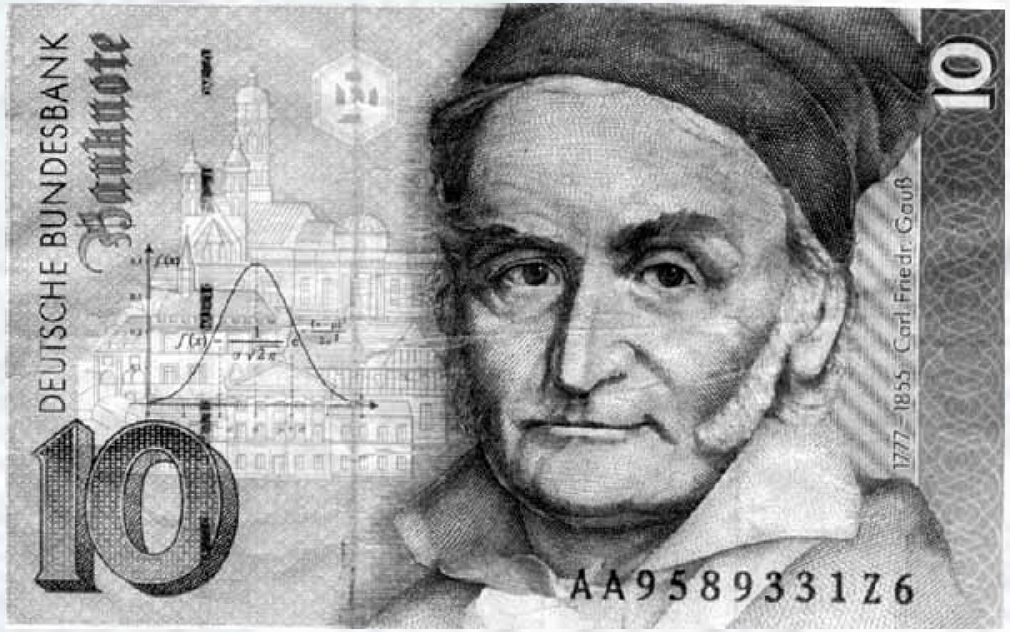
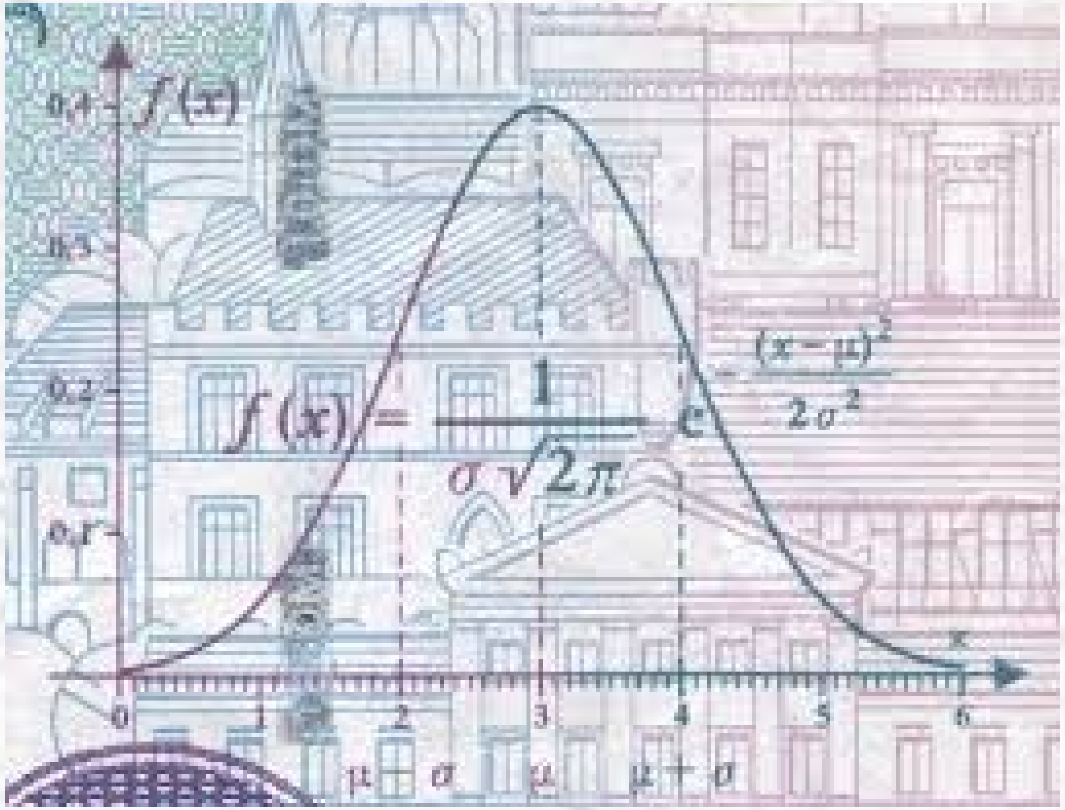


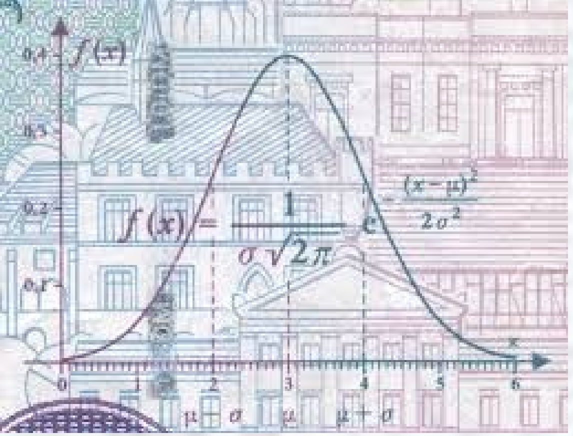
Johann Carl Friedrich Gauss (1777-1855), en el billete de 10 marcos (Alemania)



Una variable aleatoria continua tiene distribución normal de parámetros μ y (> 0) con x real (el recorrido es) si la función densidad de probabilidad es:

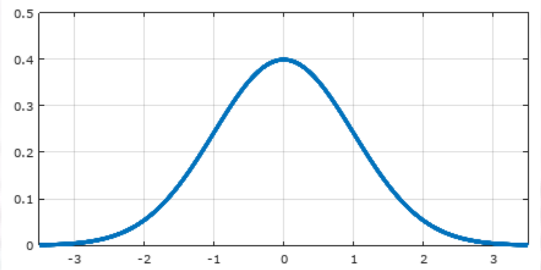
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La famosa ***campana de Gauss*** (gráfico de función densidad de probabilidad) del anverso del billete para el caso en que μ = 3 y igual a 1.



¿

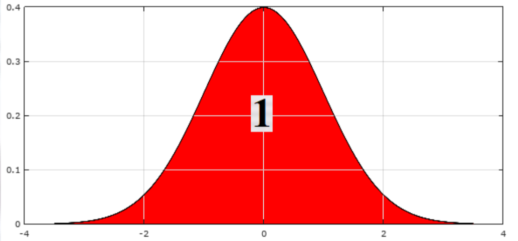
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



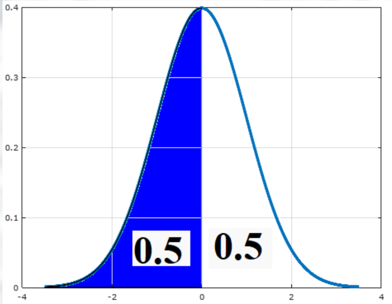
$$f(x)>0 \; \forall x \in \mathscr{R}$$

$$f(x)|x| \overset{\rightarrow}{\rightarrow} \infty \, 0$$

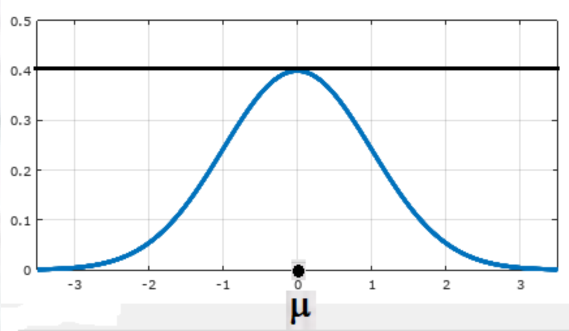
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

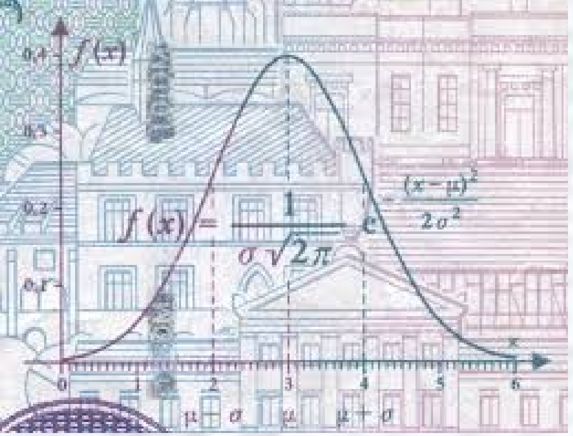


$$f(\mu+d)=f(\mu-d) \; \forall d \int_{-\infty}^{\mu} f(x)\,dx = ¿ \int_{\mu}^{\infty} f(x)\,dx = ¿ 0.5 ¿¿$$



$$f'(\mu)=0 \; y \; f''(\mu)<0 \; f(x) \leq f(\mu) \; \forall x \in \mathscr{R}$$





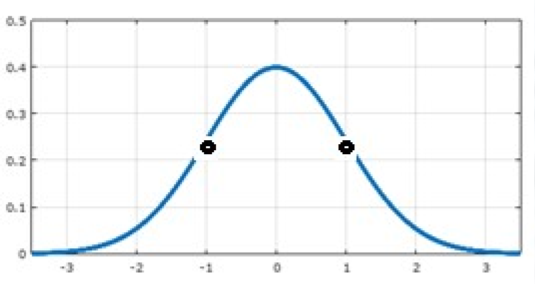
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x) > 0 \; \forall x \in \mathscr{R}$ $\qquad f(x)\, |x| \vec{\rightarrow} \infty\, 0$ $\qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$

$f(\mu+d) = f(\mu-d) \; \forall d \int_{-\infty}^{0} f(x)\,dx = ¿ \int_{-\infty}^{0} f(x)\,dx = ¿0.5¿¿$

$f'(\mu) = 0 \text{ y } f''(\mu) < 0 \; f(x) \le f(\mu) \; \forall x \in \mathscr{R}$

$f''(\mu+\sigma) = f''(\mu-\sigma) = 0$



$E(X) = \mu$

El parámetro es **media**, **mediana** y **modo**.

