \\ 1, & \text{si } x \geq 2 \end{cases}$$

y su gráfica es:



La función de distribución cumple las siguientes propiedades:

1) $0 \leq F(x) \leq 1$
2) Es una función escalonada.
3) F(x) = 0, si x es menor que el menor valor de la variable.
4) F(x) = 1, si x es mayor que el mayor valor de la variable.
5) F(x) es creciente.

La **media**, **esperanza** (o **valor esperado**) de una variable aleatoria X es igual a la suma de los productos de los valores de la variable por las probabilidades respectivas.

$$\mu = \sum_{i=1}^{n} x_i \cdot p_i$$

Así mismo, la **varianza** de una variable aleatoria X es:

$$V=\sum_{i=1}^{n}(x_i-\mu)^2\cdot p_i \qquad V=\sum_{i=1}^{n}x_i^2\cdot p_i-\mu^2$$

y la **desviación típica** se calcula mediante:

$$\sigma=\sqrt{\sum_{i=1}^{n}(x_i-\mu)^2 p_i} \qquad \sigma=\sqrt{\sum_{i=1}^{n}x_i^2\cdot p_i-\mu^2}$$

ACTIVIDADES

- **DISTRIBUCIÓN**

La distribución de probabilidad de una variable aleatoria discreta viene dada por la siguiente tabla:

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

a) Halla m para que se trate de una función de probabilidad.

b) Calcula y representa gráficamente su función de distribución.

c) Halla $p(x \le 4)$ y $p(2 \le x \le 4)$.

- **JUEGO DE CARTAS**

En un juego, una persona recibe 15 céntimos cuando saca una sota o un caballo y recibe 5 céntimos si saca un rey o un as de una baraja española con 40 cartas. Si saca cualquier otra carta tiene que pagar 4 céntimos. ¿Cuál es la ganancia esperada para una persona que entra en el juego?. ¿Cuál es la varianza? ¿Y la desviación típica?.

- **VENTAS**

Un director de ventas elabora la siguiente tabla de probabilidades de distintos niveles de ventas de un nuevo producto:

Ventas (unidades)	50	100	150	200	250	300
Probabilidad	0'10	0'30	0'30	0'15	0'10	0'15

Calcula las ventas esperadas, la varianza y la desviación típica.

2.– Distribución binomial

Hacemos n repeticiones independientes de una prueba con dos resultados posibles que son sucesos contrarios A y $\overline{A}$ (éxito y fracaso). Sean p y q las probabilidades de éxito y fracaso en una prueba. Sea X la variable aleatoria número de éxitos. La probabilidad de obtener k éxitos en las n pruebas es:

$$p(k)=\binom{n}{k}\, p^k\, q^{n-k}, \quad \text{con } p+q=1$$

Se dice que la variable X sigue una distribución binomial de parámetros n y p. Se expresa así: X ≈ B(n, p). La media, varianza y desviación típica de una variable aleatoria binomial, X, se obtienen por medio de las fórmulas:

$$\overline{X} = n \cdot p \qquad V = n \cdot p \cdot q \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Ejemplo: Un examen tipo test

Un estudiante hace un examen tipo test de elección múltiple compuesto por 10 preguntas con 5 respuestas cada una. Si no ha estudiado para el examen y señala aleatoriamente la respuesta, ¿qué resultado puede obtener?. ¿Cuál es la probabilidad de que el estudiante consiga exactamente 6 respuestas correctas?. ¿Cuál es la probabilidad de que el estudiante consiga al menos 6 respuestas correctas?.

Si X es el número de aciertos, X sigue una distribución binomial de parámetros n=10 y $p=1/5=0.2$ Es decir, X~B(10, 0.2). Para ver los posibles resultados, simulamos una binomial de parámetros n=10 y p=0.2. Para ello utilizamos la función **randBin(** de la calculadora gráfica TI–83, cuya sintaxis es:

randBin(número de pruebas, probabilidad de éxito, número de simulaciones)

Suponiendo 50 simulaciones, introducimos la función randBin(10, 0.2, 50) en la lista L_1 pulsando [MATH] [◄] [7] 10 , 0.2 , 50) [ENTER] [→] L_1

A continuación dibujamos el histograma correspondiente a la lista L_1 obteniendo dos aciertos como resultado más frecuente.

Para hallar la probabilidad de que el estudiante acierte 6 respuestas, utilizamos el menú DISTR de la TI–83. Este menú se visualiza en pantalla pulsando [2^{nd}] [VARS] La función de cuantía binomial es **binompdf(,** cuya sintaxis es la siguiente:

binompdf(número de pruebas, probabilidad de éxito, número de éxitos).

En caso de que no se indique el tercer parámetro, esta función nos dará una lista con todas las probabilidades posibles.

Así, la probabilidad de tener 6 aciertos es: binompdf(10, 0.2, 6) 0.00550524

Para hallar la probabilidad de obtener al menos 6 aciertos, utilizamos la función de distribución binomial, **binomcdf(**, cuya sintaxis es la siguiente:

binomcdf(número de pruebas, probabilidad de éxito, número de éxitos).

Si el tercer parámetro no se indica esta función nos da una lista con todas las probabilidades acumuladas posibles.

El suceso contrario de tener al menos 6 aciertos es tener, como máximo, 5 aciertos. La probabilidad de este suceso se obtiene con la función binomcdf(10, 0.2, 5). Así, la probabilidad de tener al menos 6 aciertos es: 1−binomcdf(10, 0.2, 5), que da como resultado 0.0063693824.

ACTIVIDADES

- **CONTROL DE CALIDAD**

Al controlar la cantidad de un producto envasado, se eligen tres al azar de una caja que contiene 50 envases. Por término medio, sabemos que en cada caja hay 5 cuya calidad es deficiente.

a) Determina la probabilidad de que entre los tres no haya ninguno, uno o dos deficientes.

b) Si el primero resulta deficiente, ¿cuál es la probabilidad de que entre los tres haya uno o dos deficientes?.

- **LICENCIATURA**

La probabilidad de que un estudiante que ingresa en la Universidad se licencie en 5 años, es de 0,4. Se eligen al azar 10 estudiantes. Halla: a) Probabilidad de que ninguno se licencie en 5 años. b) Probabilidad de que al menos uno se licencie en 5 años. c) Probabilidad de que todos se licencien en 5 años.

- **FILOSOFÍA**

En un cierto instituto, el curso pasado aprobaron la Filosofía el 80% de los alumnos de Bachillerato. ¿Cuál es la probabilidad de que, de un grupo de 8 alumnos elegidos al azar, sólo dos hubieran suspendido la Filosofía?.

