STATISTICA

I modelli probabilistici, II

Dal campione alla **popolazione**

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g}/\mathrm{L}$.

Sapendo che la legge impone una soglia massima per l’arsenico di 10 $\mu\mathrm{g}/\mathrm{L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Quale tipo di modello probabilistico
per la quantità di arsenico
nell’acqua di Milano?

Dal campione alla **popolazione**

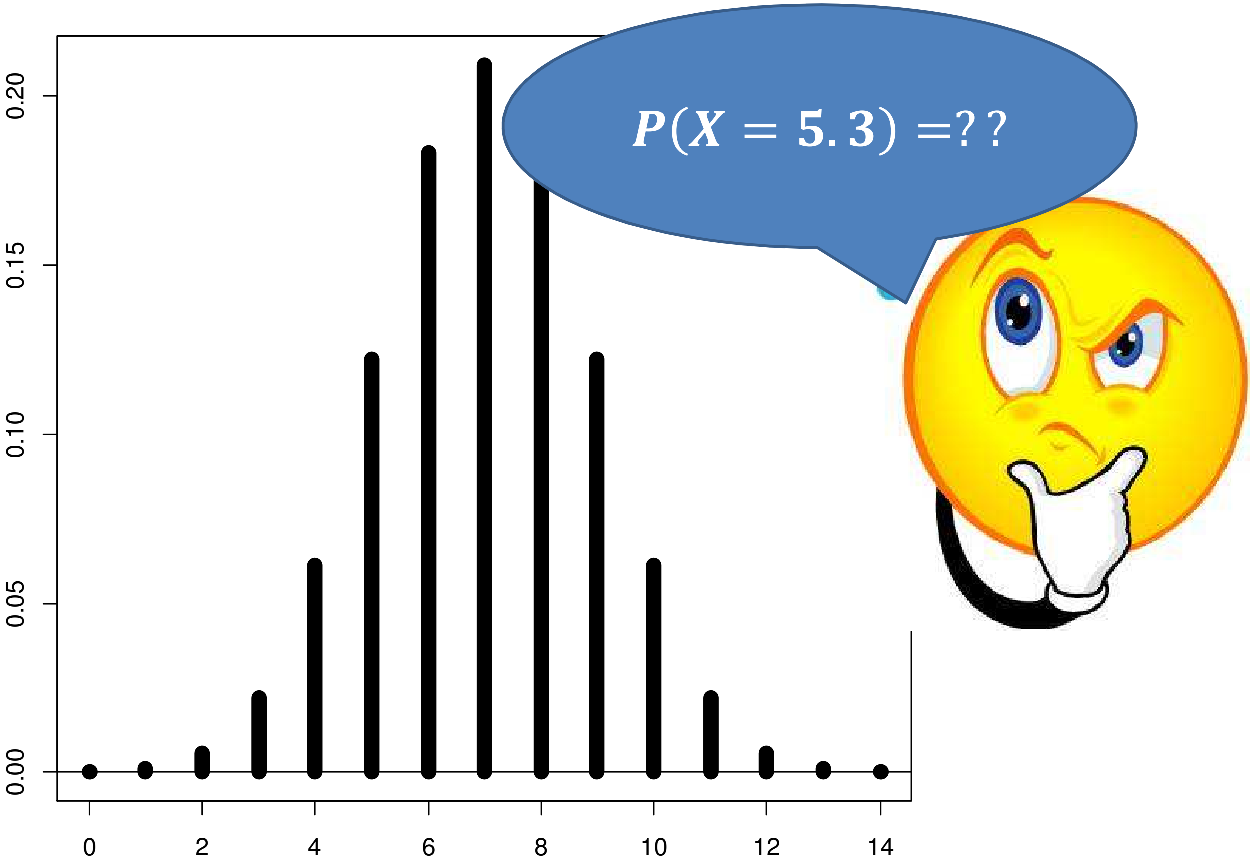
X sia la variabile che dice quanto arsenico c'è in un campione di acqua milanese, in $\mu\mathrm{g/L}$.

$$\boldsymbol{X}: \quad (\cdots) \rightarrow (\mathbf{0}, +\infty) \quad \& \quad \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = ???$$

Dal campione alla **popolazione**

X sia la variabile che dice quanto arsenico c'è in un campione di acqua milanese, in $\mu\text{g/L}$.

