Biometría

Distribuciones de probabilidad para variables aleatorias continuas

Variables aleatorias continuas

Interesa estudiar la temperatura ambiente a las 12 hs en abril en la ciudad de Buenos Aires. Utilizando un termómetro se obtiene la siguiente medición:

19°C

Pero si se hubiera utilizado un instrumento con mayor precisión, se podría haber obtenido mayor información:

19,25°C

O mejor aún:

19,25852110289........ °C

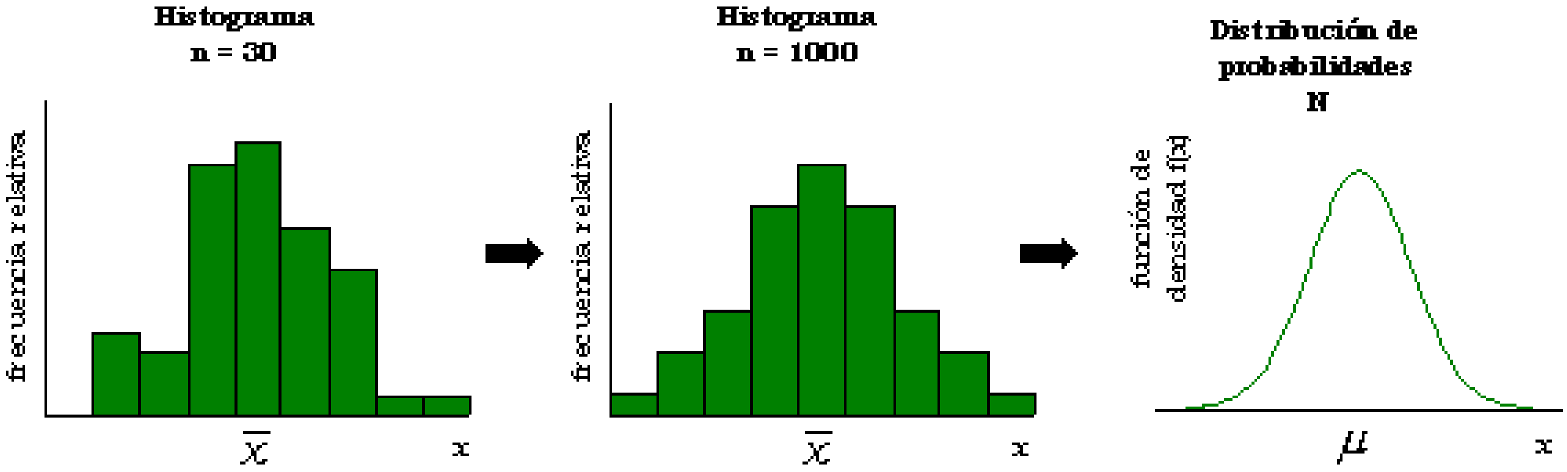
Variables aleatorias continuas

- Se trata de una V.A. continua, ya que admite infinitos valores

- La probabilidad de que un día cualquiera de abril la temperatura sea exactamente igual a 19,25852110289......... °C vale …

- También puede deducirse que sería imposible que dos valores de la variable sean exactamente iguales (en sus infinitos decimales)

Distribuciones de probabilidad

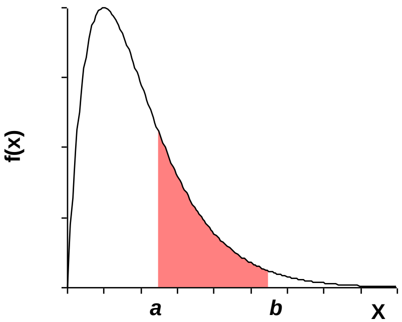


Función de densidad

- **función de densidad** f(x) es una función que describe la distribución de probabilidades de la variable aleatoria continua x.

- Una función de densidad debe cumplir con los siguientes requisitos:
  - $f(x) \geq 0$ para todo x
  - el área total bajo la curva **= 1**

Cálculo de probabilidades



- La probabilidad de que un evento ocurra en un intervalo $[a,b]$ es el área bajo la curva de la función en ese intervalo
- Se calcula utilizando integrales:

$$P(a < x < b) = P(x \in (a,b)) = \int_a^b f(x)\,dx$$

- Por esto, la probabilidad de que la V.A. tome un valor exacto es nula:

$$P(x = a) = \int_a^a f(x)\,dx = 0$$

- Se define la **función de distribución de la v.a. continua**, F(x), como la probabilidad de que *X* sea menor o igual que a

$$F[x] = \int_{-\infty}^{a} f[x]\,dx$$

Esperanza y varianza

V.A. DISCRETAS	V.A.CONTINUAS
$E(x) = \Sigma x_i p_i$	$E(x) = \int_{-\infty}^{\infty} x \ f_{(x)} dx$
$\sigma^2(x) = \Sigma (x_i - \mu)^2 p_i$	$\sigma^2(x) = \int_{-\infty}^{\infty} (x - \mu_{(x)})^2 \ f_{(x)} dx$

Propiedades de la esperanza o media y de la varianza

- Sea k una constante cualquiera. Entonces:

$$E(k) = k$$
$$\sigma^2(k) = 0$$

- Si se suma k a cada uno de los valores de una variable aleatoria X entonces:

$$E(X + k) = E(X) + k$$
$$\sigma^2(X + k) = \sigma^2(X)$$

- Si se multiplica por k a cada uno de los valores de una variable aleatoria X entonces:

$$E(X \, . \, k) = E(X) \, . \, k$$
$$\sigma^2(X \, . \, k) = \sigma^2(X) \, . \, k^2$$

