Tema 11: Distribuciones binomial y normal.

11.1 Variables aleatorias.

- **Variable aleatoria** X: es la función o ley que asocia a cada elemento del espacio muestral un número real. Esto permite sustituir los resultados de una prueba o experimento por números y los sucesos por partes del conjunto de los números reales.
  Las variables aleatorias pueden ser discretas o continuas, según que tomen un número finito o infinito de valores.
  Por ejemplo en el experimento aleatorio de lanzar tres monedas el espacio muestral es E = [CCC, CCX, CXC, XCC, CXX, XCX, XXC, CCC].
  Supongamos que a cada suceso le asignamos el número de caras obtenidas.
  Esta ley o función que acabamos de construir es una variable aleatoria (discreta). Consideremos el experimento que consiste en elegir al azar 100 judías de una plantación y medir su longitud. La ley que asocia a cada judía su longitud es una variable aleatoria continua.
- **Función de probabilidad** (de una variable aleatoria) es la ley que asocia a cada valor de la variable aleatoria x_i su probabilidad $p_i = p(X = x_i)$.
- **Función de distribución F(x)** (de una variable aleatoria) es la ley que asocia a cada valor de la variable aleatoria, la probabilidad acumulada de este valor.
  $$F(x) = p\,(X \leq x)$$
- **Media** de una variable aleatoria discreta: $\boxed{\mu = \sum_{i=1}^{n} x_i p_i}$
- **Varianza** de una variable aleatoria discreta: $\boxed{\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i}$

Ejemplo: Apuntar el nº de caras al lanzar al aire tres monedas.
La función de probabilidad es:

x_i	0	1	2	3
p_i	1/8	3/8	3/8	1/8

La función de distribución es:

x_i	0	1	2	3
p_i	1/8	4/8	7/8	8/8

La media es: $0\cdot(1/8)+1\cdot(3/8)+2\cdot(3/8)+\cdot(1/8) = 1'5$ y la varianza es:
$(0 - 1'5)^2\cdot 1/8+(1 - 1'5)^2\cdot 3/8+(2 - 1'5)^2\cdot 3/8+(3 - 1'5)^2\cdot 1/8 = 0'6667$

Ejemplo: en una bolsa hay bolas numeradas: 9 bolas con un 1, 5 con un 2 y 6 con un 3. Sacamos una bola y vemos que número tienen.
La función de probabilidad es:

x_i	1	2	3
p_i	9/20	5/20	6/20

La función de distribución es:

x_i	1	2	3
p_i	9/20	14/20	20/20

La media es $1\cdot(9/20)+2\cdot(5/20)+3\cdot(6/20) = 1'85$ y la varianza es:
$(1 - 1'85)^2 \cdot 9/20 + (2 - 1'85)^2 \cdot 5/20 + (3 - 1'85)^2 \cdot 6/20 = 0'72$

11.2 Distribución Binomial

- Una variable aleatoria discreta es **binomial** si cumple las siguientes características:
  1. Los elementos de la población se clasifican en dos categorías, éxito o fracaso (A y $\overline{A}$ con probabilidades p y $q = 1 - p$).
  2. El resultado obtenido en cada prueba es independiente de los resultados anteriores.
  3. La probabilidad de éxito y fracaso es siempre constante: p y $q = 1 - p$.

  Ejemplo: fumadores de una población, nº de aprobados de la clase, días de lluvia a lo largo de un año, nº de caras al tirar una moneda, etc.
- **Función de probabilidad:** $\boxed{P(X = r) = \binom{n}{r} p^r \cdot q^{n-r}}$ donde p es la probabilidad de éxito, q la probabilidad de fracaso, n el numero total de pruebas y r el número de éxitos. Las expresiones $\binom{n}{r}$ se llaman números combinatorios y se obtienen a partir de factoriales: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$, siendo el factorial de un número: $n! = n \cdot (n-1) \ldots \cdot 3 \cdot 2 \cdot 1$, ejemplo $3! = 3 \cdot 2 \cdot 1 = 6$.
- **Función de distribución** $\boxed{P(X \le x) = \sum_{r=1}^{x} \binom{n}{r} p^r \cdot q^{n-r}}$
- **Media** $\boxed{\mu = n \cdot p}$
- **Varianza** $\boxed{\sigma^2 = n \cdot p \cdot q}$
- **Desviación típica** $\boxed{\sigma = \sqrt{n \cdot p \cdot q}}$

Ejemplo: Se lanza una moneda 11 veces:

- ¿Cuál es la probabilidad de obtener 5 caras?

$$p(X = 5) = \binom{11}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^6 = 0'2256$$

- ¿Cuál es la probabilidad de obtener 5 o menos caras?

$$p(X \le 5) = \binom{11}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{11} + \binom{11}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{10} + \binom{11}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^9 +$$

$$+\binom{11}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^8 + \binom{11}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^7 + \binom{11}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^6 = 0'5$$

- ¿Cuántas caras se obtienen por término medio?