$V(X) = \sigma^2 \sigma(X) = \sigma$

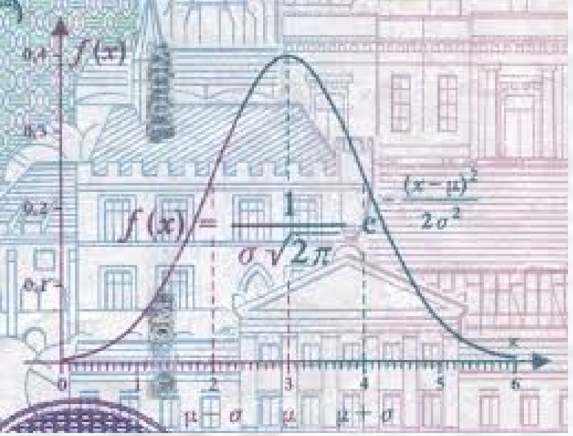
El parámetro es el desvío estándar de X y fija una escala horizontal. Su recíproca fija la escala vertical.

$\gamma(X) = 0$

La función densidad de probabilidad es simétrica respecto de entonces es nulo el coeficiente de simetría

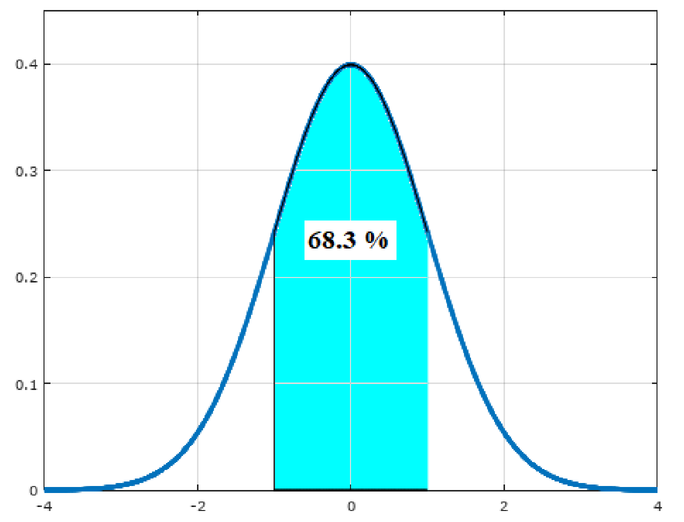
$\kappa(X) = 3$

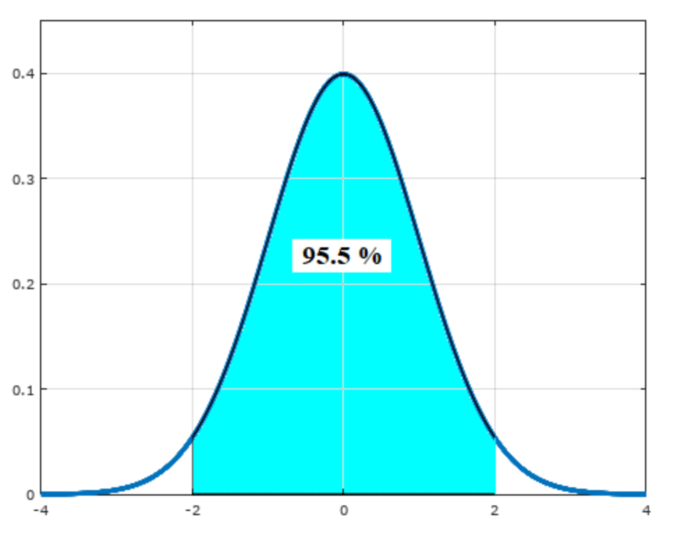
El coeficiente de kurtosis 3 de la normal se usa para comparar la forma de otras densidades simétricas respecto de la normal.

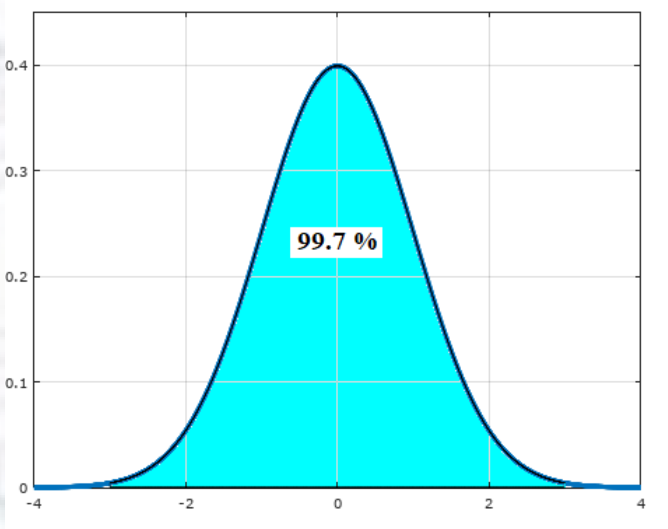


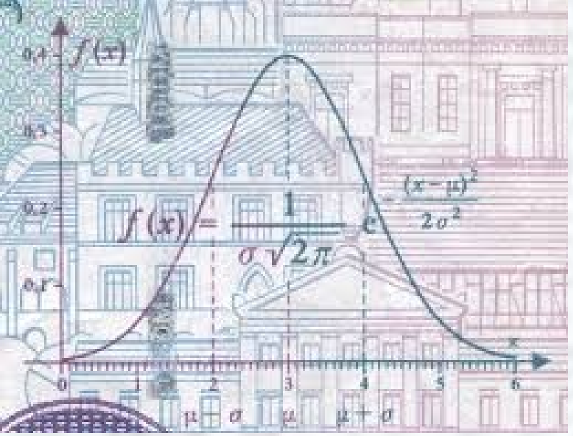
$$f(x) = \frac{1}{\sqrt{2\,\pi\,\sigma}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

1	0.683 (68.3%)
2	0.955 (95.5 %)
3	0.997 (99.7 %)