Propiedades de la esperanza o media y de la varianza

Sean dos variables aleatorias independientes *X1* y *X2*, siendo sus esperanzas *E(X1)* y *E(X2)* y sus varianzas $\sigma^2(X1)$ y $\sigma^2(X2)$ respectivamente

- Si se define la VA *X* + *X* entonces

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$\sigma^2(X_1 + X_2) = \sigma^2(X_1) + \sigma^2(X_2)$$

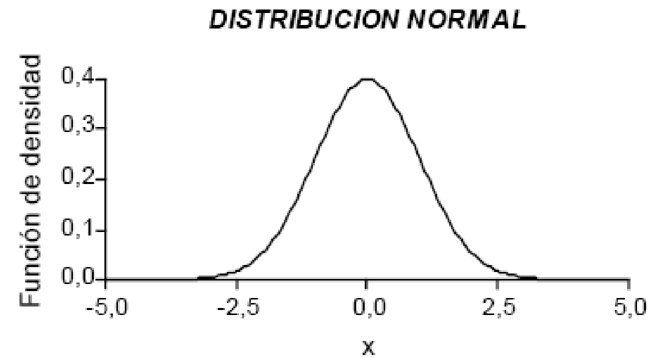
- Si se define la VA *X1* - *X2* entonces

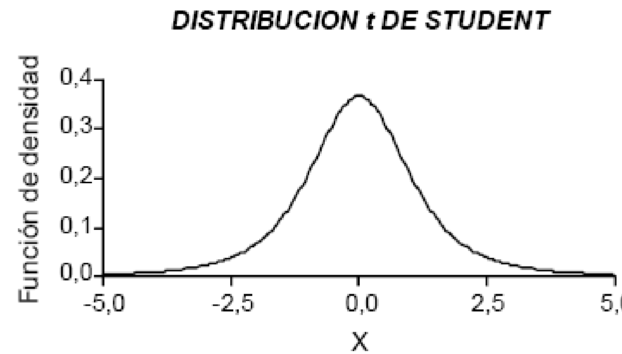
$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

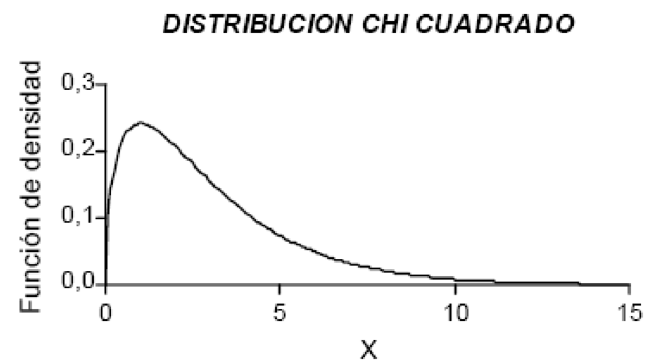
$$\sigma^2(X_1 - X_2) = \sigma^2(X_1) + \sigma^2(X_2)$$

Modelos

- Existen numerosos modelos de distribución de probabilidades para V.A. continuas
- Nosotros veremos:
  - la distribución normal
  - la distribución t de Student
  - la distribución chi-cuadrado



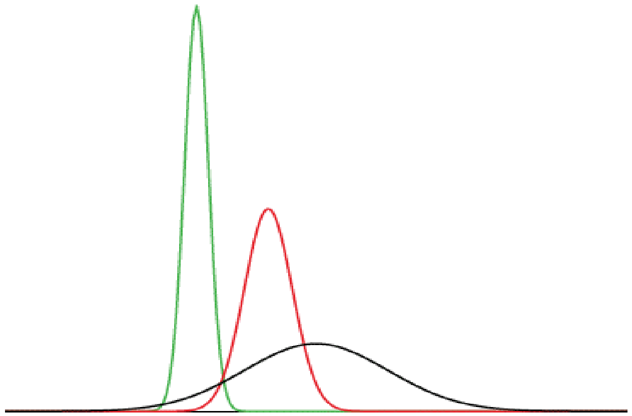




Distribución normal o de Gauss

- Aplica a Variables aleatorias continuas
- Distribución simétrica y unimodal
- Media, mediana y moda coinciden
- Asintótica al eje de las x
- Dominio de la variable es:

$$-\infty < x < \infty$$

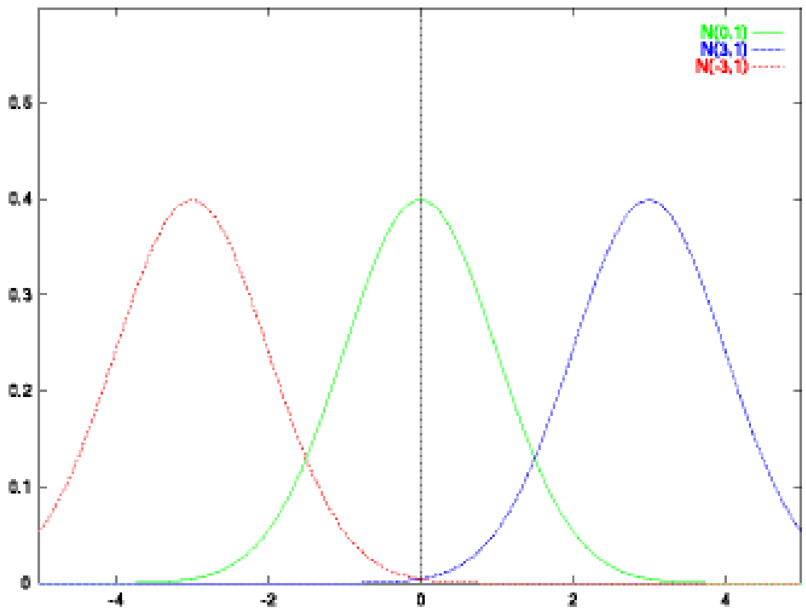


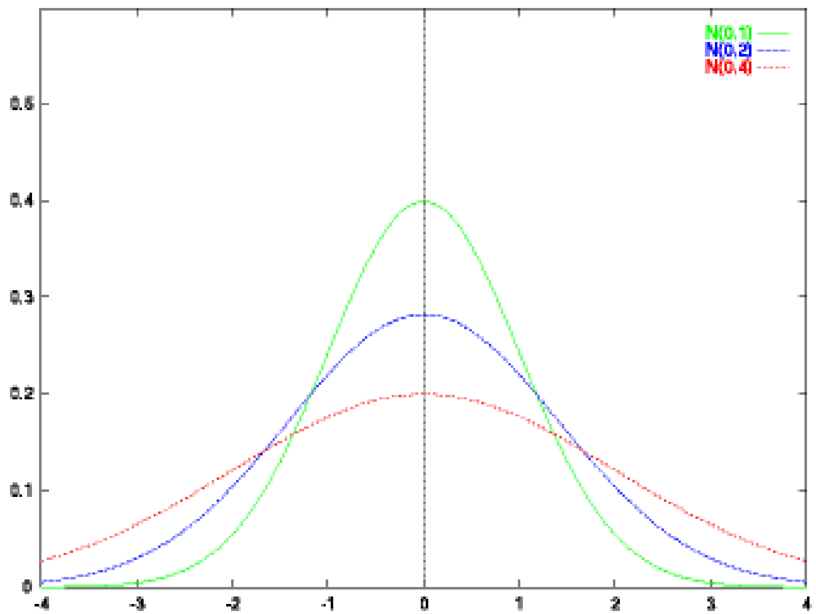
- Aparece de manera natural:
  - Altura, peso
  - Error de medición
  - En procesos donde la variable es el resultado de la acción de muchos pequeños efectos

Los parámetros de la distribución normal

Está caracterizada por dos parámetros:

- la media μ localizada en el centro de la distribución
- el desvío estándar σ localizado en cada punto de inflexión.





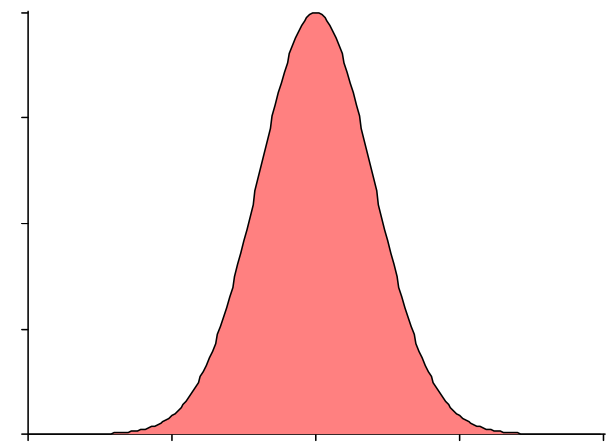
Función de densidad

La función de densidad que describe el comportamiento de la variables es:

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

donde e= 2,7183.. ∏ = 3,1416…
y μ y σ son los parámetros de la distribución

Campana de Gauss

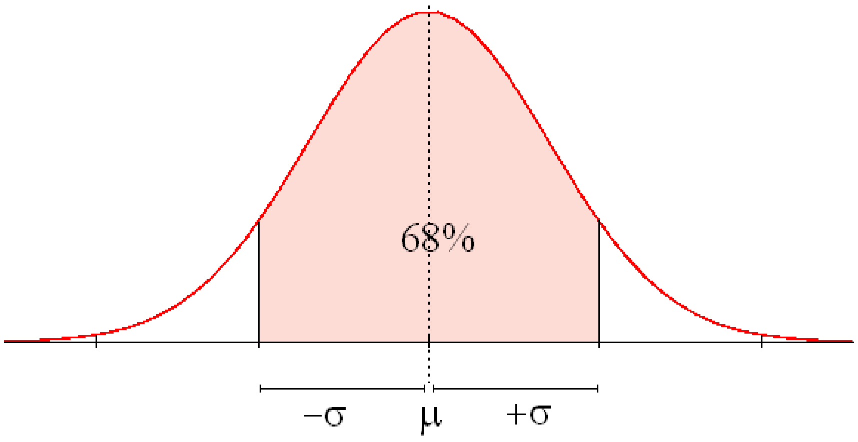


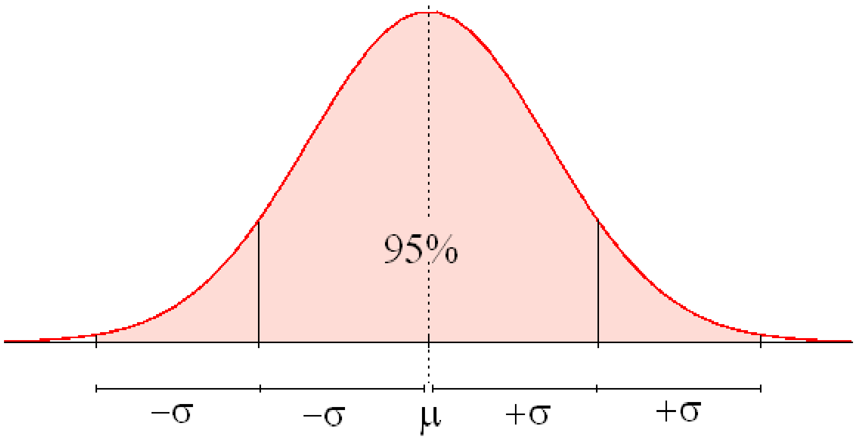
Presentada por Abraham de Moivre 1733

Formulada por:
Carl Friedrich Gauss (1777-1855)