  $\mu = n \cdot p = 11 \cdot 0'5 = 5'5$
- ¿Cuál es la desviación típica?

  $\sigma^2 = n \cdot p \cdot q = 11 \cdot 0'5 \cdot 0'5 = 2'75$ Luego $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{2'75} = 1'6583$

Otro ejemplo: un examen tipo test tiene 10 preguntas, cada una de ellas con tres opciones para elegir. Si un alumno contesta al azar, calcula la probabilidad de que:

a) No acierte ninguna. Se trata de una binomial $B\left(10, \frac{1}{3}\right)$

$$p[x=0]=\binom{10}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{10}=0'0173$$

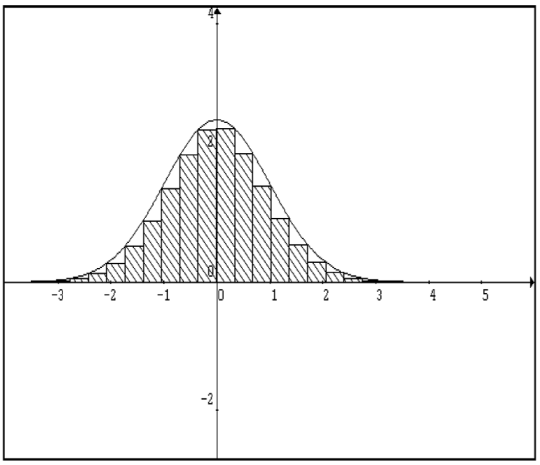
b) Acierte más de 8 preguntas.

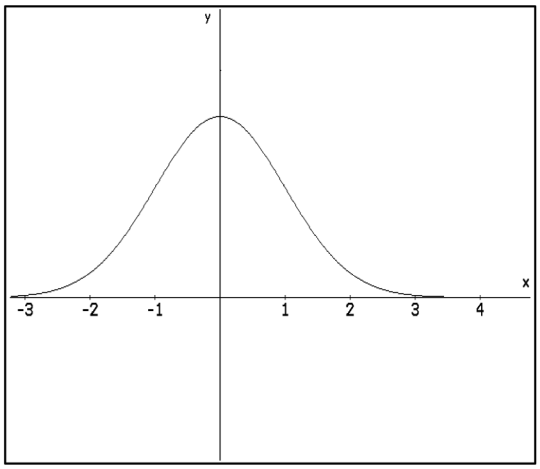
$$p[x>8]=p[x=9]+p[x=10]=\binom{10}{9}\cdot\left(\frac{1}{3}\right)^9\cdot\left(\frac{2}{3}\right)+\binom{10}{10}\cdot\left(\frac{1}{3}\right)^{10}=$$

$$=10\cdot\frac{2^9}{3^{10}}+\frac{1}{3^{10}}=0'0867$$

11.3 Campana de Gauss.

- Carl Friedrich Gauss (matemático alemán 1777 – 1855) estudiando los errores que se producen al medir reiteradamente una magnitud, demostró que se distribuían según la función: $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Siendo μ la media y σ la desviación típica de la distribución de las medidas de la magnitud. Esta función es la misma que resulta cuando en un histograma de frecuencias relativas de una variable continua aumentamos el nº de clases indefinidamente, y por tanto disminuimos indefinidamente la amplitud de los intervalos. En el siguiente gráfico se observa como el polígono de frecuencias relativas se acerca a la función f(x) anterior que llamaremos función de densidad de la distribución normal o campana de Gauss.