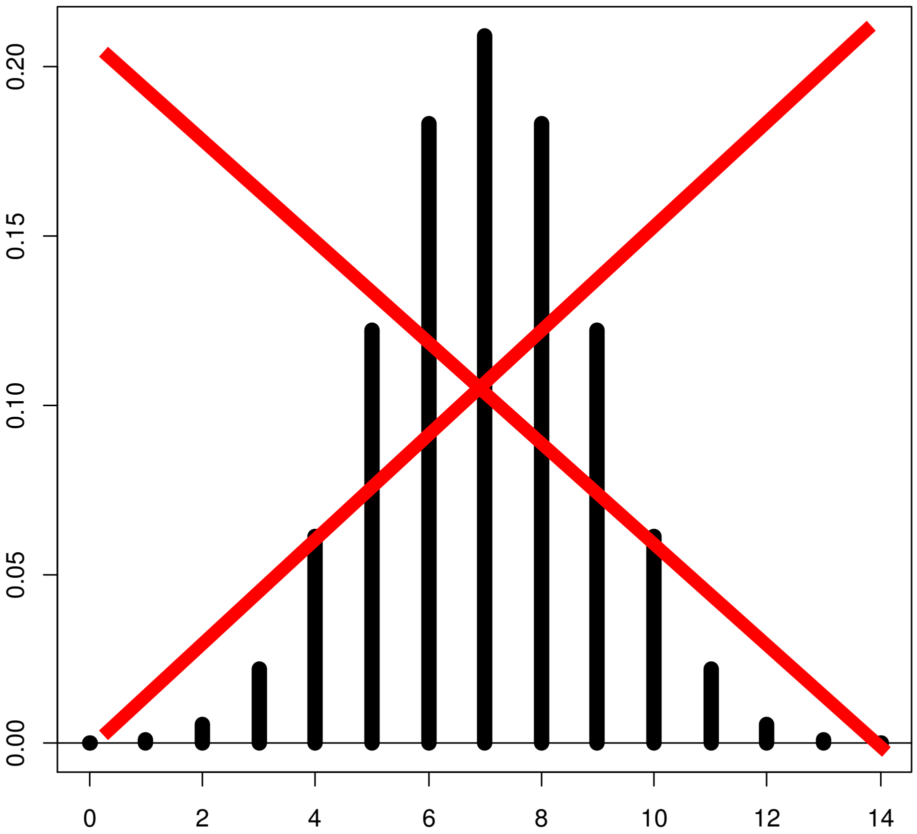
$$\boldsymbol{X}: \quad (\cdots) \rightarrow (\mathbf{0}, +\infty) \quad \& \quad \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = ???$$

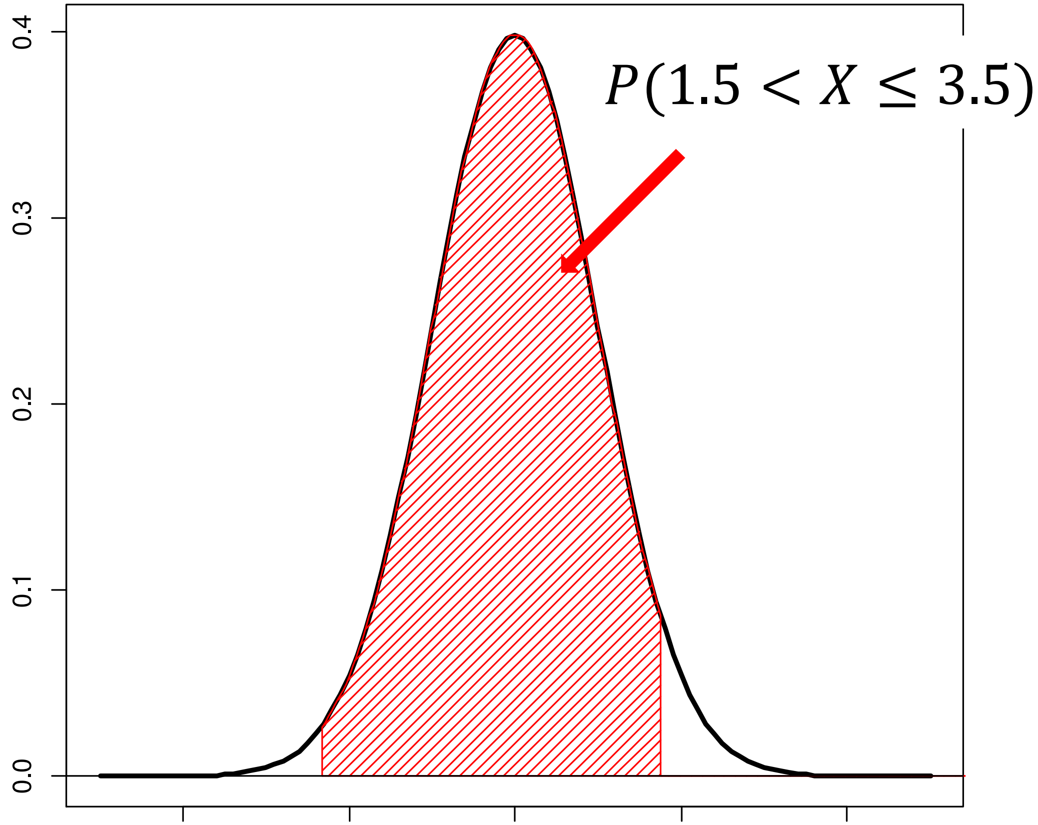


Variabili continue

X sia la variabile che dice quanto arsenico c'è in un campione di acqua milanese, in μg/L.