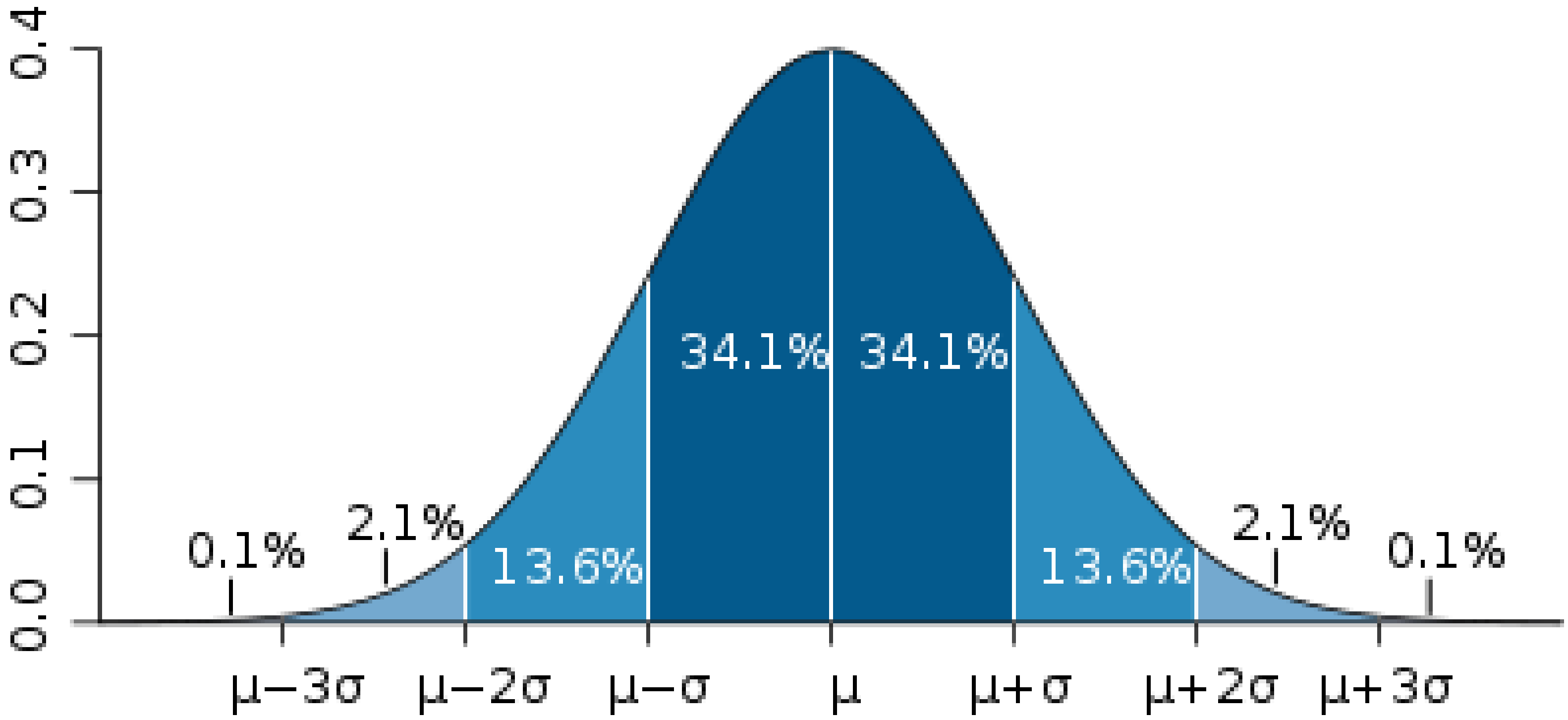
La regla 68-95-99.7

- Entre la media ± un desvío estándar tenemos siempre la misma probabilidad: aprox. 68%
- Entre la media ± dos desvíos estándar: aprox. 95%
- Entre la media ± tres desvíos estándar aprox. 99.7%





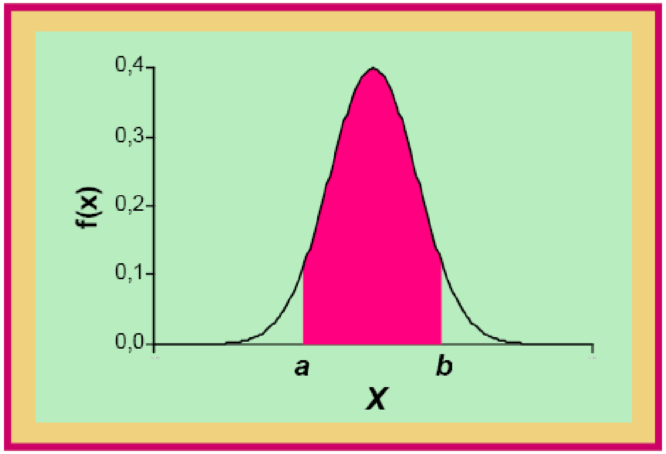
La regla 68-95-99.7



Cálculo de probabilidades

- Las probabilidades son áreas bajo la curva de f(x)

$$P(a < x < b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$



- El área total bajo la curva vale 1.