Una variable aleatoria es normal si se rige según las leyes del azar. La mayoría de las distribuciones más importantes son normales: pocos individuos en los extremos y aumento paulatino hasta llegar a la parte central en la que están la mayoría de los datos. Ejemplo: la distribución de los pesos de los individuos de cualquier especie, la estatura de una población, la temperatura del mes de agosto a lo largo de 100 años, la longitud de los tornillos que salen de una fábrica, el consumo de cierto producto por familia, el cociente intelectual, el nº de productos defectuosos etc. De hecho se ha

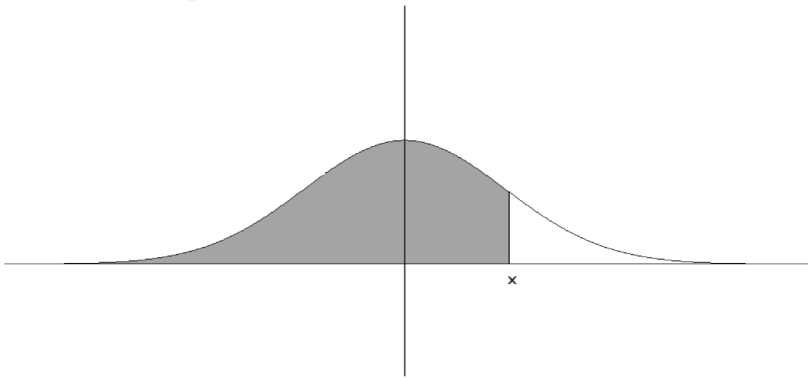
comprobado que la mayoría de los fenómenos relacionados con psicología, pedagogía, biología, etc., siguen una distribución normal.
Sin embargo, no todas las distribuciones son normales, por ejemplo si clasificamos según el nivel de renta a los ciudadanos españoles, son muy pocos los que poseen niveles de rentas altas y en cambio son muchos los que poseen niveles de rentas bajas, por tanto la distribución no sería simétrica y en consecuencia no se adapta al modelo normal.

11.4 Función de densidad y función de distribución

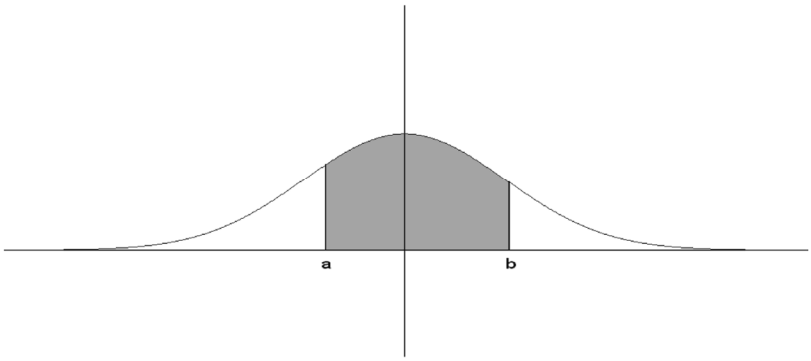
- La función de densidad o campana de Gauss anterior $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

  donde μ es la media y σ la desviación típica cumple las siguientes propiedades:
  - $f(x) \geq 0$
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$, el área encerrada bajo la curva de la función vale 1.
  - $\int_{a}^{b} f(x)dx = p(a \leq X \leq b)$, es decir, la probabilidad es el área bajo la curva correspondiente a ese intervalo.
- La función de distribución o distribución normal es:

  $F(x) = p(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ = área de la región encerrada por la función a la izquierda del número x.

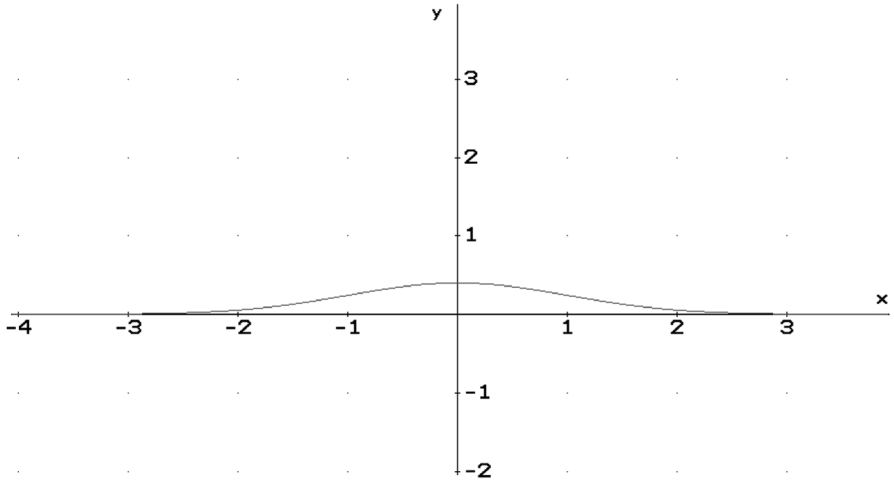


$p(a \leq X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$ = área de la región encerrada por la función entre los números a y b.