- **BOMBILLAS**

En una fábrica de bombillas se sabe que el 2% son defectuosas. Si se empaquetan en cajas de 20 unidades, calcula la probabilidad de que en una caja: a) No haya ninguna bombilla defectuosa. b) Sólo haya una defectuosa. c) Haya más de tres bombillas defectuosas.

- **CARA Y CRUZ**

Si se lanza una moneda 6 veces, ¿cuál es la probabilidad de que el resultado "cruz" no salga más veces que el resultado "cara"?.

- **LOTERÍA**

El 11% de los billetes de lotería reciben algún tipo de premio, aunque sea el reintegro. En una familia juegan a 46 números. ¿Cuál es la probabilidad de que obtengan premio, al menos, 10 de ellos?.

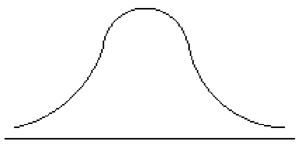
- **APUESTAS**

El jugador A apuesta que al lanzar un dado obtendrá por los menos dos seises en seis tiradas. El jugador B apuesta que al lanzar una moneda diez veces obtendrá por lo menos siete veces cara.

¿Qué probabilidad de ganar tiene cada uno?. ¿Cuál es la probabilidad de que los dos ganen sus respectivas apuestas?. ¿Cuál es la probabilidad de que solamente gane uno de los dos?. ¿Cuál es la probabilidad de que ninguno de ellos gane?.

3.– Distribución normal

Muchos histogramas pueden ajustarse por una curva que tiene la forma de una campana invertida:



Esta curva es conocida como **CURVA NORMAL** o **CURVA DE GAUSS**. Es simétrica respecto de la vertical que pasa por la media y presenta un máximo para dicho valor medio, $\overline{x}$.

Además, la desviación típica σ es la distancia del eje de simetría a cualquiera de los dos **puntos de inflexión** de la curva normal. En el **punto de inflexión**, la campana cambia de mirar hacia abajo a mirar hacia arriba (o viceversa).

Existen multitud de fenómenos de azar o procesos aleatorios que pueden representarse por la curva normal (por eso, precisamente, se llama **normal**). De manera que, para esos fenómenos, la campana de Gauss cumple el mismo papel que el histograma.

Para calcular probabilidades a partir de un histograma, entre dos valores dados, basta sumar áreas de rectángulos. Para calcular probabilidades a partir de la curva normal, hay que recurrir al cálculo de primitivas. La función que representa a la curva normal, llamada **función de densidad normal** viene dada por la fórmula:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2}$$

La dificultad reside en que esta función no tiene primitivas expresables mediante funciones elementales. Por esta razón, se han utilizado métodos de aproximación numérica (método de los rectángulos, método de Simpson, etc) para obtener valores de dicha función. Con la calculadora gráfica puedes obtener directamente estos valores, conocidos los parámetros media, $\overline{x}$ y desviación típica, σ.

También puedes obtener con la calculadora TI–83 los valores que puede tomar el área bajo la **curva normal**, de media $\overline{x} = 0$ y desviación típica $\sigma = 1$, cuya fórmula es:

$$F(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

La **función de distribución normal** es la función cuyo valor en cada punto **a** es

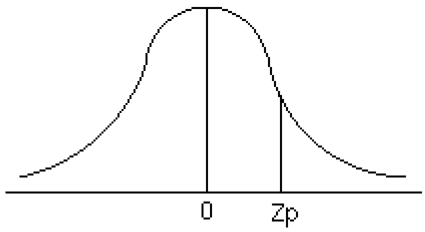
F(a) = p(x < a) = p(x ≤ a)

y verifica la propiedad: **p(a < x < b) = F(b) – F(a)**

Si X es una variable aleatoria continua que sigue una distribución normal de media $\overline{x}$ y desviación típica σ, escribimos: X~N($\overline{x}$, σ)

Si Z es una variable aleatoria continua que sigue una distribución normal de media 0 y desviación típica 1, decimos que es una variable normal típica o estándar y la representamos así: Z ~N(0, 1).

La probabilidad de que la variable Z tome valores menores o iguales que uno z_p fijado de antemano, p(Z≤ z_p), es el área bajo la curva normal hasta z_p. La probabilidad de que la variable Z tome valores comprendidos entre 0 y z_p, p(0 ≤ Z ≤ z_p), es el área bajo la curva normal desde z_p = 0 hasta z_p.



La función **normalcdf(** calcula la probabilidad de distribución normal entre el límite inferior y el límite superior para la media μ y desviación típica σ especificadas. Los valores predeterminados son μ = 0 y σ = 1. La sintaxis de esta función es la siguiente:

normalcdf(límite inferior, límite superior, μ, σ)

Así: normalcdf(−1E99, 0.5) da como resultado 0.6914624678. Por tanto: p(Z≤0'50)=0'6915. De la misma forma, normalcdf(−1E99, 1.17) da como resultado 0.8789994587. Por tanto: p(Z≤1'17)=0'8790.

Recuerda que la suma de las áreas de los rectángulos que componen un histograma es igual a la unidad. Puesto que la curva normal es una aproximación del histograma, esta propiedad también la verifica la campana de Gauss, es decir:

EL ÁREA BAJO LA CURVA NORMAL EN LA TOTALIDAD DE SU DOMINIO VALE 1.

$$\int_{-\infty}^{+\infty} F(z) = 1$$

Utilizando esta propiedad y el hecho de que la curva es simétrica respecto del eje de ordenadas, tenemos:

p(Z≤ −1) = p(Z ≥ 1) = 1 – p(Z ≤ 1) = 1 – 0'8413 = 0'1587

Así pues, se cumple que p(Z≤−1)=0'1587=15'87 %. Pero este resultado también se podría haber obtenido directamente, mediante la función:

normalcdf(−1E99, −1) que da como resultado 0.1586552596.

- **EJEMPLOS**

a) Utilizando las propiedades de la curva de Gauss, calcula a partir de la tabla de la función de distribución normal, las siguientes probabilidades:

p(Z≤2) ; p(Z>2) ; p(Z>−2) ; p(Z<−2) ; p(0'5<Z<1'5) ; p(−0'5≤Z≤0'5)

¿Puedes usar este procedimiento para calcular p(Z=1'5)?.

En general, podemos hacer la aproximación: **p(Z = a) = p(a – 0'5 < Z ≤ a + 0'5)**

En nuestro caso, p(Z=1'5) = p(1<Z<2) = normalcdf(1,2) = 0.1359051975 ≈ 0.136.