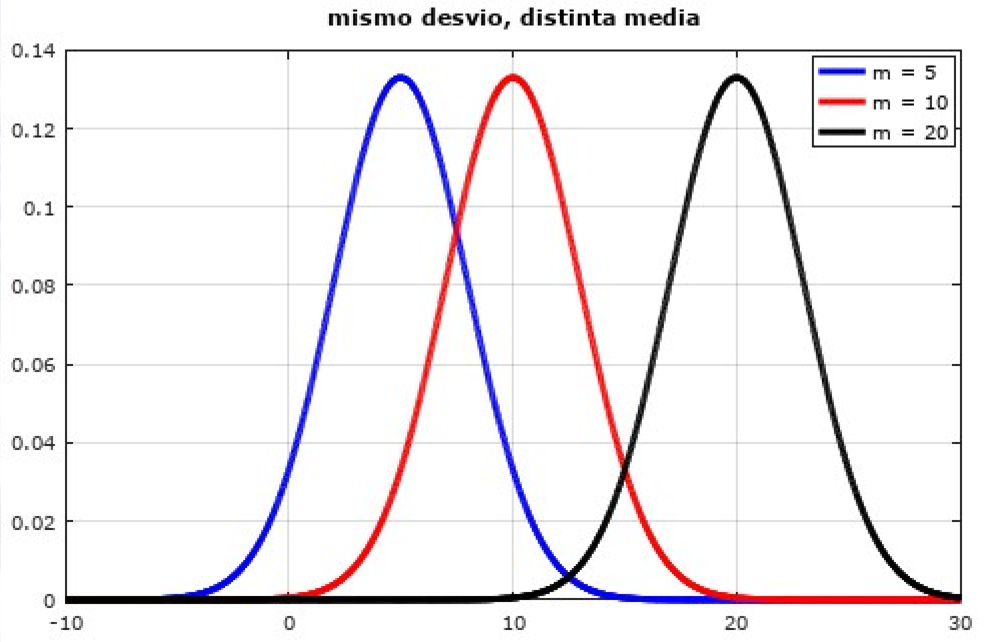


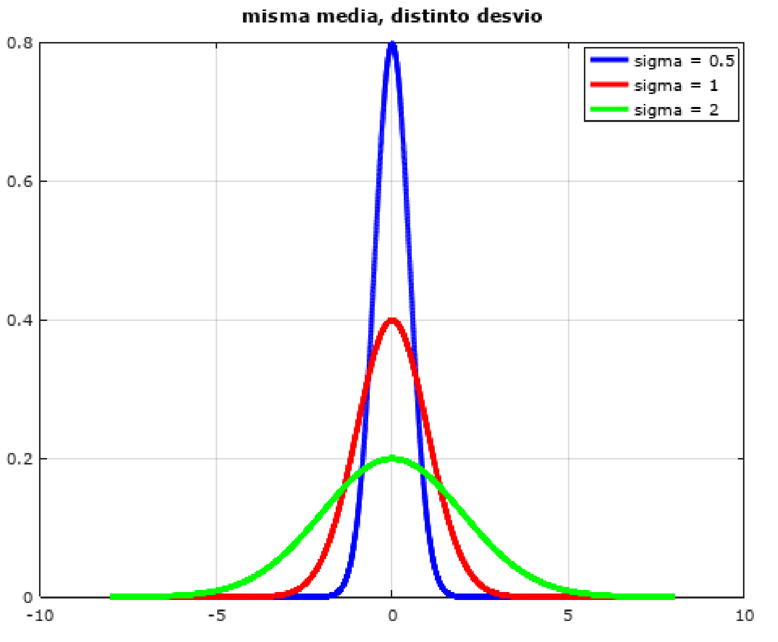






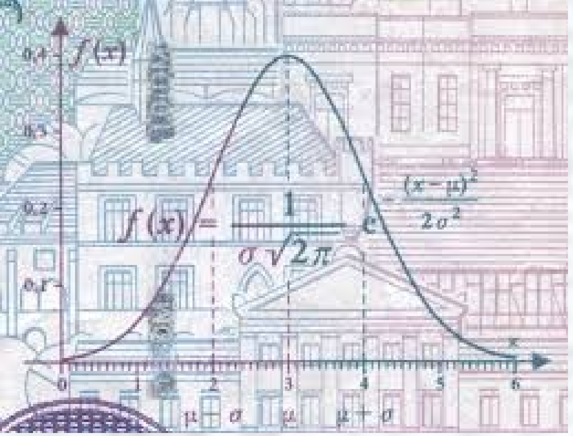
$$f(x) = \frac{1}{\sqrt{2\,\pi\,\sigma}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



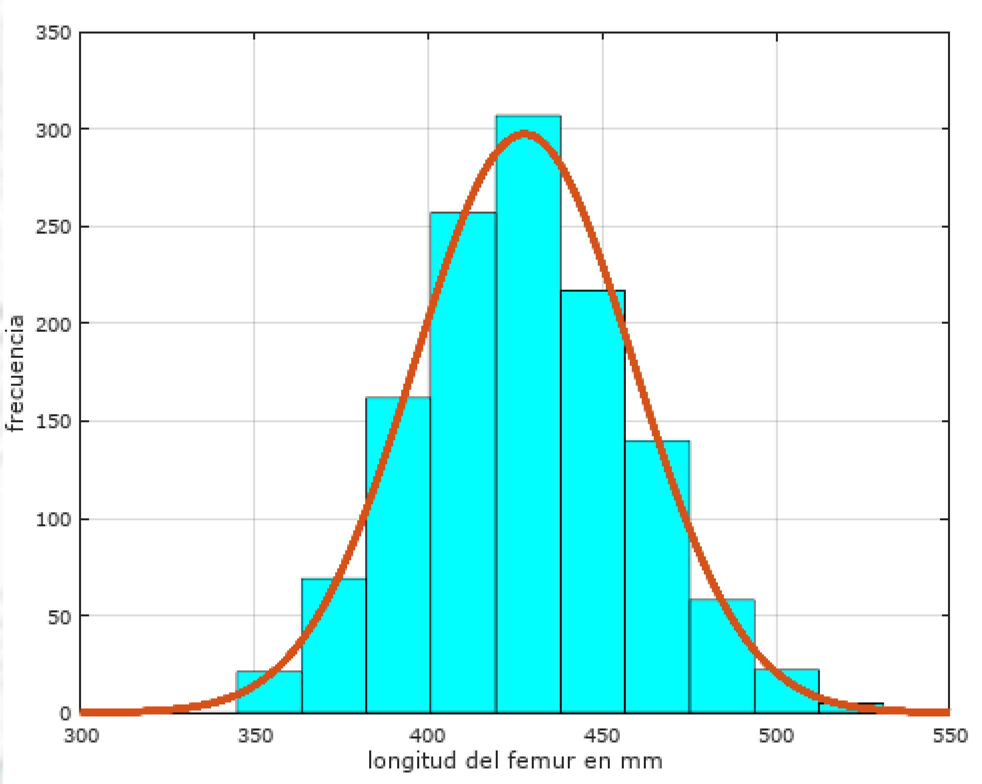


Como el parámetro μ es la media, mediana y moda
al aumentar se traslada “la campana” .
Este parámetro es de **posición**.

Al variar el parámetro   se modifican las escalas
horizontal y vertical. El intervalo de valores de interés
de la variable tiene ancho 6



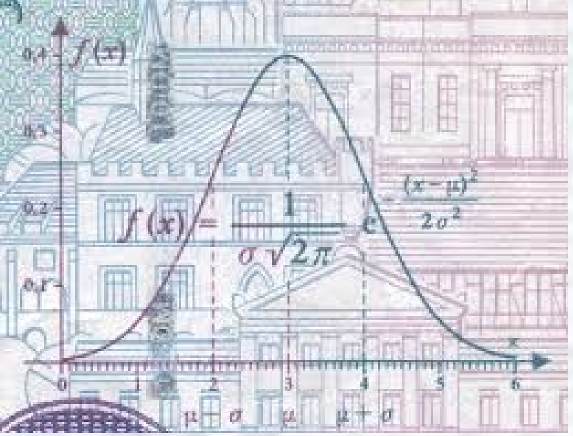
La **distribución normal** es muy importante ya que permite modelar muchos fenómenos naturales, sociales, fisiológicos y psicológicos. En estos fenómenos hay muchas variables intervinientes y ocurre que las observaciones pueden asumirse que son suma de no demasiadas causas independientes (ya lo veremos al presentar el teorema Central del Límite).



Longitud del fémur izquierdo en mm en 1200 personas (The Goldman Osteometric Data Set)

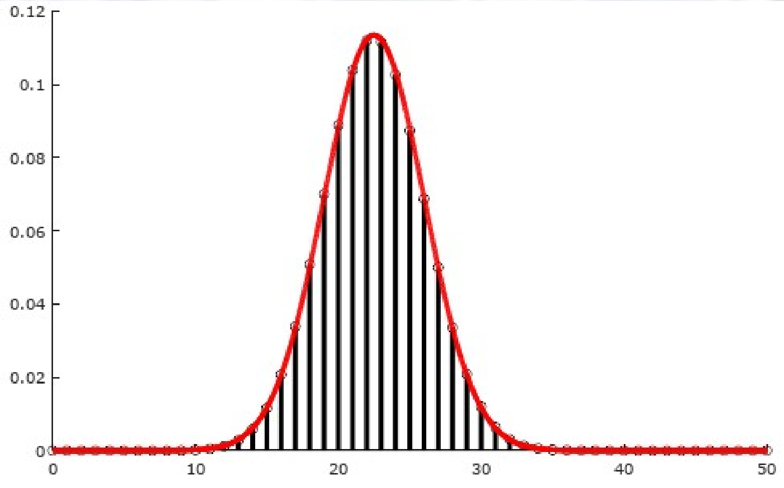
Algunos ejemplos de variables asociadas a fenómenos naturales y sociales pueden ser estos:

a) Características de personas tales como el **peso, estatura** y otras de un mismo grupo.
b) Efecto de medicamentos en las personas
c) El **consumo de ciertos productos** por un grupo de personas
d) El cociente intelectual (**IQ**)
e) El **ruido de señales** en telecomunicaciones
f) Errores cometidos en la medición de magnitudes experimentales.
g) Otras muchas …

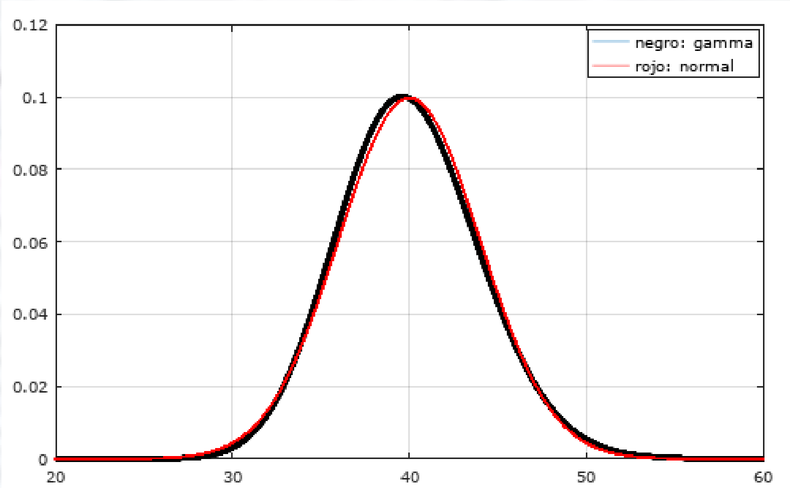


Encontraremos esta **distribución normal en muchas áreas de la estadística**. La distribución de probabilidades de la **media de una muestra aleatoria** veremos que es normal si proviene de medir una variable normal y aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal. La **distribución normal es la más extendida en estadística** y muchas pruebas estadísticas están basadas en la **normalidad de la variable** aleatoria bajo estudio.