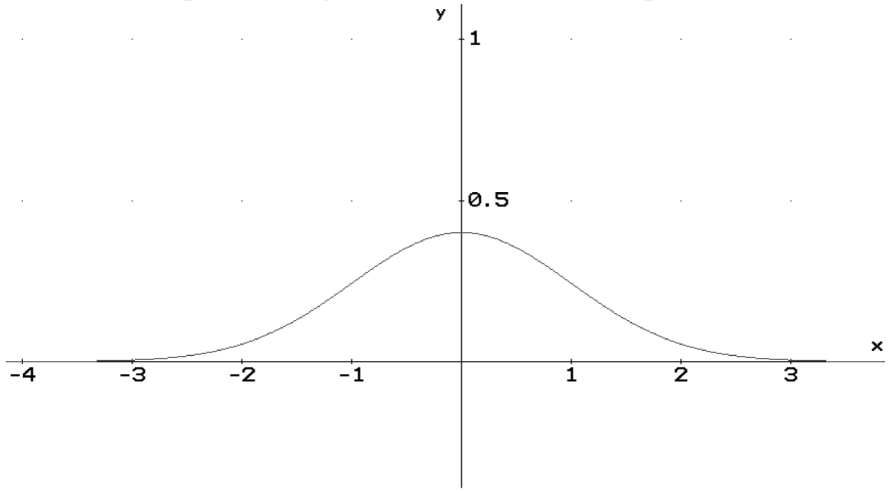


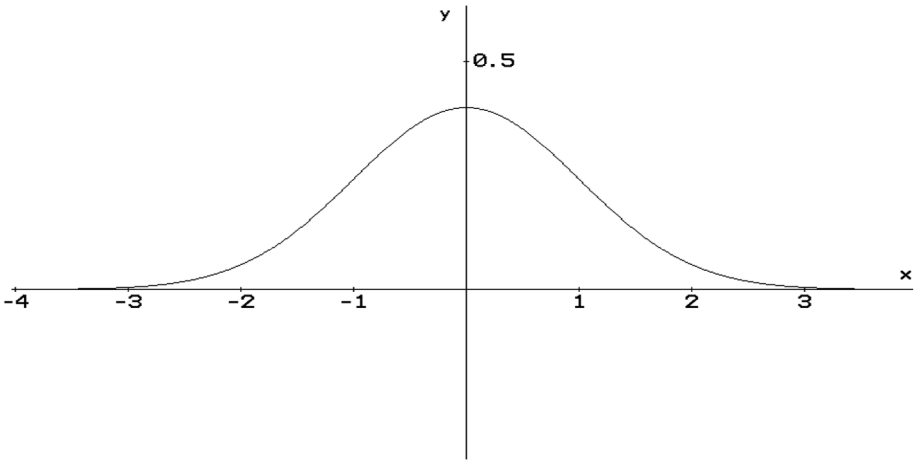
11.5 Tipificación de la variable.

- Cuando $\mu = 0$ y $\sigma = 1$, N(0,1) se dice que tenemos una distribución normal reducida, estándar, o tipificada. La función de densidad es $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$



En la práctica se utiliza un dibujo de la función un poco deformada en el eje de ordenadas, para conseguir una forma mas acampanada.





La distribución N(0,1) se encuentra tabulada, lo cual permite un cálculo rápido de las probabilidades asociadas a esta distribución. Pero en general la media no suele ser 0, ni la varianza 1, por lo que se hace una transformación que se llama **tipificación de la variable**, que consiste en hacer el siguiente cambio de variable:

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

A partir de este cambio de variable obtenemos una variable Z que sí es N(0,1) y que por lo tanto se pueden calcular sus probabilidades utilizando las tablas.

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \sigma \cdot dz$$

Haciendo una utilización conjunta de μ y σ y buscando en las tablas, obtenemos unas relaciones muy importantes**:**

- En $(\mu \pm \sigma)$ está el 68'26% de los datos ya que:

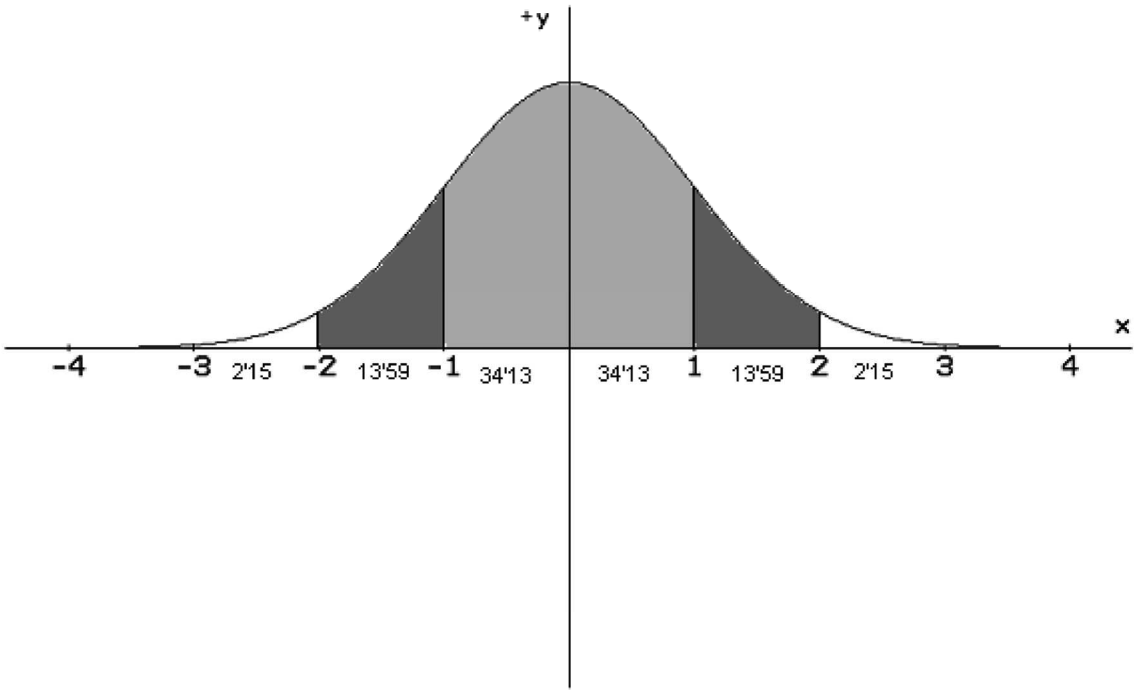
$$p(\mu - \sigma, \mu + \sigma) = p\left(\frac{\mu - \sigma - \mu}{\sigma} < Z < \frac{\mu + \sigma - \mu}{\sigma}\right) = p(-1 < Z < 1) = 0'6826 .$$

- En $(\mu \pm 2\sigma)$ está el 95'44% de los datos ya que:

$$p(\mu - 2\sigma, \mu + 2\sigma) = p\left(\frac{\mu - 2\sigma - \mu}{\sigma} < Z < \frac{\mu + 2\sigma - \mu}{\sigma}\right) = p(-2 < Z < 2) = 0'9544$$

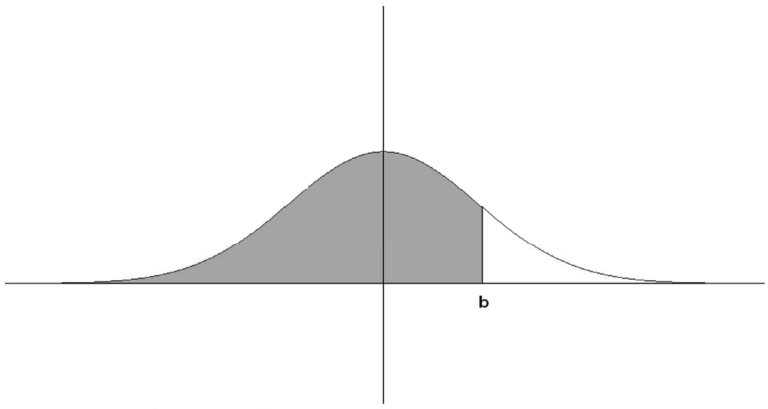
- En $(\mu \pm 3\sigma)$ está el 99'73% de los datos ya que:

$$p(\mu - 3\sigma, \mu + 3\sigma) = p\left(\frac{\mu - 3\sigma - \mu}{\sigma} < Z < \frac{\mu + 3\sigma - \mu}{\sigma}\right) = p(-3 < Z < 3) = 0'9973 .$$



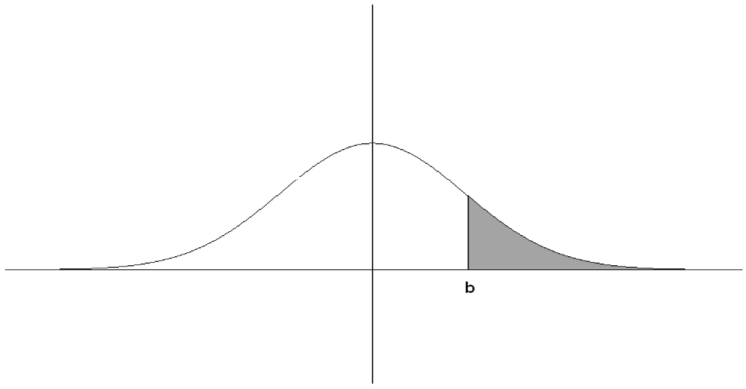
11.6 Manejo de tablas.