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

- Están tabuladas las áreas (integrales) para una dada variable normal, denominada normal estándar

Estandarización

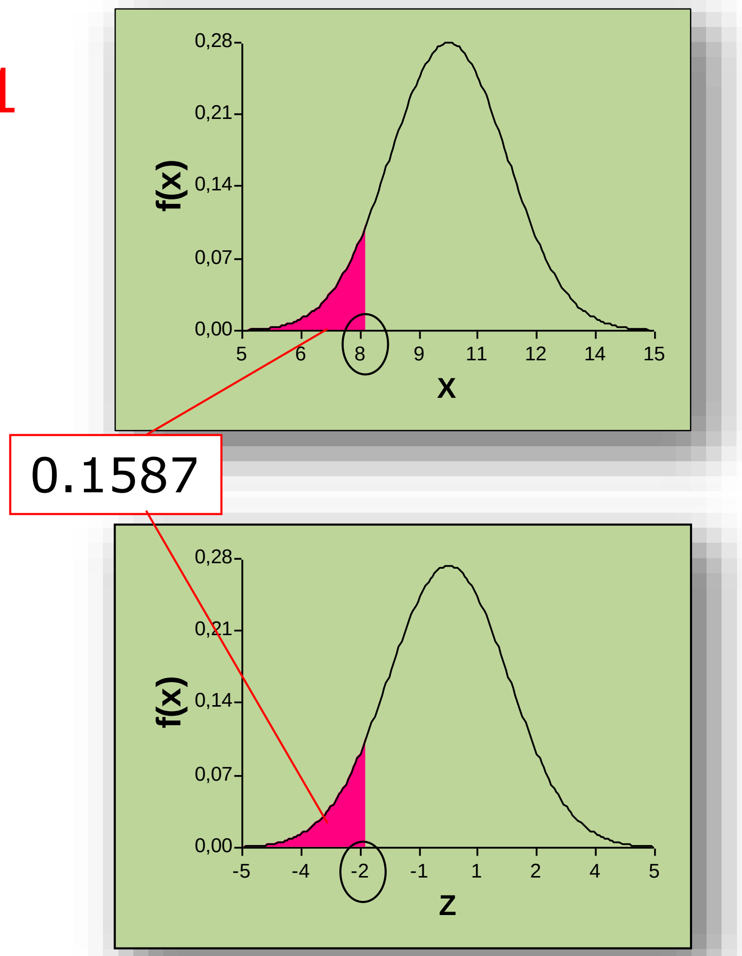
- La variable en estudio debe ser transformada (reescalada) en esta variable normal estándar
- Dada una variable de media μ y desviación estándar σ, se denomina valor estandarizado o valor **Z**, de una observación x, a la distancia (con signo) con respecto a la media, medido en desviaciones estándar, es decir

$$z = \frac{x - \mu}{\sigma}$$

- Cuanto más grande sea el valor de Z, más lejos estará el valor de la media.
- es un valor sin dimensiones, y por lo tanto es una medida útil para comparar valores de datos de dos poblaciones distintas, para saber cuál de los dos es más extremo.

La distribución normal estándar

- Media = 0; Desvío estandar = 1
- Cuando $x = \mu$, $z = 0$
- Simétrica con respecto a $z = 0$
- Valores de z a la izquierda del centro son negativos
- Valores de z a la derecha del centro son positivos
- El área total bajo la curva es 1.
- Áreas en Tabla



Ejemplo

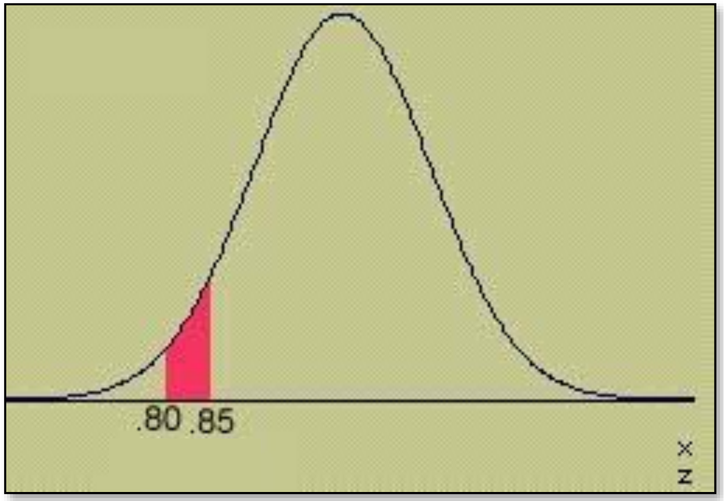
- El diámetro de las matas de la forrajera *Poa pratensis* se distribuye normalmente con un promedio de 1 dm y un desvío estándar de 0.1 dm.

- ¿Cuál es la probabilidad de que un ejemplar elegido al azar mida entre 0.8 y 0.85 dm?

$$P(0.80 < x < 0.85) =$$

$$P(-2 < z < -1.5) =$$

$$0.0668 - 0.0228 = 0.0440$$



Ejemplo

- El diámetro de las matas de la forrajera *Poa pratensis* se distribuye normalmente con un promedio de 1 dm y un desvío estándar de 0.1 dm.