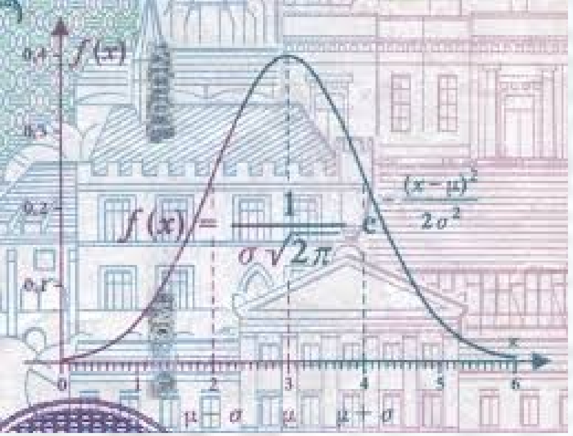
En el cálculo de probabilidades, la **distribución normal aparece como el límite de varias distribuciones de probabilidad** continuas y discretas.



Función de probabilidad de una variable binomial $n = 50$, $p = 0.45$ y función de densidad de probabilidad normal de igual media y desvío estándar. La máxima diferencia de ordenadas es menor que 1e-3.



Función de densidad de probabilidad de una variable Gama (con r = 100, L = 0.4) y función de densidad de probabilidad normal de igual media y desvío estándar. La máxima diferencia de ordenadas es menor que 1e-2.



¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

$$P(a<X<b)=\int_a^b f_X(x)\,dx=F_X(b)-F_X(a)$$

$$F_X(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{¿¿}\,¿$$

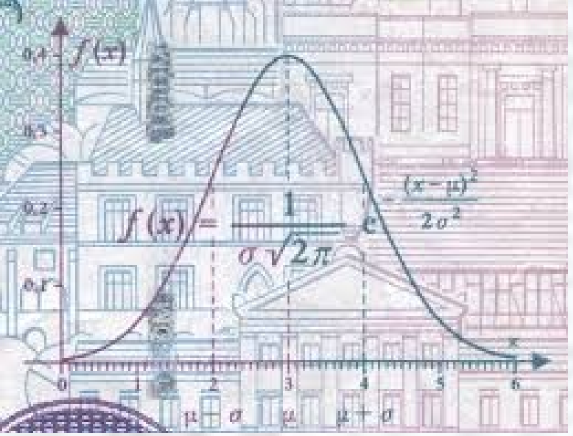
La evaluación de esta integral presenta dos problemas. Sin duda el integrando es una función continua por lo tanto existe su primitiva. La cuestión es que no existe una primitiva que se pueda evaluar en forma exacta en un número finito de operaciones. La segunda cuestión es que depende de dos parámetros y no podría reseñarse en una tabla la solución numérica de la integral …

Hay una transformación del integrando (un cambio de variable) que permite expresar el cálculo de esta función de distribución en términos de otra **libre de parámetros**: y esa es la que se resuelve numéricamente …

Estandarización de la variable X: la variable Z se denomina estandarizada , es **adimensional**, tiene **valor esperado 0** y **varianza 1**.

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

Estandarización de la variable X: la variable Z mide desvíos de X usando el desvío estándar como unidad de medida. Por ejemplo Z = 1.5 es un desvío a la derecha de μ de magnitud 1.5 .



¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

$$P(a < X < b) = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{¿¿} ¿$$

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

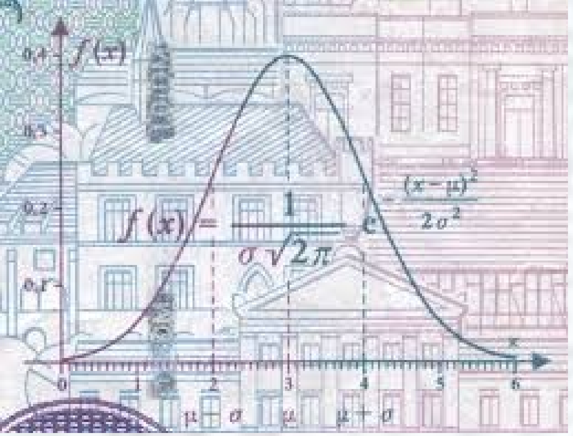
Estandarización de la variable X: la variable Z sugiere el cambio de variable a realizar en el integrando (su cuadrado está en el exponente)…

Introduciendo el cambio de variable: $z = \frac{t-\mu}{\sigma}$

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu¿/\sigma} e^{-z^2/2}\,dz = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

donde: $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du$

La función Φ es la función de distribución de Z y comparándola con la función *F* resulta ser la de una variable aleatoria normal de media μ = 0 y = 1. La variable aleatoria Z se denomina **normal estándar** y su función de distribución Φ está **libre de parámetros**.



¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

$$P(a<X<b)=\int_a^b f_X(x)\,dx=F_X(b)-F_X(a)$$

$$F_X(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{¿¿}¿$$

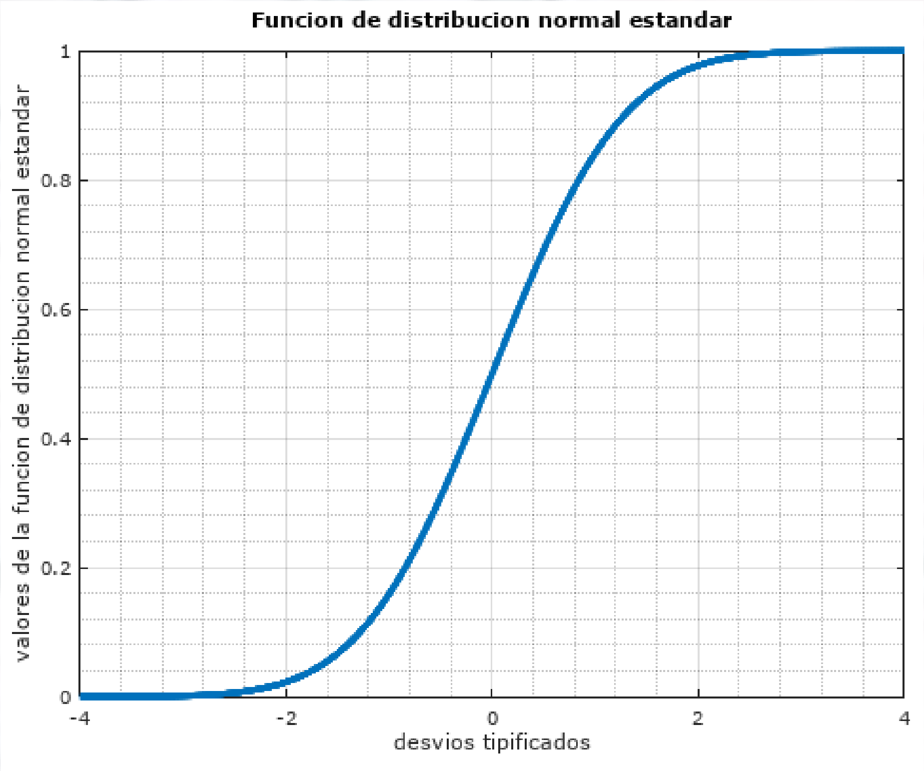
$$Z=\frac{X-\mu}{\sigma}$$

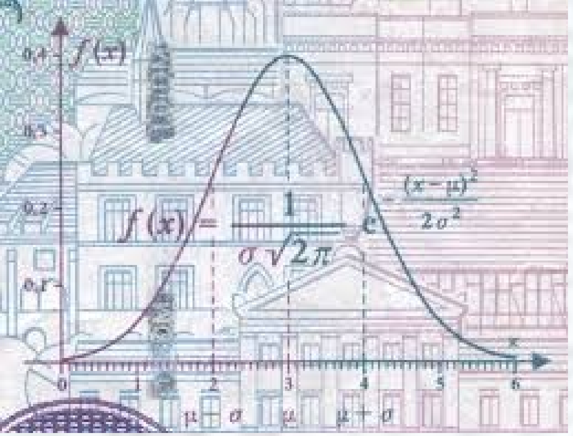
Estandarización de la variable X: la variable Z se denomina variable aleatoria normal estándar.

$$F_X(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-\mu¿/\sigma} e^{-z^2/2}dz=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

donde:

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-u^2/2}du$$

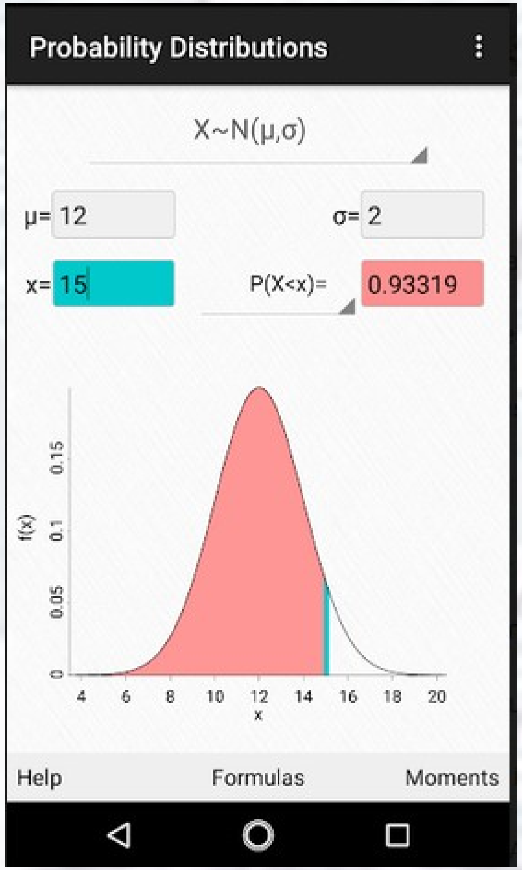


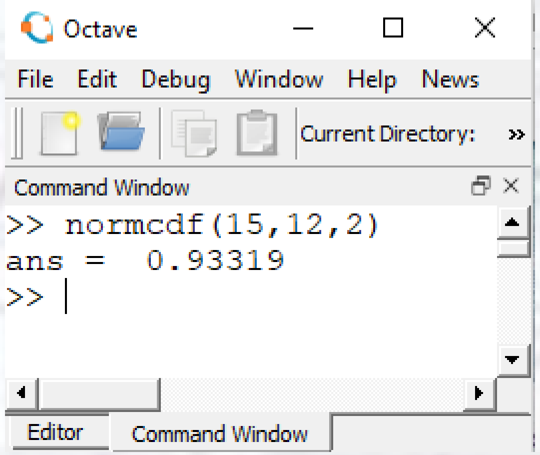


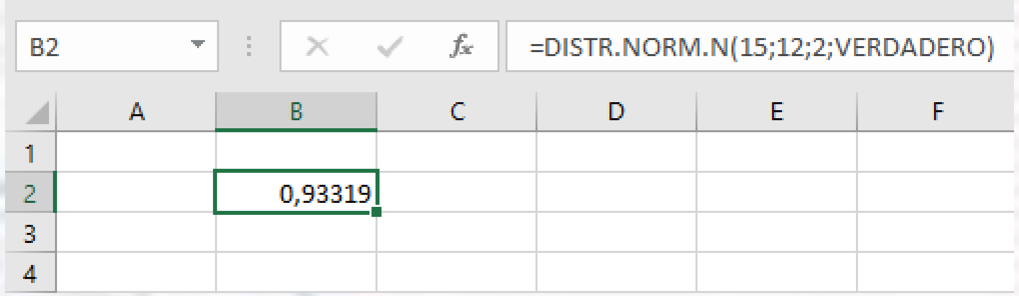
¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

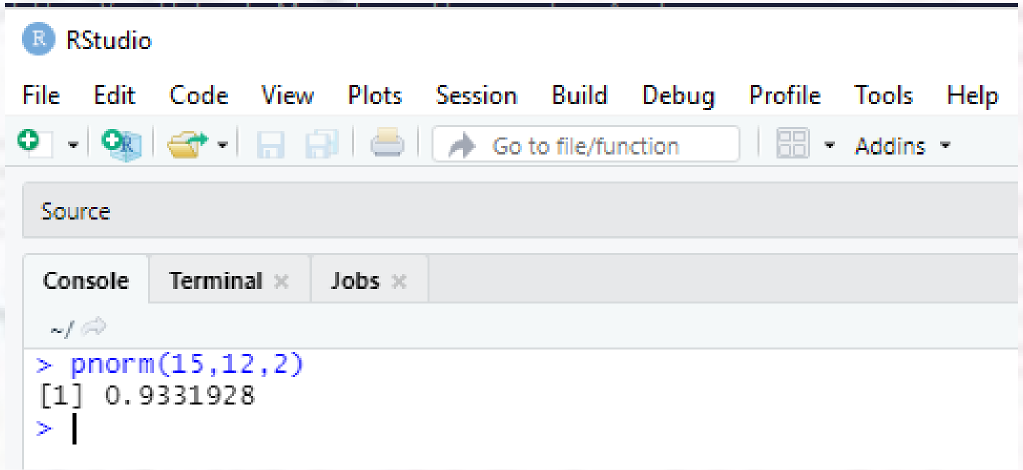
$$P(a<X<b)=\int_a^b f_X(x)\,dx=F_X(b)-F_X(a)$$

Supongamos que X tiene distribución normal de media 12 y desvío estándar 2. ¿ cuál es la probabilidad de que X sea menor que 15?

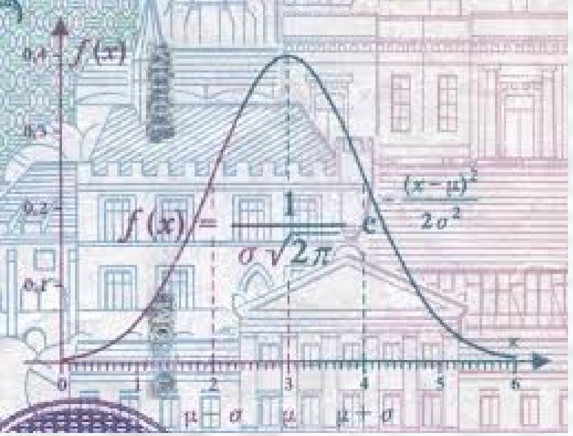








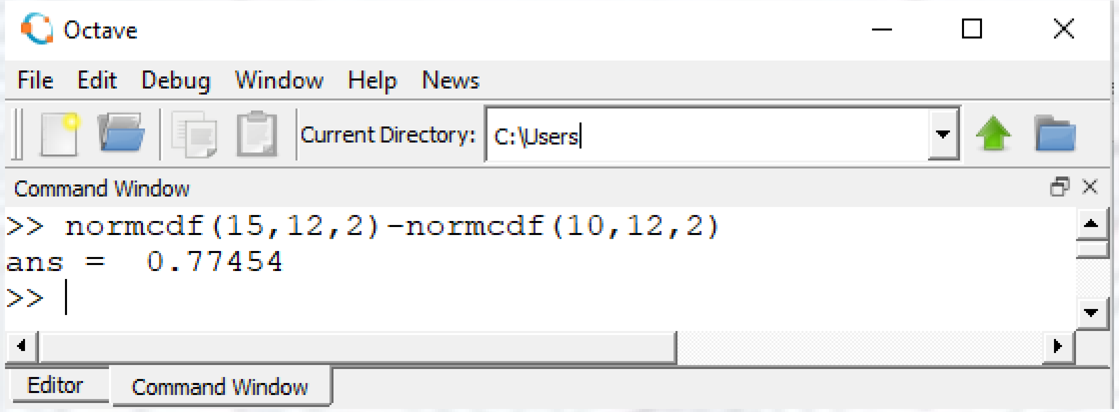
Evaluando la función de distribución de X usando una aplicación de celular, Octave, la planilla de cálculo y R.



¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

$$P(a<X<b)=\int_a^b f_X(x)\,dx=F_X(b)-F_X(a)$$

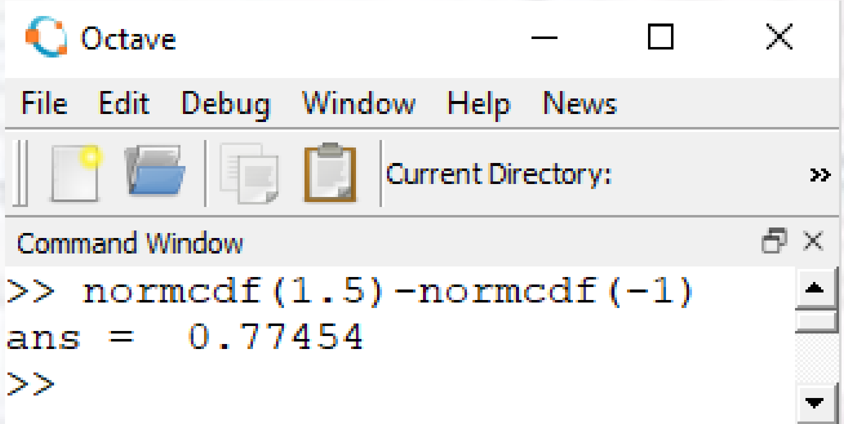
Supongamos que X tiene distribución normal de media 12 y desvío estándar 2. ¿ cuál es la probabilidad de que X tome valores entre 10 y 15?

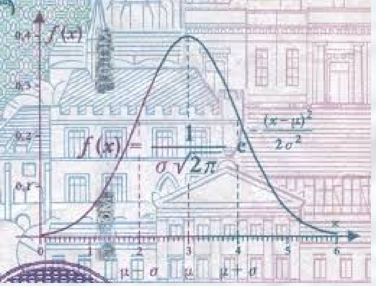


Usando la distribución de la variable normal estándar hay que estandarizar los valores de X:

$$P(a<X<b)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

)





¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

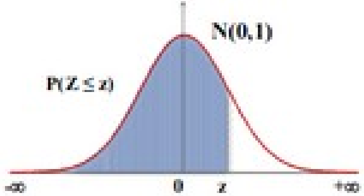
$$P(a<X<b)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

Existen tablas que reseñan valores de la función Φ en un arreglo de puntos equiespaciados como muestra la figura adjunta. Es una tabla de doble entrada. Por la primera columna se ingresa la parte entera y el primer decimal y el segundo decimal se ubica por la primera fila. En la intersección de fila y columna se obtiene el valor de Φ.

z	0,00	0,01	0,02	0,03	0,04	0,05
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749

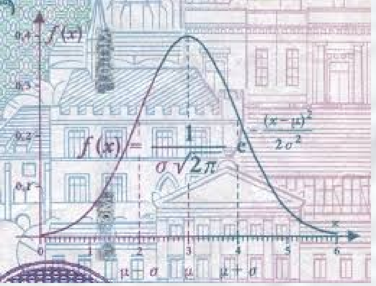
$$\Phi(0.64)\approx 0.7389$$

FUNCIÓN DE DISTRIBUCIÓN NORMAL N(0,1)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Nota: En el interior de la tabla se da la probabilidad de que la variable aleatoria Z, con distribución N(0,1), esté por debajo del valor z.