- $p(Z < b)$. En este caso se busca en la tabla directamente.



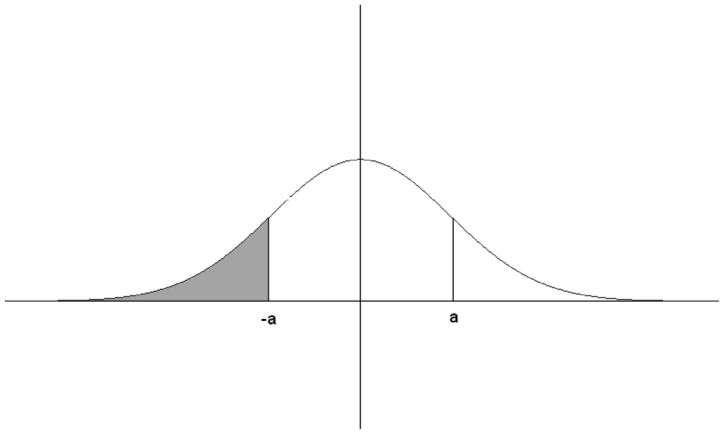
Ejemplo: $p(Z < 1'45) = 0'9265 = 92'65\%$.

- $p(Z > b) = 1 - p(Z < b)$. Aquí hay que tener en cuenta que $Z > b$ y $Z < b$ son complementarios, por lo que sus probabilidades suman 1.



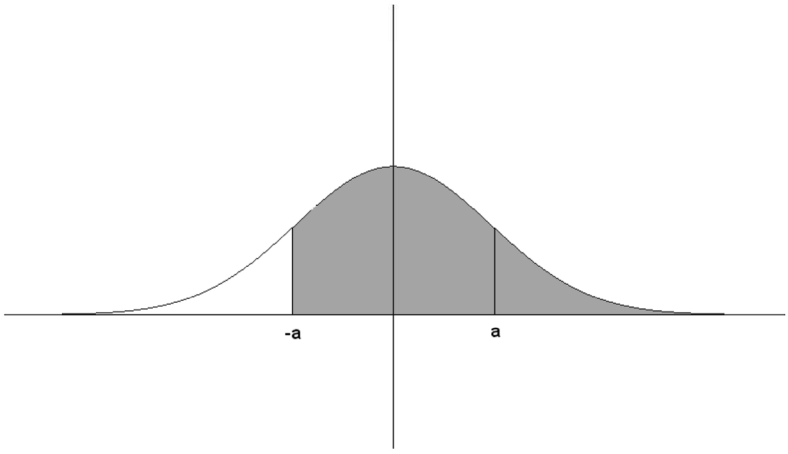
Ejemplo: $p(Z > 1'45) = 1 - p(Z < 1'45) = 1 - 0'9265 = 0'0735 = 7'35\%$.

- $p(Z < -a) = p(Z > a) = 1 - p(Z < a)$.



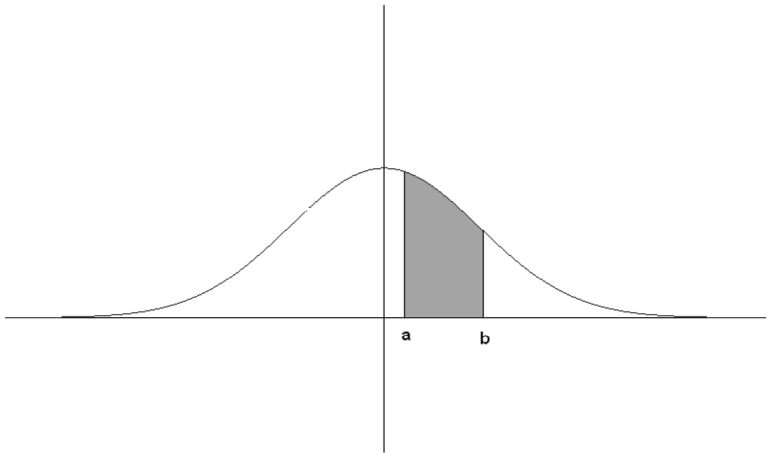
Ejemplo:
$p(Z < -0'75) = p(Z > 0'75) = 1 - p(Z < 0'75) = 1 - 0'7734 = 0'2266 = 22'66\%$