La relación que existe entre la curva normal tipificada (media 0 y desviación típica 1) y cualquier curva normal (media distinta de 0 y desviación típica distinta de 1) es la siguiente:

Si X es la variable de un fenómeno aleatorio representado por una curva normal de media $\overline{x}$ y desviación típica σ, entonces la variable

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

es una **normal tipificada**, es decir, de media 0 y desviación típica 1. Es decir:

$$\text{Si } X \sim N(\overline{X},\sigma) \text{ entonces } Z = \frac{X-\overline{X}}{\sigma} \sim N(0, 1)$$

b) **Una variable X está representada por una curva normal de media $\overline{X}=2$ y desviación típica σ = 0’5. ¿Qué posibilidades tenemos de que dicha variable X tome valores comprendidos entre 1’8 y 2’2 ?. ¿Y de que la variable X tome valores inferiores a 1’8 o superiores a 2’2 ?.**

- **COCIENTE INTELECTUAL**

Los cocientes intelectuales de una población de individuos siguen una distribución normal de media 100 y desviación típica 15. a) Utiliza la calculadora gráfica para obtener una muestra aleatoria de tamaño 100. b) Utiliza la calculadora gráfica para representar diversas curvas de Gauss. c) ¿Cuál es la probabilidad de que una persona elegida al azar tenga un cociente intelectual comprendido entre 90 y 115 ?.

La función **randNorm(** de la calculadora TI–83 genera un número real aleatorio de una distribución normal especificada. Su sintaxis es:

randNorm(media, desviación típica, número de pruebas)

Para obtener una lista con una muestra de tamaño 100 utilizaremos esta función, pulsando:
MATH ◄ [6] 100 , 15 , 100) → L_1

Comprueba que la media y la desviación típica de la lista L_1 son las esperadas.

La función **normalpdf(** situada en el menú DISTR de la TI–83 permite obtener la función de densidad de una variable normal. Su sintaxis es la siguiente:

normalpdf(valor, media, desviación típica)

Si no se indican la media y la desviación típica, se sobreentiende 0 y 1, es decir una normal tipificada.

Podemos dibujar las gráficas de diversas curvas normales, utilizando esta función. Así, en el menú Y= introducimos las funciones Y_1 = normalpdf(X), Y_2 = normalpdf(X, 0, 2), Y_3 = normalpdf(X, 2, 0.5). Extrae conclusiones.

La función **normalcdf(** situada en el menú DISTR de la TI–83 calcula la distribución de probabilidad normal acumulada entre la cota inferior y la cota superior. Su sintaxis es la siguiente:

normalcdf(cota inferior, cota superior, media, desviación típica)

Si no se especifican la media y desviación típica, se sobreentiende 0 y 1, es decir, una normal tipificada.

Para hallar la probabilidad de que una persona elegida al azar tenga un cociente intelectual entre 90 y 115 utilizamos la función **normalcdf(90, 115, 100, 15),** pulsando [2nd] [VARS] eligiendo la función en la lista de opciones, introduciendo los parámetros y pulsando ENTER.

También podemos calcular cuantiles correspondientes a la distribución normal estándard, es decir valores k de la variable para los que la probabilidad $p(Z \le k)$ toma un determinado valor fijado de antemano. Esto se puede hacer mediante la función **invNorm(** de la TI–83, cuya sintaxis es la siguiente:

invNorm(probabilidad)

Por ejemplo, si queremos hallar k con la condición de que $p(Z \le k)=0.95$, utilizamos la función invNorm(0,95), para lo que hay que pulsar [2nd] DISTR [3] 0.95) ENTER. En pantalla aparece el resultado, k = 1.644853626.

Si queremos hallar el cociente intelectual que supera al 95 % de la población, suponiendo que la distribución es normal de media 100 y desviación típica 15, basta tener en cuenta la tipificación. Así:

$Z = \dfrac{X-100}{15} \approx N(0,1)$. Como el valor k=1.644853626 es tal que $p(Z \le k)=0.95$, el valor del cociente intelectual X buscado debe cumplir: $\dfrac{X-100}{15} = 1.64485$. Por tanto, X = 1.64485 × 15 + 100 = 124.6728 ≈ 125.

En general, si $X \approx N(\mu, \sigma)$, el valor del cuantil k, tal que $p(X \le k)=\alpha$, se obtiene así:

$$k = \text{invNorm}(\alpha) \times \sigma + \mu$$

ACTIVIDADES

- **VENTAS**

La media de ventas diarias de un vendedor de unos grandes almacenes es de 950 euros y la desviación típica es 200 euros. Suponiendo que la distribución de ventas es normal, ¿cuál es la probabilidad de vender más de 1250 euros en un día?.

- **CULTURA GENERAL**

Tras realizar un test de cultura general entre los habitantes de cierta población, se observa que las puntuaciones siguen una distribución normal, de media 68 y desviación típica 18. Se desea clasificar a los habitantes en tres grupos (de baja cultura general, de cultura general aceptable, de cultura general excelente), de manera que el primer grupo abarque un 20% de la población, el segundo un 65% y el tercero el 15% restante. ¿Cuáles son las puntuaciones que marcan el paso de un grupo a otro?.

- **ELECTRODOMÉSTICO**

Se sabe que la vida media de un electrodoméstico es de 10 años con una desviación típica de 0,7 años. Suponiendo que dicha vida media sigue una distribución normal, calcula: a) La probabilidad de que el electrodoméstico dure más de 9 años. b) La probabilidad de que dure entre 9 y 11 años.

- **BATAS**

Una gran empresa debe reponer las batas de sus 1000 operarios. Se sabe que la talla media es de 170 cm, con una desviación típica de 3 cm. Las batas se confeccionan en tres tallas válidas para estaturas entre 155 y 165 cm, 165 y 175 cm y, finalmente, entre 175 y 185 cm. ¿Cuántas batas de cada talla ha de adquirir?.

- **EDADES**

En la ciudad A, la edad de sus 400000 habitantes sigue una distribución normal de media 41 años y desviación típica 12 años. En la ciudad B, con el doble de habitantes, la edad se distribuye normalmente con media 47 años y desviación típica 8 años. ¿En cuál de las dos ciudades es mayor la proporción de habitantes mayores de 65 años?. ¿Cuál de las dos ciudades tiene mayor número de habitantes con edad superior a 65 años?.

- **FLUIDEZ VERBAL**

Se ha aplicado un test de fluidez verbal a 500 alumnos de primero de ESO de un centro de secundaria. Se supone que las puntuaciones obtenidas se distribuyen según una normal de media 80 y desviación típica 12. Se pide:

a) ¿Qué puntuación separa el 25% de los alumnos con menos fluidez verbal?.

b) ¿A partir de qué puntuación se encuentra el 25% de los alumnos con mayor fluidez verbal?.