¿Cómo se calcula la probabilidad de que una variable aleatoria normal tome valores en un intervalo?

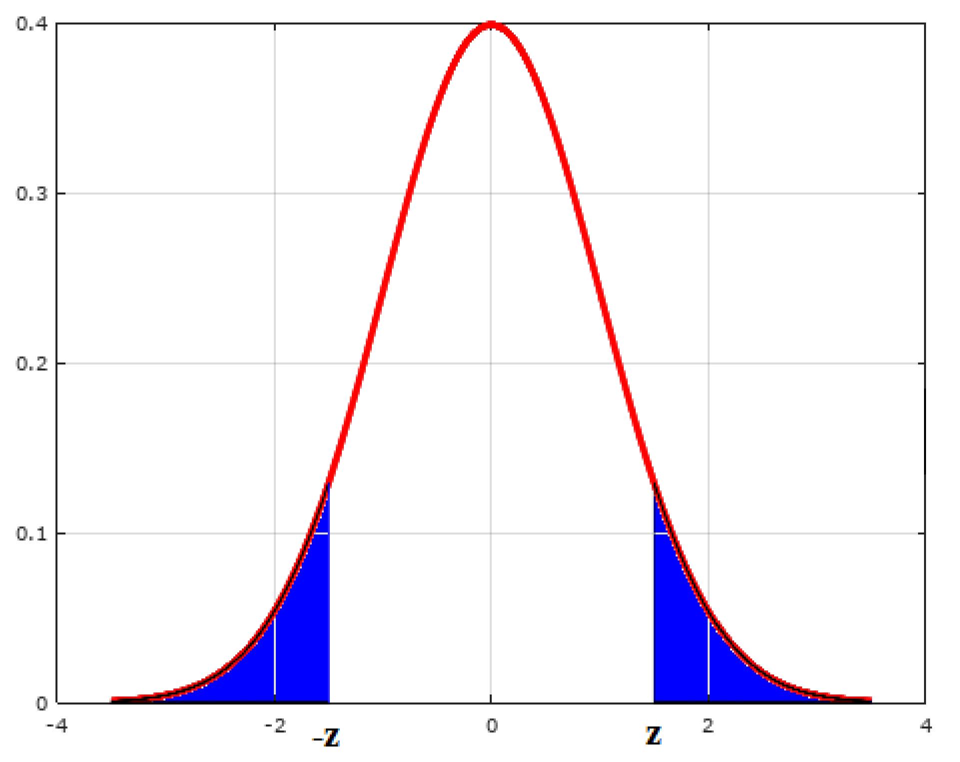
$$P\left(a<X<b\right)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

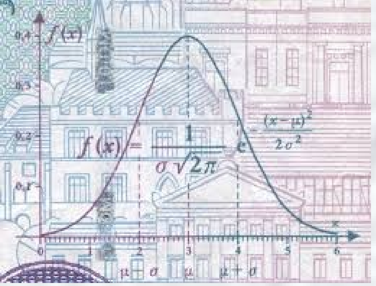
Esa tabla solo reseña valores de la función Φ para un conjunto de valores positivos. Esto es por una propiedad de esta función. Como la función densidad de la variable aleatoria normal estándar es par entonces el área a la izquierda de $-z$ $(z>0)$ y el área a la derecha de z son iguales. Esto es:

Con los valores de la función Φ para argumentos positivos alcanza.

Un resultado muy usado:

$$P\left(-a<Z<a\right)=\Phi\left(a\right)-\Phi\left(-a\right)=2\Phi\left(a\right)-1$$





$$p_r = P(-r < Z < r) = \Phi(r) - \Phi(-r) = 2\Phi(r) - 1$$

1	0.8413	0.6826
2	0.9773	0.9546
3	0.9987	0.9974

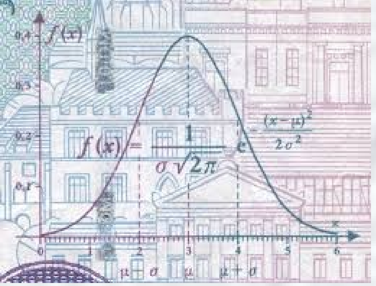
Veamos un ejemplo:

El volumen ***V*** que una máquina de llenado automático dispensa en latas de una bebida gaseosa tiene distribución normal de media 348 ml y desvío estándar 2.8 ml. Si se desechan todas las latas que tienen menos de 340 ml ó más de 350 ml de líquido, ¿cuál es la proporción *p* de latas desechadas?.

Para poder usar la función Φ hay que estandarizar los valores de la variable V.

$$p = \Phi\left(\frac{340-348}{2.8}\right) + 1 - \Phi\left(\frac{350-348}{2.8}\right) \approx \Phi(-2.88) + 1 - \Phi(0.71)$$

(24%)



Hay muchas situaciones en las que se conoce una probabilidad acumulada de un valor x de una variable aleatoria normal ***X***., esto es se conoce un valor de su función de distribución y se desea conocer cual es ese valor de x . Como la función de distribución normal es continua y estrictamente monótona entonces tiene función inversa y ese problema tiene solución única.

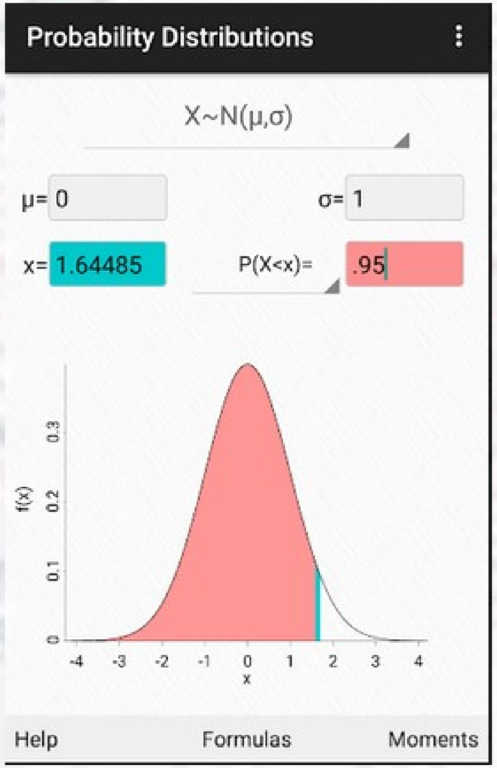
$$F_X(x_\alpha)=\alpha$$ se denomina el fractil de ***X.***

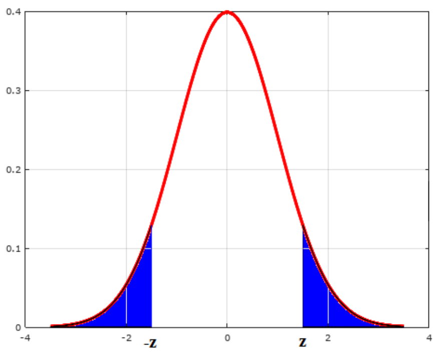
$$F_X(x_\alpha)=\Phi\left(\frac{x_\alpha-\mu}{\sigma}\right)=\alpha$$

$$Si\,\Phi(z_\alpha)=\alpha$$ se denomina el fractil de ***Z.***

Entonces se tiene:

se puede consultar en la misma tabla de la función de distribución normal estándar (tal vez se deba interpolar) o bien con un programa o aplicación conveniente. Por ejemplo:

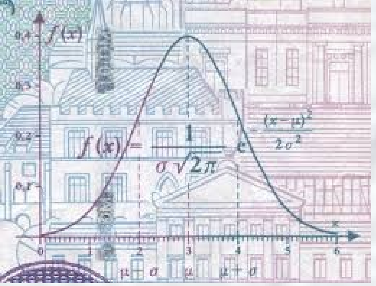




tiene una propiedad vinculada con la simetría ya comentada de la función densidad normal estándar:

$$z_\alpha=-z_{1-\alpha}$$

$$z_{0.95}\approx 1.6449 \qquad z_{0.05}\approx -1.6449$$



El volumen V que una máquina de llenado automático deposita en latas de una bebida gaseosa tiene distribución normal de media 348 ml y desvío estándar 2.8 ml.

a) Se desean rotular todas las latas con un volumen v de manera tal que el 90 % de todas las latas tengan como mínimo ese volumen. ¿Cuál es el valor de v?

b) La media de la operación de llenado puede ajustarse con facilidad, pero la desviación estándar sigue siendo de 2.8 ml. ¿Qué valor debe darse a la media para que el 95% de todas las latas contengan más de 345 ml.?