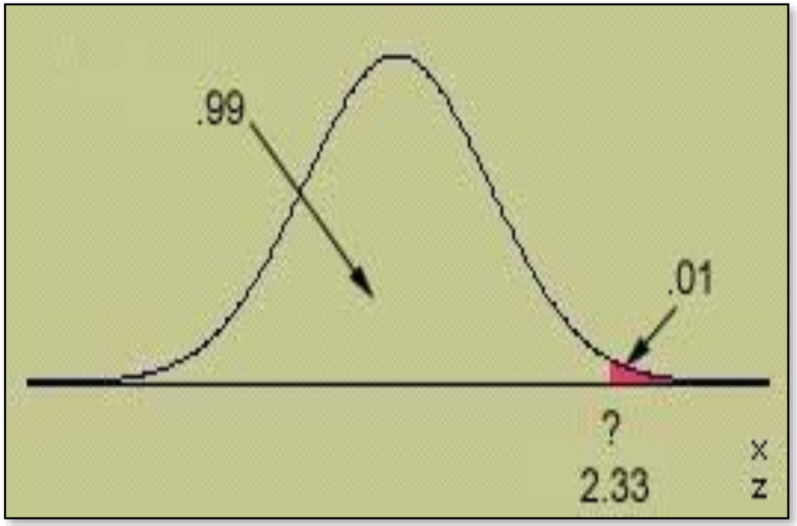
- ¿Cuál es el diámetro tal que sólo el 1% de las matas excede dicho valor?

$$P(x > ?) = 0.01$$

$$P(x < k) = 0.99$$

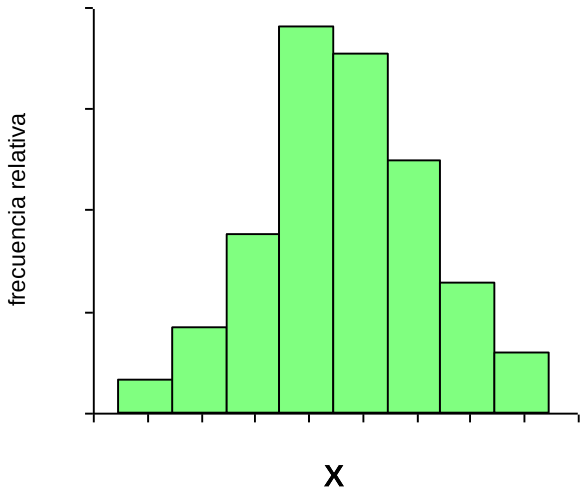
$$z = 2.33 = \frac{x - \mu}{\sigma} = \frac{x - 1}{0.1}$$

$$x = 1.233\,dm$$

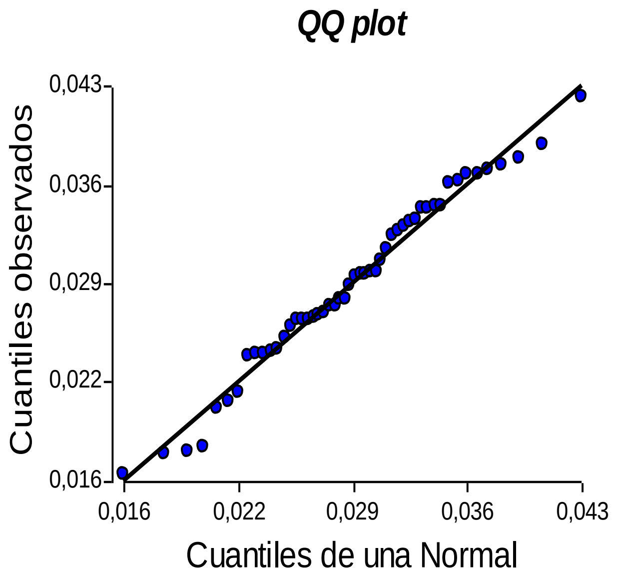


¿Cómo saber si una variable tiene distribución normal?

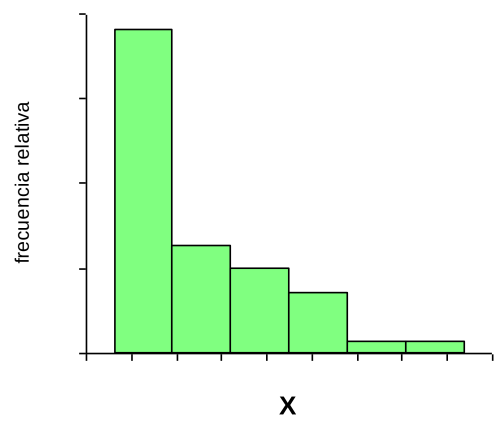
- Métodos analíticos
  - Coeficiente de asimetría g_1
  - Coeficiente de kurtosis g_2
  - Otros
- Métodos gráficos
  - Q-Q plot
  - otros