- $p(Z > -a) = 1 - p(Z < -a) = 1 - p(Z > a) = 1 - [1 - p(z < a)] = 1 - 1 + p(Z < a) = p(z < a)$



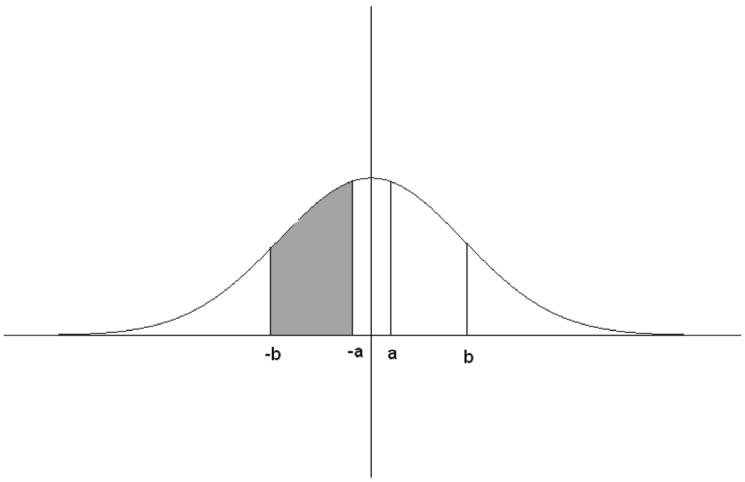
Ejemplo:
$p(Z > -0'75) = 1 - p(Z < -0'75) = 1 - p(Z > 0'75) = 1 - [1 - p(z < 0'75)] =$
$= 1 - 1 + p(Z < 0'75) = p(z < 0'75) = 0'7734 = 77'34\%$

- $p(a < Z < b) = p(Z < b) - p(Z < a)$ y se busca en las tablas.



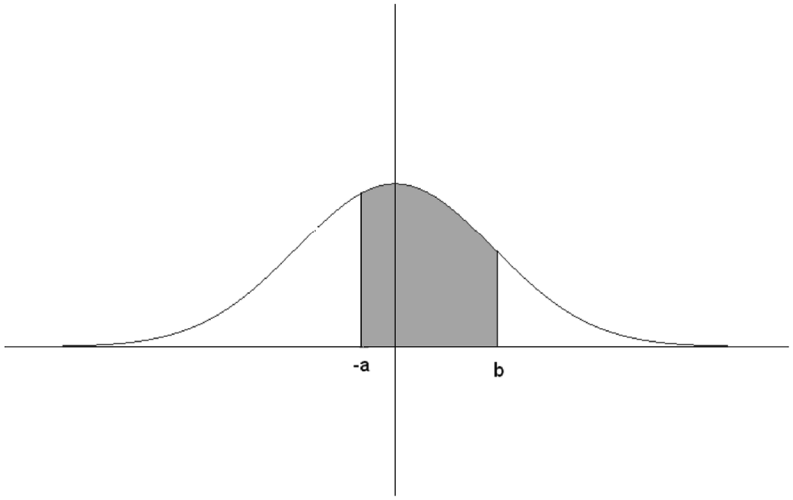
Ejemplo:
$p(1'25 < Z < 2'57) = p(Z < 2'57) - p(Z < 1'25) = 0'9949 - 0'8944 =$
$= 0'1005 = 10'05\%$

- $p(-b < Z < -a) = p(a < Z < b)$



Ejemplo: $p(-2'57 < Z < -1'25) = p(1'25 < Z < 2'57) = 0'1005 = 10'05\%$.

- $p(-a < Z < b) = p(Z < b) - p(Z < -a) = p(Z < b) - p(Z > a) = p(Z < b) - [1 - p(Z < a)] =$
  $= p(Z < b) - 1 + p(Z < a)$ y se busca en las tablas.