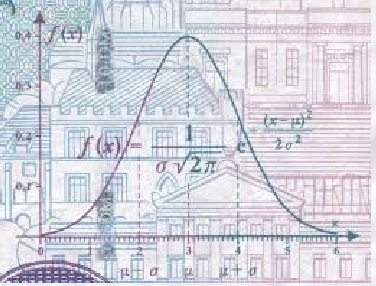
a) $P(V>v)=0.9 \rightarrow 1-\Phi\left(\frac{v-348}{2.8}\right)=0.9$ entonces se tiene: $\frac{v-348}{2.8}=z_{0.1}=-1.282$

$$v=344.4\,ml$$

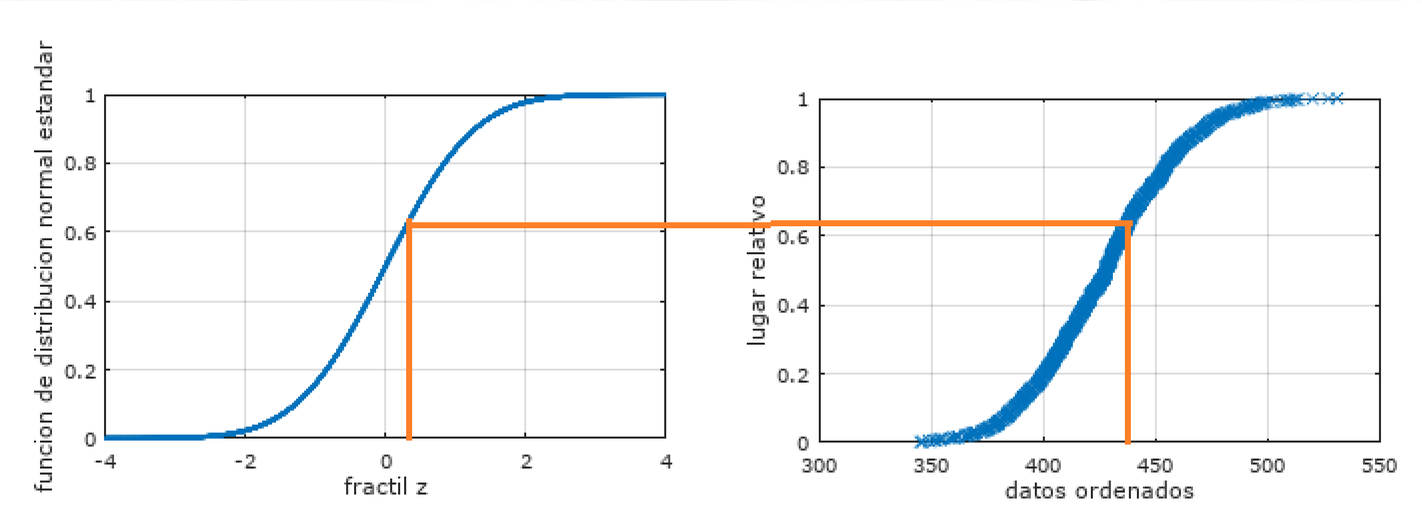
b) $P(V>345)=0.95 \rightarrow 1-\Phi\left(\frac{345-\mu}{2.8}\right)=0.95$ $\quad \frac{345-\mu}{2.8}=z_{0.05}=-1.645$

$$\mu=349.6\,ml$$

https://bookdown.org/aquintela/EBE/ejemplos-de-la-distribucion-normal.html (2/9/2020)

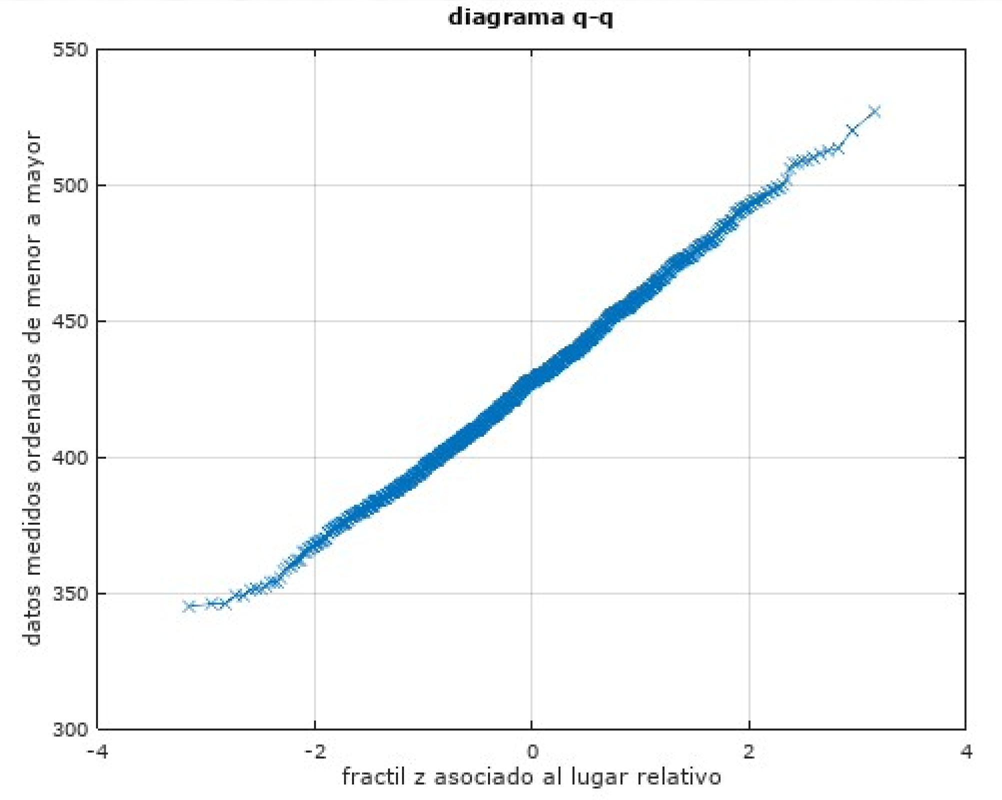


La **distribución normal** es muy importante ya que permite modelar muchos fenómenos naturales, sociales, fisiológicos y psicológicos. En muchas oportunidades se invoca la suposición de que un conjunto de datos proviene de haber medido una variable aleatoria normal. Veamos un gráfico (**qqplot**) que en una primera vista pueda dar plausibilidad a esa afirmación.



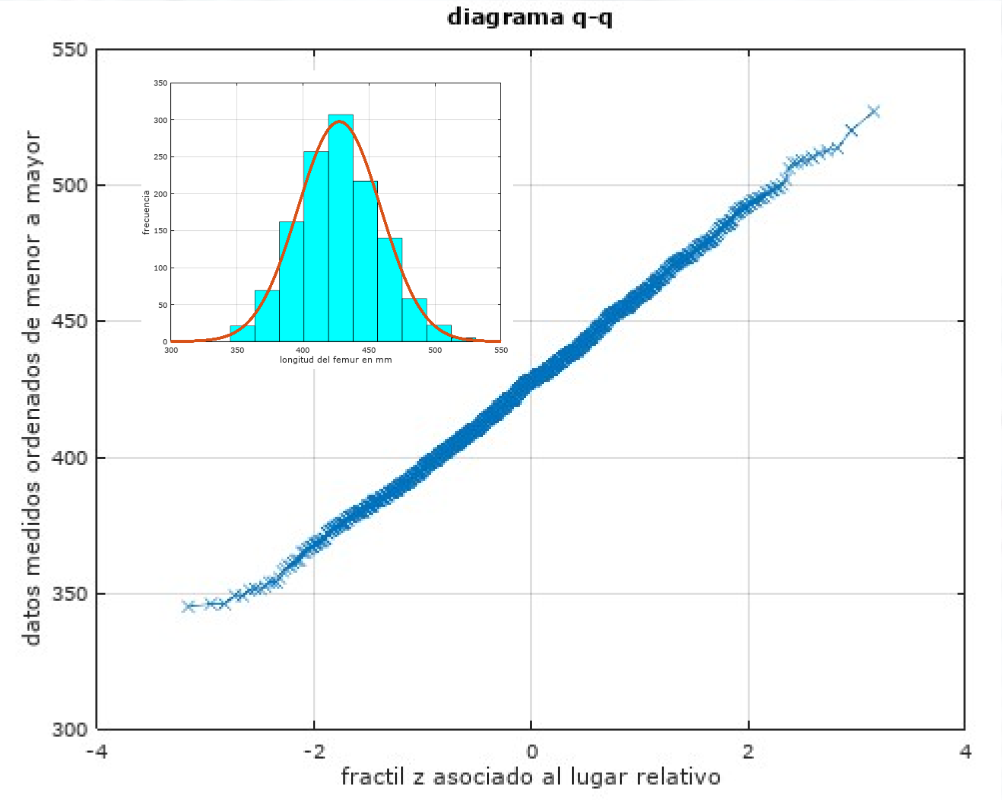
distribución normal estándar que se usa para comparar

de los datos: lugar relativo de un dato vs el dato en orden ascendiente. Dado un dato éste tiene un lugar relativo y le corresponde un fractil de la distribución con la que se compara.