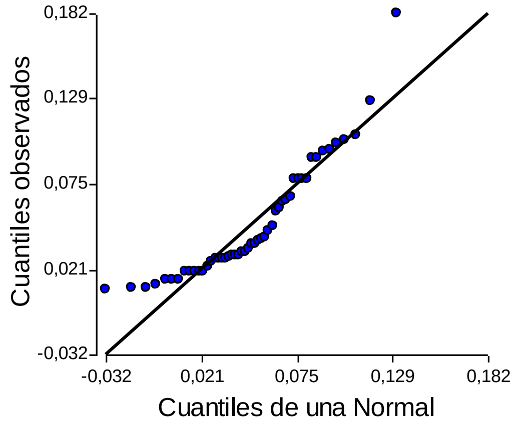
$$g_1 \approx 0$$
$$g_2 \approx 0$$

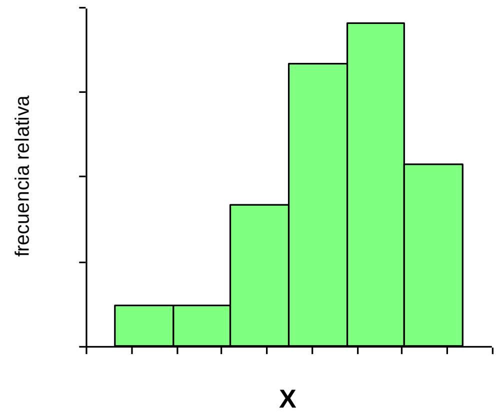


Desviaciones de la normalidad

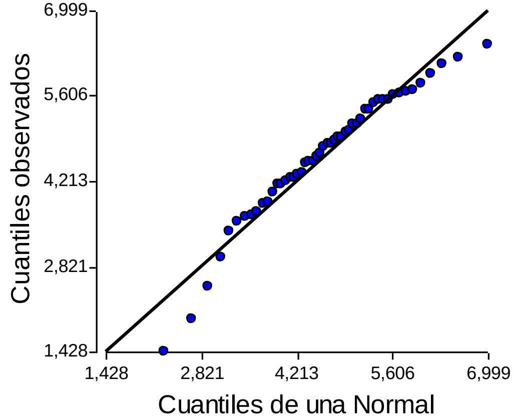


$$g_1 > 0$$





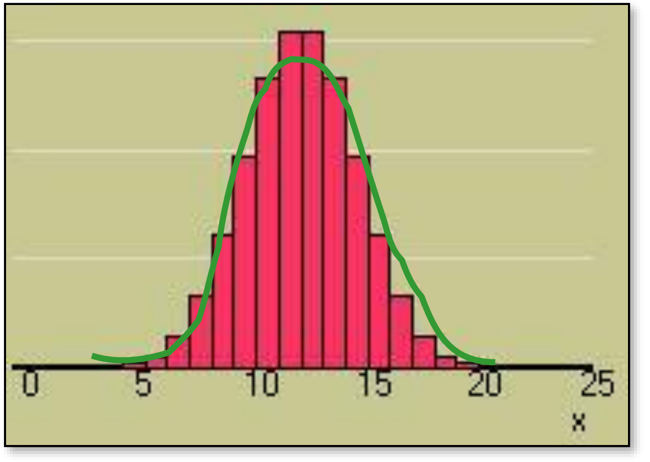
$$g_1 < 0$$



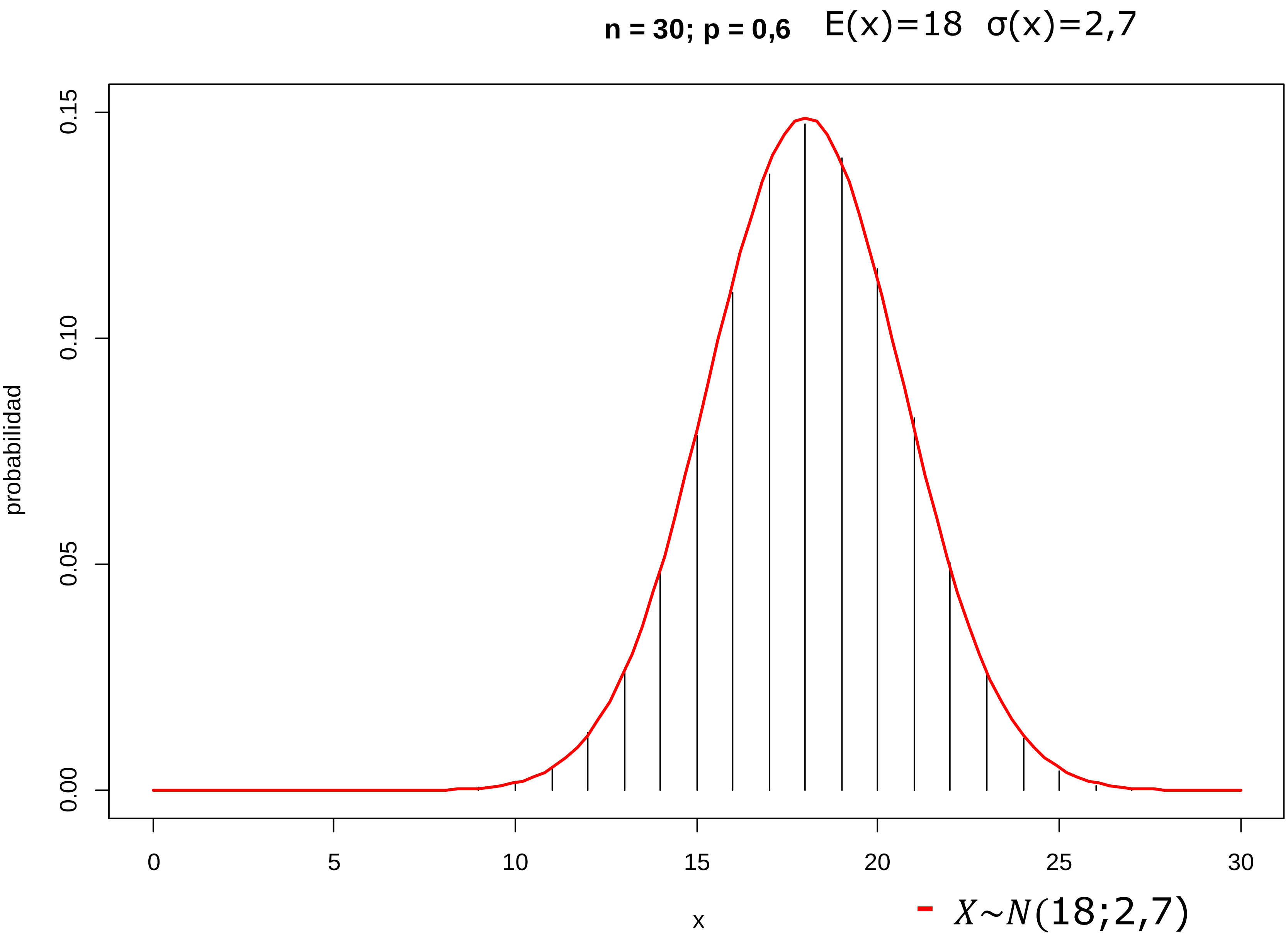
Aproximación de la binomial a la normal

- El cálculo de probabilidades en la distribución binomial se complica cuando n es grande
- Se puede recurrir a
  - Software
  - Aproximaciones