- **UN PROBLEMA DE ALTURAS**

Un país está habitado por dos grupos étnicos, M y N, que se encuentran en las proporciones 75% y 25%, respectivamente. Se sabe que la talla de los individuos adultos varones es $N(\mu, \sigma)$, con $\mu = 170$ y $\sigma = 5$ cm para el grupo M, $\mu = 175$ y $\sigma = 5$ cm para el grupo N. Se conviene en que un individuo es alto si tu talla es superior a 180 cm. Se pide:

a) Porcentaje de individuos altos en el grupo M.

b) Porcentaje de altos en el grupo N.

c) Porcentaje de altos en el país.

d) Si un individuo es alto, ¿cuál es la probabilidad de que pertenezca al grupo N?.

4.– Aproximación de la distribución binomial por la normal

Toda **distribución binomial** puede ajustarse por una **distribución normal** con la misma media, $\overline{X} = n\,p$ y la misma desviación típica σ = $\sqrt{n\,p\,q}$.

El sentido de la palabra “ajustar” es que las áreas encerradas por la curva normal en cada intervalo y el área de los rectángulos del histograma que corresponden al mismo intervalo son “casi” iguales.

En general se demuestra que el ajuste es bueno siempre que se cumplan las condiciones: $n\,p\,q \geq 9,\ n\,p \geq 5,\ n\,q \geq 5$.

- **VACUNA**

Se conoce, por estudios previos, que la proporción de reses que enfermarán después de suministrarles una determinada vacuna es del 2%. Una granja tiene 600 reses que son vacunadas.

a) Determina el número esperado de reses que no enfermarán.

b) Halla la probabilidad de que el número de reses que enferman sea, como máximo, 20.

c) Determina la probabilidad de que el número de reses que no enferman sea, como mínimo, 590.

Sea X el número de reses que no enfermarán. X sigue una distribución binomial de parámetros n=600 y p = 0’98. Es decir, X ≈ B(600, 0’98).

a) El número medio de reses que no enfermarán es $\overline{X} = n \cdot p = 600 \cdot 0'98 = 588$.

b) La probabilidad de que el número de reses que enferman sea 20 como máximo es:

p(600–X≤20) = p(X≥600–20) = p(X≥580) = p(X=580)+p(X=581)+p(X=582)+…+p(X=600) =
= 1–binomcdf(600, 0’98, 579) = 0.9891689919 ≈ 0.99.

Además, la media es $\overline{X} = np = 600 \cdot 0'98 = 588$, la varianza es V=n·p·q=600·0’98·0’02=11’76 y la desviación típica es: σ=$\sqrt{n \cdot p \cdot q} = \sqrt{11'76} = 3'42928564$ ≈3’4. Como se cumple que n·p·q≥9, n·p≥5 y n·q≥5, la variable binomial X se puede aproximar por una variable normal Y de media 588 y desviación típica 3’4. Es decir: Y≈N(588, 3’4). Entonces la probabilidad pedida es:

p(X ≥ 580) = p(Y ≥ 579’5) = normalcdf(579’5, 1E99, 588, 3.4) = 0.9937903201 ≈ 0.99.

c) La probabilidad de que como mínimo no enfermen 590 reses es:

p(X ≥ 590) = p(Y ≥ 589.5) = normalcdf(589.5, 1E99, 588, 3.4) = 0.3295426481 ≈ 0.33

ACTIVIDADES

- **VACACIONES EN LA PLAYA**

El 90% de los miembros de un club pasan sus vacaciones en la playa. ¿Cuál es la probabilidad de que, de un grupo de 60 miembros del club, 50 o menos vayan a la playa a pasar sus vacaciones?.

- **BOMBILLAS**

Una fábrica de bombillas ha comprobado que el 10% de la producción tiene algún defecto. Las bombillas se empaquetan en cajas de 100.

a) Calcula la probabilidad de que una caja contenga más de 5 bombillas defectuosas.

b) Calcula la probabilidad de que el número de bombillas defectuosas esté comprendido entre 7 y 13.

c) ¿Cuál es el número esperado de bombillas defectuosas en cada caja por término medio?.

- **EL DADO**

Se lanza un dado 600 veces. Halla la probabilidad de obtener un seis, más de 90 veces y menos de 110.

- **PROCESO DE FABRICACIÓN**

El 5% de las piezas obtenidas en cierto proceso de fabricación resultan defectuosas. ¿Cuál es la probabilidad de que en 1000 piezas fabricadas resulten defectuosas menos de 35?.

- **PRUEBA DEPORTIVA**

Se sabe, después de una larga serie de observaciones, que sólo superan una cierta prueba deportiva el 10% de los atletas presentados. De un conjunto de 500 aspirantes, ¿cuál es la probabilidad de que superen la prueba más de 80?.

- **BEBÉS**

La probabilidad de que un bebé sea chico es 0'52. Si en una clínica han nacido 184 bebés, ¿cuál es la probabilidad de que haya 100 chicos o más?.

5.– Aproximación de la binomial a la normal por simulaciones

- **APROXIMACIÓN DE LA BINOMIAL POR LA CURVA NORMAL**

Toda **distribución binomial** puede ajustarse por una **distribución normal** con la misma media, $\overline{X} = n\,p$ y la misma desviación típica $\sigma = \sqrt{n\,p\,q}$, en el sentido de que las áreas encerradas por la curva normal en cada intervalo y el área de los rectángulos del histograma que corresponden al mismo intervalo son "casi" iguales. El ajuste es especialmente bueno si se cumple que $n\,p\,q \geq 9,\ n\,p \geq 5,\ n\,q \geq 5$.

Vamos a comprobar, a través de diversos ejemplos, que cuando n toma valores suficientemente grandes, el histograma de la distribución binomial se aproxima a la curva normal de la misma media y desviación típica.

- **Lanzamiento de una moneda**

Efectuamos n lanzamientos de una moneda. Llamamos X = nº de caras obtenidas. La variable X sigue una distribución binomial de parámetros n y p = 1/2, es decir: $X \approx B(n, 1/2)$. En este caso, la probabilidad de obtener k caras en n lanzamientos viene dada por la expresión:

$$p(X = k) = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^n$$

Con la calculadora gráfica vamos a simular n=20, n=80 y n=100 lanzamientos de la moneda. En cada caso haremos 100 simulaciones, construiremos el histograma de la variable X=nº de caras obtenidas y estudiaremos si se aproxima o no a una normal.