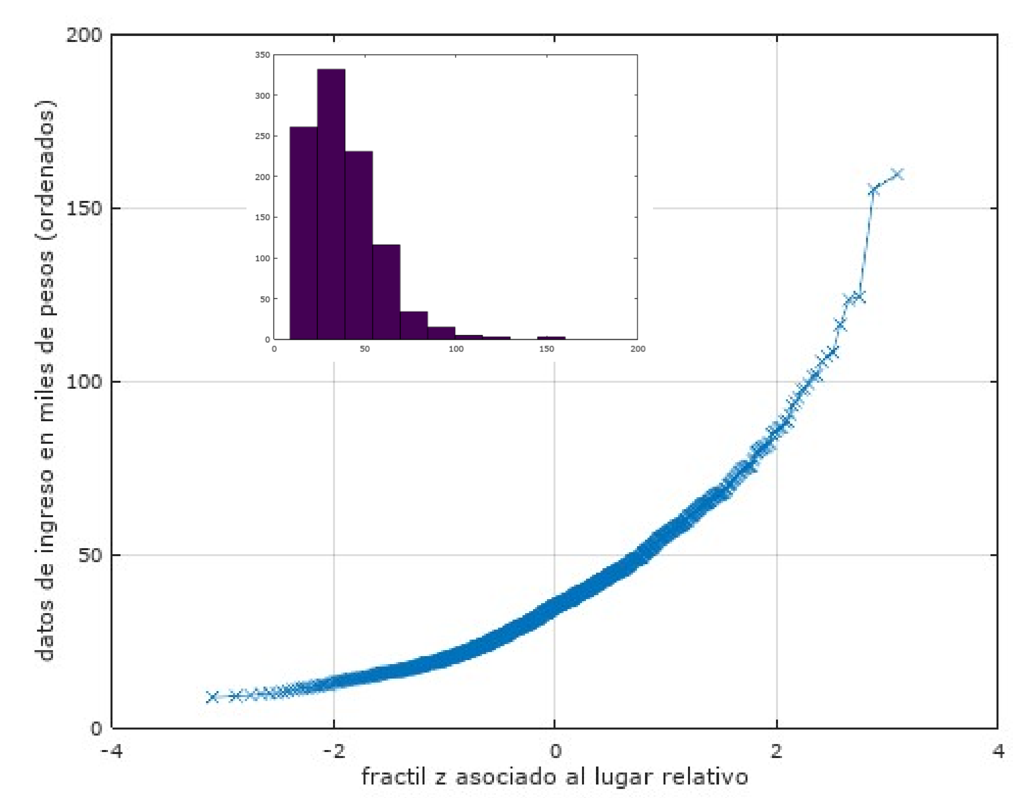


de la normal en este caso. Si están los puntos alineados entonces es indicio de normalidad en este caso.

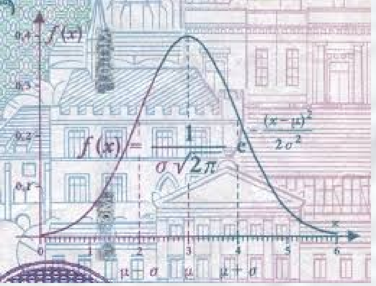
La **distribución normal** es muy importante ya que permite modelar muchos fenómenos naturales, sociales, fisiológicos y psicológicos. En muchas oportunidades se invoca la suposición de que un conjunto de datos proviene de haber medido una variable aleatoria normal. Veamos un gráfico (**qqplot**) que en una primera vista pueda dar plausibilidad a esa afirmación.



(se puede presuponer distribución normal).

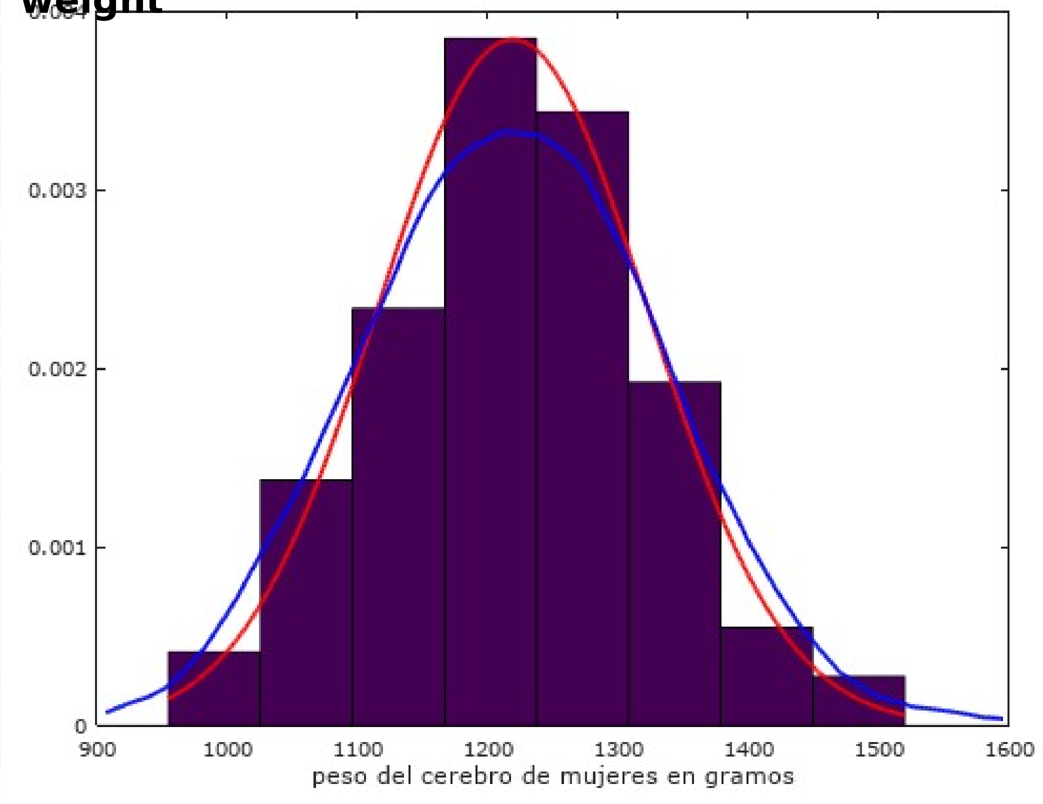


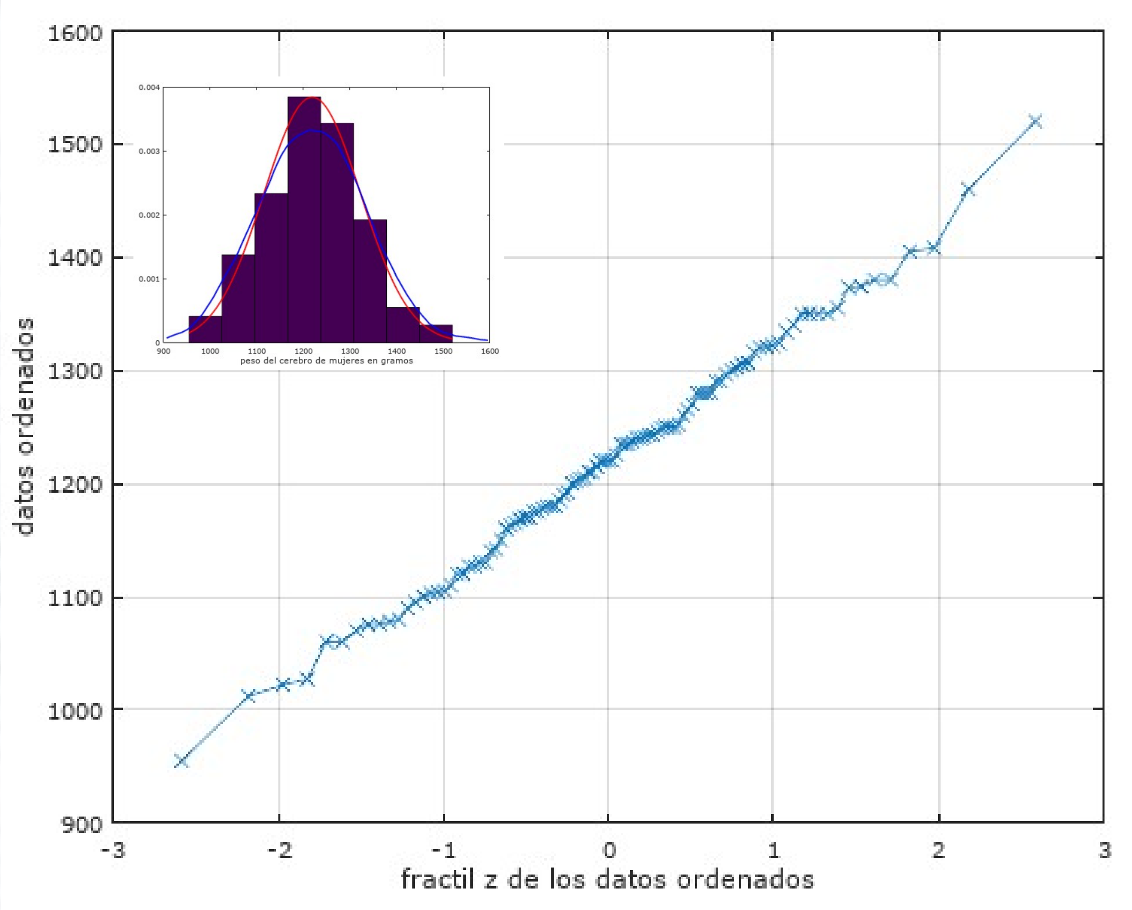
(NO se puede presuponer distribución normal).



La **distribución normal** es muy importante ya que permite modelar muchos fenómenos naturales, sociales, fisiológicos y psicológicos.

https://www.kaggle.com/rashmiek99/head-size-vs-brain-weight





103 mujeres. Histograma de 8 clases, la función densidad de probabilidad normal (para la misma media y desvío que los datos) en rojo y la curva de densidad
(kernel density plot) en azul (ajuste del histograma).