- Cuando n es grande ($n>20$) y p no demasiado pequeño ($np>5$) ni grande ($nq>5$), la distribución binomial se aproxima a una distribución normal con media np y varianza npq

- Cuando p es muy pequeño o muy grande es mejor usar la aproximación del modelo de Poisson.

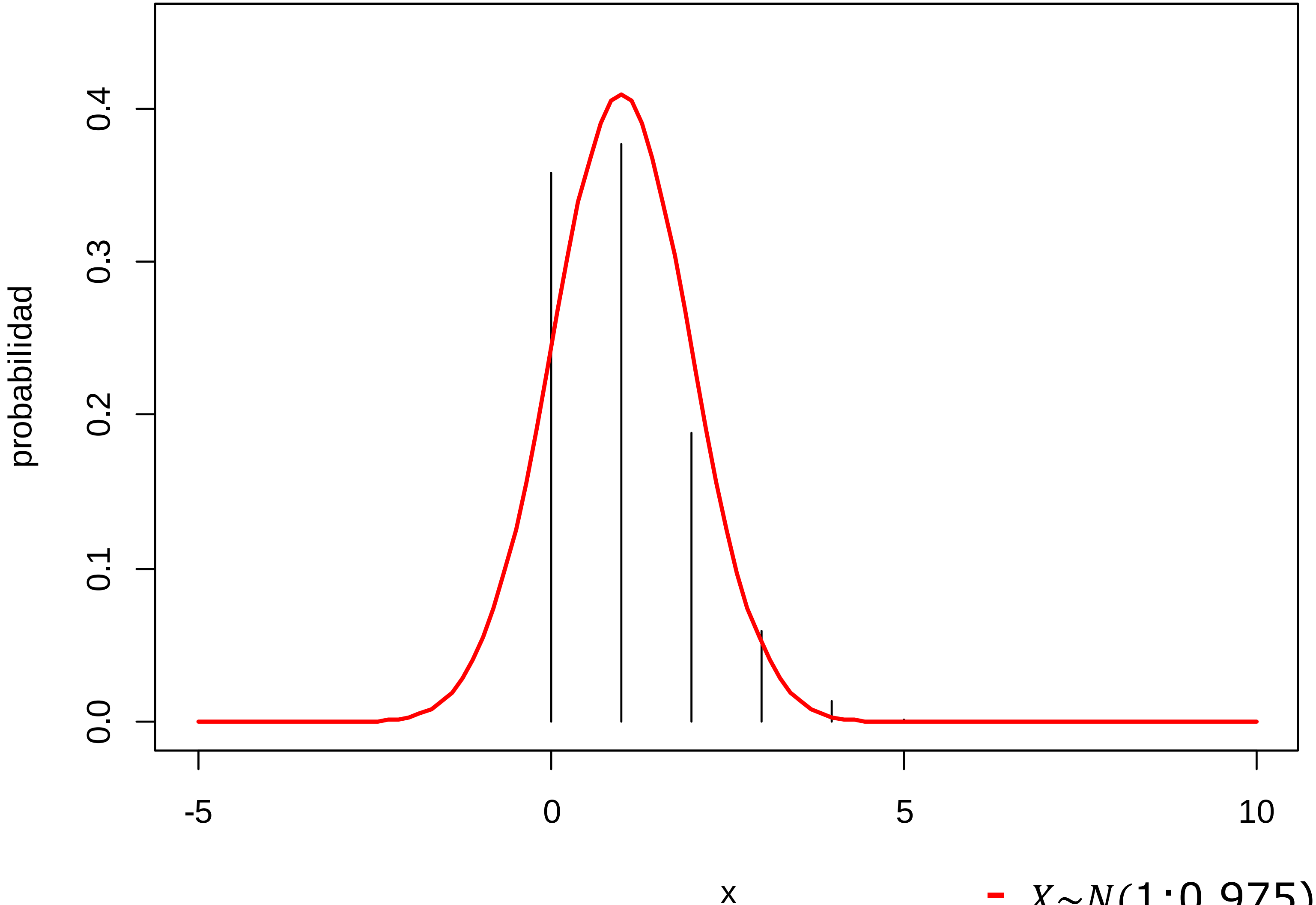


Una distribución Binomial con E(X)>5
Se aproxima bien a una distribución Normal



Una distribución Binomial con E(X)<5
No se aproxima a una distribución Normal

n= 20; p=0.05 E(x)=1 σ(x)=0,975



$X \sim N(1;0{,}975)$

Una distribución Binomial con E(X)<5
Se aproxima mejor a una distribución Poisson

n= 20; p=0.05 **E(x)=1 σ(x)=0,975**