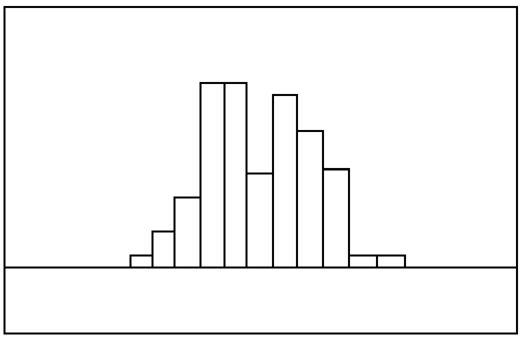
- **20 lanzamientos**

Pulsa [STAT] [ENTER] para iniciar el editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_1 y pulsa [CLEAR] para borrar su contenido. Pulsa [MATH] [◄] 7 para seleccionar el comando **7: randBin(**. Pulsa 20 [,] .5 [,] 100 [ )]. De esta forma hemos definido la lista **L_1=randBin(20, .5, 100)**. Al pulsar [ENTER] se genera la lista L_1.

Pulsa [2nd] [STATPLOT] [ENTER] para definir el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist:	L_1
Freq:	1

Pulsa [WINDOW] y define los parámetros de visualización: Xmin=0, Xmax=20, Xscl=1. Pulsa [GRAPH] y obtendrás el histograma correspondiente.



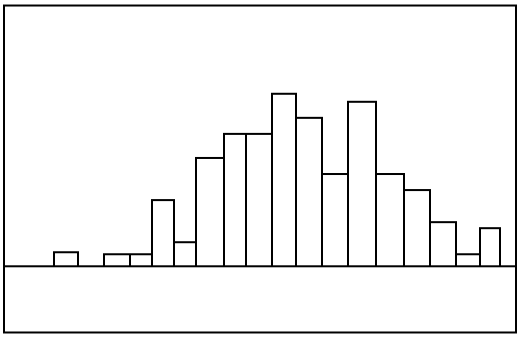
- **80 lanzamientos**

Pulsa [STAT] [ENTER] para iniciar el editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_1 y pulsa [CLEAR] para borrar su contenido. Pulsa [MATH] [◄] 7 para seleccionar el comando **7: randBin(**. Pulsa 80 [,] .5 [,] 100 [ )]. De esta forma hemos definido la lista **L_1=randBin(80, .5, 100)**. Al pulsar [ENTER] se genera la lista L_1.

Pulsa [2nd] [STATPLOT] [ENTER] para definir el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist:	L_1
Freq:	1

Pulsa [WINDOW] y define el parámetro de visualización: Xscl=1. Pulsa [GRAPH] y obtendrás el histograma correspondiente.



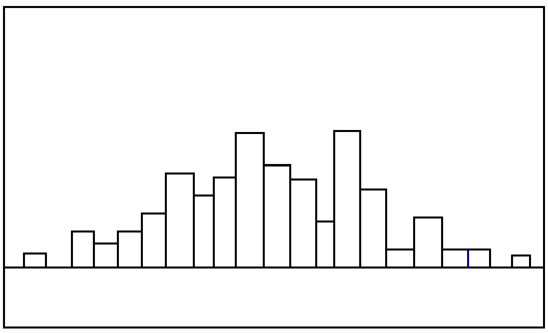
- **100 lanzamientos**

Pulsa [STAT] [ENTER] para iniciar el editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_1 y pulsa [CLEAR] para borrar su contenido. Pulsa [MATH] [◄] 7 para seleccionar el comando **7: randBin(**. Pulsa 100 [,] .5 [,] 100 [ )]. De esta forma hemos definido la lista **L_1=randBin(100, .5, 100)**. Al pulsar [ENTER] se genera la lista L_1.

Pulsa [2nd] [STATPLOT] [ENTER] para definir el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist:	L_1
Freq:	1

Pulsa [WINDOW] y define el parámetro de visualización: Xscl=1. Pulsa [GRAPH] y obtendrás el histograma correspondiente.



- **Comparación con la curva normal**

Las curvas normales asociadas a cada una de las simulaciones anteriores son:

a) Para 20 lanzamientos, la media es $\overline{X} = n\,p = 20 \cdot 0'5 = 10$ y la desviación típica es $\sigma = \sqrt{n\,p\,q} = \sqrt{20 \cdot 0'5 \cdot 0'5} = \sqrt{5}$. Por tanto, la variable X=nº de caras se puede aproximar por una curva normal N(20, $\sqrt{5}$).

b) Para 80 lanzamientos, la media es $\overline{X} = n\,p = 80 \cdot 0'5 = 40$ y la desviación típica es $\sigma = \sqrt{n\,p\,q} = \sqrt{80 \cdot 0'5 \cdot 0'5} = \sqrt{20}$. Por tanto, la variable X=nº de caras se puede aproximar por una curva normal N(40, $\sqrt{20}$).

c) Para 100 lanzamientos, la media es $\overline{X} = n\,p = 100 \cdot 0'5 = 50$ y la desviación típica es $\sigma = \sqrt{n\,p\,q} = \sqrt{100 \cdot 0'5 \cdot 0'5} = \sqrt{25} = 5$.

Vamos a dibujar las gráficas de las tres curvas normales, **Y_1=normalpdf(X, 10, $\sqrt{5}$)**, **Y_2=normalpdf(X, 40, $\sqrt{20}$)**, **Y_3=normalpdf(X, 50, 5)**. Para ello sigue los siguientes pasos:

Pulsa [Y=] [▲] [ENTER] para desactivar el gráfico estadístico Plot1. Pulsa [▼] para situar el cursor en Y_1. Pulsa [2nd] [DISTR] [ENTER] para seleccionar el comando **1: normalpdf(**. Pulsa [X,T,θ,n] [,] 10 [,] [2nd] [$\sqrt{\ }$] 5 [ )] [ )] para definir la función **Y_1=normalpdf(X, 10, $\sqrt{5}$).**

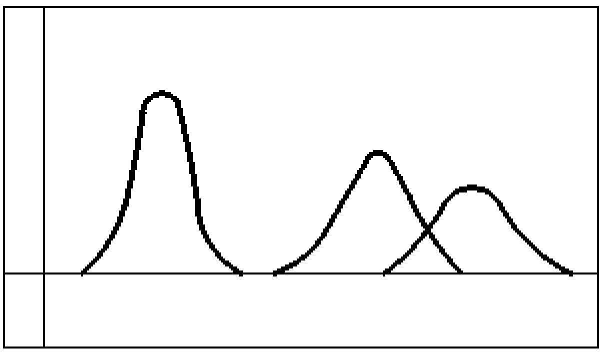
Pulsa [ENTER] para pasar a la función Y_2.