Ejemplo:
$p(-0'53 < Z < 2'46) = p(Z < 2'46) - p(Z < -0'53) = p(Z < 2'46) - p(Z > 0'53) =$
$= p(Z < 2'46) - [1 - p(Z < 0'53)] = p(Z < 2'46) - 1 + p(Z < 0'53) =$
$= 0'9931 - 1 + 0'7019 = 0'6950 = 69'50\%.$

Ejemplo: Si tenemos una distribución normal N(2,4). Calcular $p(X < 7)$:

$$p(X < 7) = p\left(\frac{X-2}{4} < \frac{7-2}{4}\right) = p\left(Z < -\frac{5}{4}\right) = p(Z > 1'25) = 1 - p(Z < 1'25) =$$
$$= 1 - 0'8944 = 0'1056 = 10'56\%.$$

Ejemplo: una cadena hotelera quiere ofrecer a un grupo de personas nuevos destinos turísticos. Para realizar la selección, tiene en cuenta dos factores: la edad y los ingresos mensuales. Se selecciona aleatoriamente un grupo de personas cuyas edades e ingresos siguen unas distribuciones N(44, 5) y N(1900, 150) respectivamente. Calcula el porcentaje de personas cuya edad está comprendida entre 38 y 50 años.

$$p(38 < X < 50) = p\left(\frac{38-44}{5} < \frac{X-44}{5} < \frac{50-44}{5}\right) = p(-1'2 < Z < 1'2) =$$
$$= p(Z < 1'2) - p(Z < -1'2) = p(Z < 1'2) - p(Z > 1'2) = p(Z < 1'2) - [1 - p(Z < 1'2)] =$$
$$= p(Z < 1'2) - 1 + p(Z < 1'2) = 0'8849 - 1 + 0'8849 = 0'7698 = 76'98\%.$$

¿Qué porcentaje de personas tienen ingresos mensuales entre 1675€ y 2095€?

$$p(1675 < X < 2095) = p\left(\frac{1675-1900}{150} < \frac{X-1900}{150} < \frac{2095-1900}{150}\right) = p(-1'5 < Z < 1'3) =$$
$$= p(Z < 1'3) - p(Z < -1'5) = p(Z < 1'3) - p(Z > 1'5) = p(Z < 1'3) - [1 - p(Z < 1'5)] =$$
$$= p(Z < 1'3) - 1 + p(Z < 1'5) = 0'9032 - 1 + 0'9332 = 0'8364 = 83'64\%.$$

Ejemplo: el C.I. de los 5600 alumnos de una provincia se distribuyen según una distribución normal N(112,6). Calcular aproximadamente cuántos de ellos tienen: (practica ahora tú y comprueba los resultados)

a) más de 112 .....2800 alumnos.....la mitad de los alumnos
b) entre 106 y 118 .....3823 alumnos .....este es el caso : $(\mu \pm \sigma)$
c) entre 106 y 112 .....1911 alumnos
d) menos de 100 .....128 alumnos
e) más de 130 .....7 alumnos
f) entre 118 y 124 .....761 alumnos

11.7 Aproximación de la distribución binomial.

- Cuando los valores para calcular las probabilidades de la distribución binomial superan a los valores de las tablas, se puede obtener un resultado aproximado utilizando la distribución normal. Según demostró De Moivre en la Binomial B(n,p) si $np \geq 5$ y $nq \geq 5$, podemos aproximar las probabilidades mediante la distribución normal $N\left(np, \sqrt{npq}\right)$. Y tipificando tendremos que $Z = \dfrac{X - np}{\sqrt{npq}}$ que es una distribución N(0,1). Este resultado es tanto más fiable cuanto mayor es el tamaño de la muestra n y cuanto más cerca está p de $0'5$.

  Ejemplo: Se ha comprobado que la probabilidad de tener un individuo los ojos marrones es $0'6$. Sea X la variable aleatoria que representa el nº de individuos que tienen los ojos marrones de un grupo de 1100. Calcular $p(X > 680)$ y $p(X = 680)$.

  $$p(X > 680) = 1 - p(X < 680) = 1 - p\left(Z < \frac{680 - 1100 \cdot 0'6}{\sqrt{1100 \cdot 0'6 \cdot 0'4}}\right) = 1 - p(Z < 1'23) =$$

  $$= 1 - 0'8907 = 0'1093 = 10'93\%.$$

  $$p(X = 680) = p(679'5 < X < 680'5) = p\left(\frac{679'5 - 1100 \cdot 0'6}{\sqrt{1100 \cdot 0'6 \cdot 0'4}} Z < \frac{680'5 - 1100 \cdot 0'6}{\sqrt{1100 \cdot 0'6 \cdot 0'4}}\right) =$$

  $$= p(1'20 < Z < 1'26) = p(Z < 1'26) - p(Z < 1'20) = 0'8962 - 0'8849 = 0'0113 = 1'13\%.$$

  Se debe hacer así, ya que en una variable continua no tiene sentido calcular probabilidades de valores puntuales.