Pulsa [2nd] [DISTR] [ENTER] para seleccionar el comando **1: normalpdf(**. Pulsa [X,T,θ,n] [,] 40 [,] [2nd] [$\sqrt{\ }$] 20 [ )] [ )] para definir la función **Y_1=normalpdf(X, 40, $\sqrt{20}$).**

Pulsa [ENTER] para pasar a la función Y_3.

Pulsa [2nd] [DISTR] [ENTER] para seleccionar el comando **1: normalpdf(**. Pulsa [X,T,θ,n] [,] 50 [,] 5 [ )] [ )] para definir la función **Y_1=normalpdf(X, 50, 5)**.

Pulsa [WINDOW] e introduce los siguientes valores de los parámetros de ventana: Xmin=0, Xmax=70, Ymin=–0'05, Ymax=0'3. Pulsa [GRAPH] para obtener la representación gráfica.



Observa que las curvas normales obtenidas tienen formas parecidas a los histogramas, pero solo de forma aproximada. Esto es debido al número de simulaciones realizadas (en este caso, 100 simulaciones). En la medida en que el número de simulaciones aumenta (y aumente también el valor de n), la forma de los histogramas se asemeja más a las curvas normales respectivas. De hecho, si en lugar de utilizar simulaciones, construyésemos los histogramas correspondientes a la distribución binomial teórica, la forma sería casi idéntica a la de la curva normal.

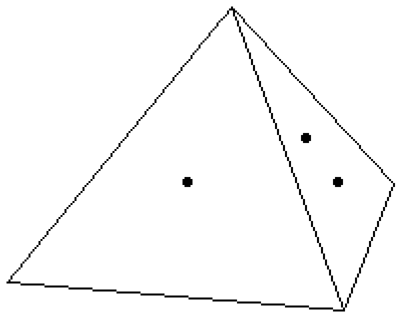
ACTIVIDADES

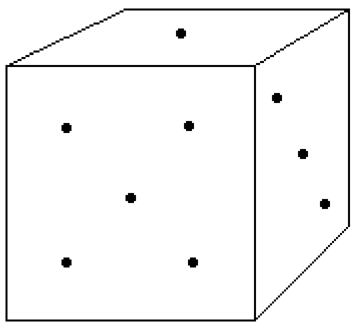
- **TETRAEDRO**

Un dado tetraédrico tiene sus caras marcadas con 1, 2, 3 y 4. Simula con la calculadora gráfica 100 lanzamientos de este dado tetraédrico y dibuja el histograma correspondiente a la variable X=nº de unos. ¿Se puede aproximar este histograma por una curva normal?. En caso afirmativo, dibuja dicha curva normal y compárala con el histograma obtenido.

- **LANZAMIENTO DE UN CUBO**

Simula con la calculadora gráfica 100 lanzamientos de un dado cúbico y dibuja el histograma correspondiente a la variable X=nº de seises. ¿Se puede aproximar este histograma por una curva normal?. En caso afirmativo, dibuja dicha curva normal y compárala con el histograma obtenido.





MODELOS Y DISTRIBUCIONES ESTADÍSTICAS CON LA CALCULADORA GRÁFICA CLASSPAD 300 DE CASIO

Introducción

A partir de los datos muestrales se asigna una probabilidad a cada uno de los datos, usando las frecuencias relativas. La forma de los histogramas correspondientes permite introducir el concepto de modelo probabilístico. El modelo que aparece con más frecuencia es el definido por la curva normal. El estudio de la Estadística inferencial es realmente muy difícil si no se utilizan recursos apropiados.

La calculadora ClassPad 300 permite obtener con facilidad números combinatorios, factoriales, áreas bajo la curva normal y valores de la distribución binomial. También permite obtener estimaciones de parámetros, determinar intervalos de confianza, validar hipótesis, etc.

En esta sesión estudiaremos algunas de las posibilidades de la ClassPad 300 para el estudio de la Probabilidad y la Inferencia Estadística en ESO y Bachillerato

1. Modelos probabilísticos

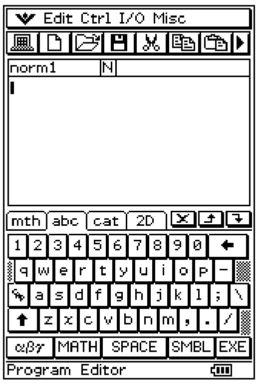
1. DISTRIBUCIÓN NORMAL

Todos los cálculos estadísticos con distribuciones se hacen dentro de un programa. Es decir, para obtener probabilidades utilizando la distribución normal, hay que iniciar previamente la aplicación Programas.

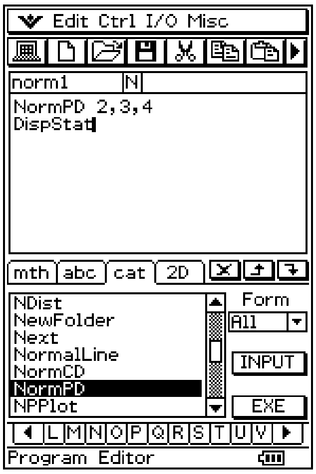
- **Densidad de probabilidad normal**
- El comando NormPD (que se obtiene en el teclado virtual [cat], dentro de la aplicación Programas) calcula la densidad de probabilidad de la distribución normal para un valor x. Utiliza para ello la expresión $f(x)=\frac{1}{\sqrt{2\pi\sigma}}\cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$. La sintaxis del comando es: **NormPD x, σ,** $\bar{x}$ siendo x el dato, σ la desviación típica y $\bar{x}$ la media.
- Si X es una variable aleatoria normal de media 4 y desviación típica 2, calcula p(X=3). Sigue los siguientes pasos:

1) En el menú de aplicaciones toca el botón 

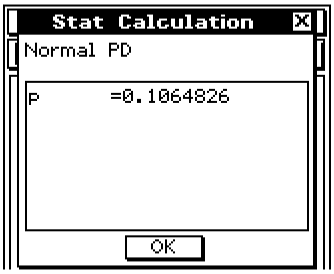
. Se inicia la aplicación programas, mostrando el editor de programas. Selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo norm1 y toca el botón [Acep.].



2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona NormPD y toca el botón [INTRO] para introducir la función NormPD en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando NormPD 3,2,4. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, norm1. Haz clic en el botón [▶] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico.



- Si X es una variable aleatoria normal estándar, calcula p(X=0,6). Utiliza para ello un programa con los comandos NormPD 0.6, 0, 1 y DispStat.

- Representa gráficamente las curvas normales N(7, 1’5), N(10, 2) y N(14, 3). Compara las curvas obtenidas. Calcula respectivamente en cada caso p(X=7), p(X=10) y p(X=14). Sigue los siguientes pasos:

1) En el menú de aplicaciones toca el botón Graph&Tab... para abrir la aplicación Gráficos y Tablas. Toca la primera línea para situar el cursor junto a Y1=.

2) Pulsa [KEYBOARD] para ver el teclado virtual. Toca el botón [2D] y utiliza dicho teclado para introducir la fórmula de la función de densidad normal N(7, 1’5), es decir: $Y1=\frac{1}{1'5\times\sqrt{2\pi}}\cdot e^{-\frac{(x-7)^2}{2\times 1'5^2}}$. Pulsa el cuadro de marcación de Y1 para seleccionar dicha función.

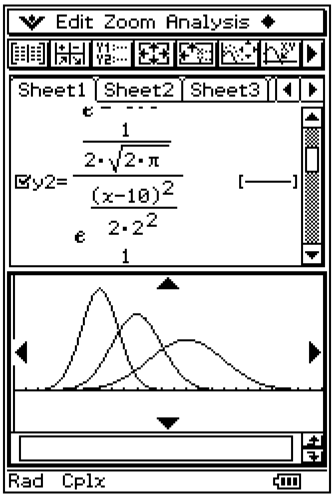
3) Usa el mismo procedimiento para introducir en las líneas Y2 e Y3 las fórmulas de las funciones de densidad normales N(10, 2) y N(14, 3), es decir: $Y2=\frac{1}{2\times\sqrt{2\pi}}\cdot e^{-\frac{(x-10)^2}{2\times 2^2}}$, $Y3=\frac{1}{3\times\sqrt{2\pi}}\cdot e^{-\frac{(x-14)^2}{2\times 3^2}}$. Selecciona los cuadros de marcación de las funciones Y2 e Y3.

4) Toca el botón para acceder al cuadro de configuración de la ventana de visualización. Introduce los siguientes valores: Xmin=0, Xmax=25, Ymin=–0,1, Ymax=0,3. Toca el botón [Acep.].

5) Toca el botón para dibujar las curvas normales seleccionadas.

6) Selecciona el comando Análisis / Trazo. Utiliza las teclas de flecha para situar el cursor en la curva normal N(7, 1’5). Pulsa [7] y toca el botón [Acep.] del cuadro de diálogo Introducir valor. Observa que aparece en pantalla el valor de p(X=7) siendo X una variable normal N(7, 1’5).

7) De la misma forma, calcula el valor de p(X=10) en la curva N(10, 2) y el valor de p(X=14) en la curva N(14, 3).



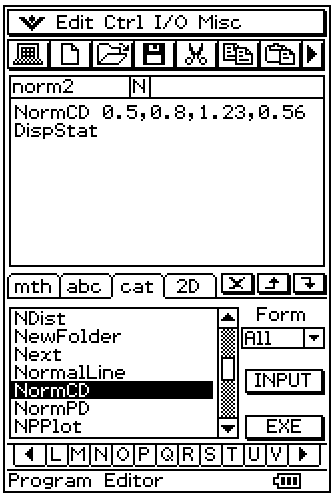
- **Probabilidad normal acumulada**

- El comando NormCD (situado en el teclado virtual [cat]) calcula la probabilidad de que una variable aleatoria normal tome valores comprendidos entre a y b, utilizando la expresión: $p = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_a^b e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \cdot dx$. La sintaxis del comando es: **NormCD a, b, σ,** $\bar{x}$, siendo a y b los extremos inferior y superior del intervalo.

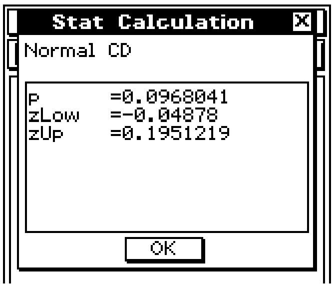
- Si X es una variable aleatoria normal de media 0,56 y desviación típica 1,23, calcula p(0,5≤X≤0,8). Sigue los siguientes pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo norm2 y toca el botón [Acep.].

2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona NormCD y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando NormCD 0.5, 0.8, 1.23, 0.56. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, norm2. Haz clic en el botón [▸] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico, en el que se indica que p(0,5≤X≤0,8)=0.0968041.



- Si X es una variable aleatoria normal de media 4 y desviación típica 2, calcula p(2≤X≤4). Repite el mismo cálculo para una variable aleatoria normal de media 6 y desviación típica 3. Utiliza para ello programas con los comandos NormCD 2, 4, 2, 4 , NormCD 2, 4, 3, 6 y DispStat.

- Representa gráficamente las curvas normales N(4, 2) y N(6, 3) y compáralas. Sigue los siguientes pasos:

1) En el menú de aplicaciones toca el botón Graph&Tab para abrir la aplicación Gráficos y Tablas. Toca la primera línea para situar el cursor junto a Y1=.

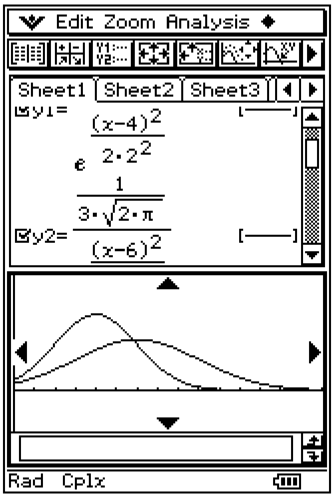
2) Pulsa [KEYBOARD] para ver el teclado virtual. Toca el botón [2D] y utiliza dicho teclado para introducir la fórmula de la función de densidad normal N(4, 2), es decir: $Y1=\frac{1}{2\times\sqrt{2\pi}}\cdot e^{-\frac{(x-4)^2}{2\times 2^2}}$. Pulsa el cuadro de marcación de Y1 para seleccionar dicha función.

3) Usa el mismo procedimiento para introducir en la línea Y2 la fórmula de la función de densidad normal N(6, 3), es decir: $Y2=\frac{1}{3\times\sqrt{2\pi}}\cdot e^{-\frac{(x-6)^2}{2\times 3^2}}$. Selecciona el cuadro de marcación de la función Y2.

4) Toca el botón  para acceder al cuadro de configuración de la ventana de visualización. Introduce los siguientes valores: Xmin=0, Xmax=15, Ymin=–0,1, Ymax=0,3. Toca el botón [Acep.].

5) Toca el botón  para dibujar las curvas normales seleccionadas.

6) Con la ventana de gráficos activa, selecciona el comando Análisis / Trazo. Utiliza las teclas de flecha para situar el cursor en cada una de las gráficas. Compara los valores obtenidos en cada una de ellas.



- **Cuantiles de una distribución normal**

- El comando InvNorm (situado en el teclado virtual [cat]) calcula los cuantiles de una variable aleatoria normal cuando se dan como datos las probabilidades respectivas. Es decir, suponiendo que introducimos como dato la probabilidad de la normal p, se trata de hallar los límites del intervalo, tal como se indica en la siguiente figura:

$\int_{-\infty}^{\alpha} f(x)dx = p$	$\int_{\alpha}^{+\infty} f(x)dx = p$	$\int_{\alpha}^{\beta} f(x)dx = p$
Cola: Izquierda Límite superior del intervalo de integración α = ?	Cola: Derecha Límite inferior del intervalo de integración α = ?	Cola: Central Límites superior e inferior del intervalo de integración α = ? β = ?

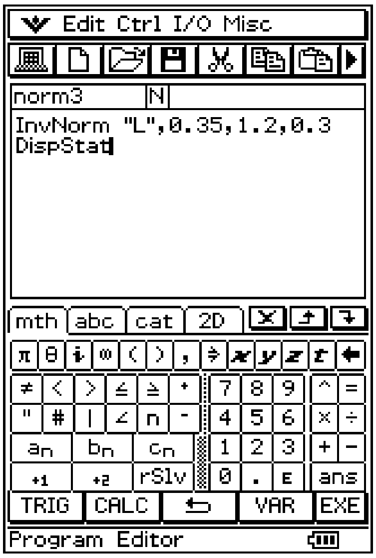
- La sintaxis del comando es: InvNorm “cola”, área(probabilidad), σ, $\overline{x}$

$$\text{siendo cola=}\begin{cases}\text{L, si es cola a la izquierda}\\ \text{R, si es cola a la derecha}\\ \text{C, si es cola central (intervalo finito)}\end{cases} \qquad \text{y siendo área = probabilidad}$$

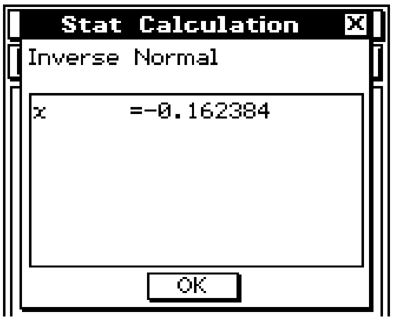
- Si X es una variable aleatoria normal de media 0,3 y desviación típica 1,2, calcula el valor A, tal que p(X≤A)=0,35. Sigue los siguientes pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo norm3 y toca el botón [Acep.].

2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona InvNorm y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando InvNorm “L”, 0.35, 1.2, 0.3. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, norm3. Haz clic en el botón [▶] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico.



- Si X es una variable aleatoria normal de media 4 y desviación típica 2, calcula los extremos A y B, de forma que p(A≤X≤B)=0,75. Utiliza para ello un programa con los comandos InvNorm “C”, 0.75, 2,4 y DispStat.

- Si X es una variable aleatoria normal de media 5 y desviación típica 1,5, calcula el valor B, tal que p(X≥B)=0,57. Utiliza un programa con los comandos InvNorm “R”, 0.57, 1.5, 5 y DispStat.

2. DISTRIBUCIÓN BINOMIAL

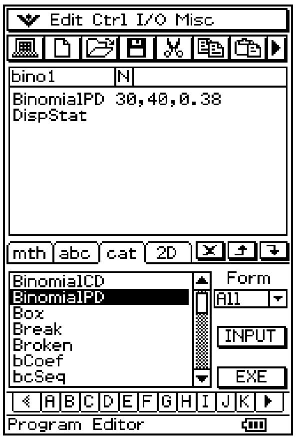
- **Densidad de probabilidad binomial**

- El comando BinomialPD (que se obtiene en el teclado virtual [cat]) calcula la densidad de probabilidad de la distribución binomial para un valor x. Utiliza para ello la expresión

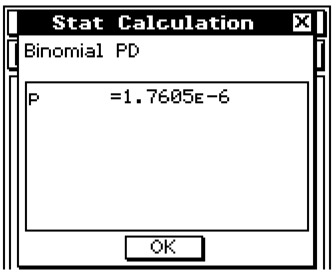
$f(x) = nCx \cdot p^{x} \cdot (1-p)^{n-x} = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$, con x=0, 1, 2, ..., n; siendo p=probabilidad de éxito en una prueba y n el número de pruebas independientes. La sintaxis del comando es: **BinomialPD x, n, p** siendo x el dato, n el número de pruebas independientes y p la probabilidad de éxito en una prueba.

- Si X es una variable aleatoria que sigue una distribución binomial de parámetros n=40 y p=0,38, calcula p(X=30), es decir, la probabilidad de obtener 30 éxitos en 40 pruebas. Sigue los siguientes pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo bino1 y toca el botón [Acep.].
2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona BinomialPD y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando BinomialPD 30, 40, 0,38. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, bino1. Haz clic en el botón [▸] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico.



- Lanzamos 14 monedas perfectamente equilibradas. ¿Cuál es la probabilidad de obtener exactamente 9 caras?. Utiliza para ello un programa con los comandos BinomialPD 9, 14, 0.5 y DispStat.

- Una urna contiene 3 bolas rojas y 7 verdes. Se saca una al azar, se anota el color y se reintegra a la urna. Esta experiencia se realiza 5 veces. ¿Cuál es la probabilidad de obtener tres bolas rojas?. Utiliza un programa con los comandos BinomialPD 3, 5, 0.3 y DispStat.

- **Probabilidad binomial acumulada**

- El comando BinomialCD (situado en el teclado virtual [cat]) calcula la probabilidad de que una variable aleatoria que sigue una distribución binomial tome valores menores o iguales que un valor dado a. La sintaxis del comando es: **BinomialCD a, n, p**, siendo a el valor dado, n el número de pruebas independientes y p la probabilidad de éxito en una prueba.

- Si X es una variable aleatoria que sigue una distribución binomial de parámetros n=40 y p=0,38, calcula $p(X \leq 30)$, es decir, la probabilidad de obtener como máximo 30 éxitos en 40 pruebas. Utiliza para ello un programa con los comandos BinomialCD 30, 40, 0,38 y DispStat.

- En una fábrica de bombillas se sabe que el 2% son defectuosas. Si se empaquetan en cajas de 20 unidades, calcula la probabilidad de que en una caja: a) no haya ninguna defectuosa, b) sólo haya una defectuosa, c) haya más de tres bombillas defectuosas.

- Una urna contiene 3 bolas rojas y 7 verdes. Se saca una al azar, se anota el color y se reintegra a la urna. Esta experiencia se realiza 5 veces. Calcula la probabilidad de obtener: a) menos de 3 bolas rojas, b) más de 3 bolas rojas, c) alguna bola roja.