$$\boldsymbol{X}: \quad (\cdots) \to (\mathbf{0}, +\infty) \quad \& \quad \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = ???$$

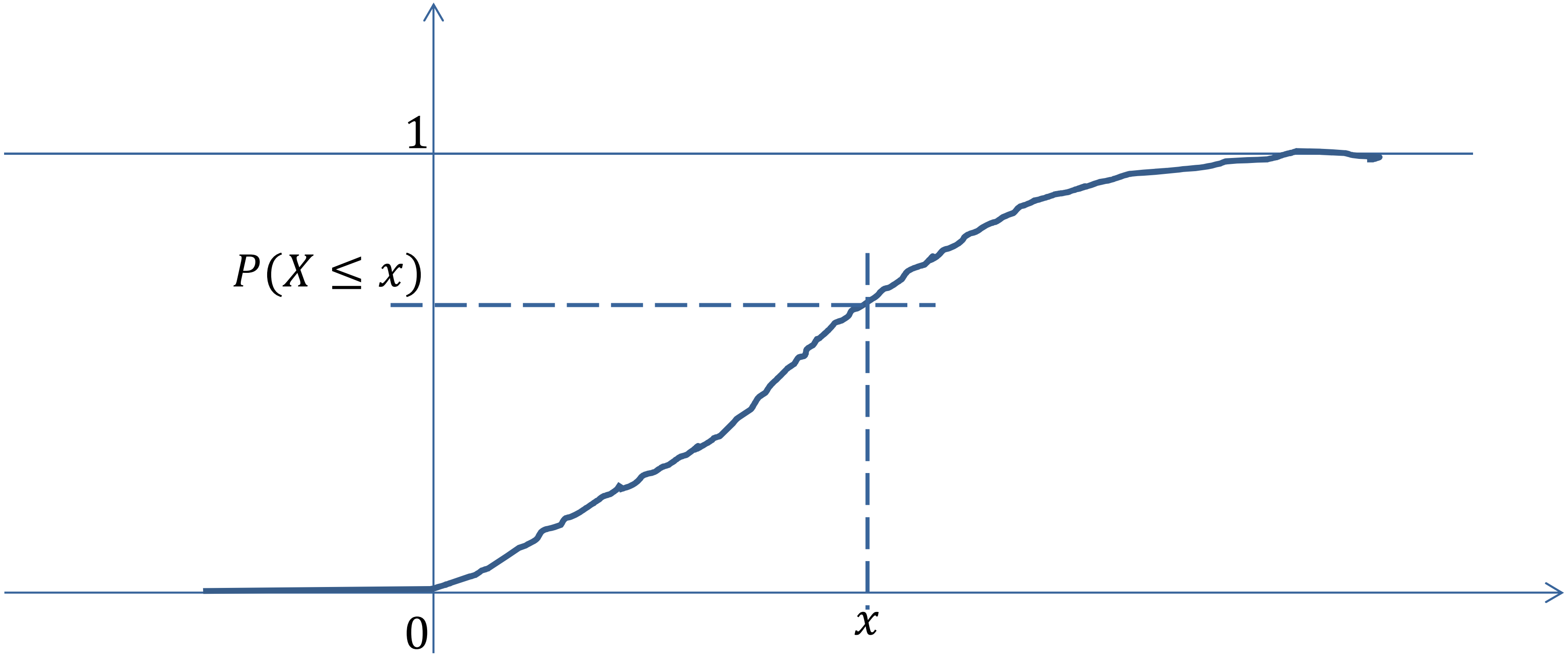




Variabili **continue**

$$\boldsymbol{X}: \quad (\cdots) \to \mathbb{R}$$

distribuzione: $P(X \leq x)$ **per ogni valore di** $x \in \mathbb{R}$



Variabili **continue**

$$\boldsymbol{X} : \quad (\cdots) \to \mathbb{R}$$

distribuzione: $\boldsymbol{P(X \leq x)}$ **per ogni valore di** $\boldsymbol{x} \in \mathbb{R}$

densità

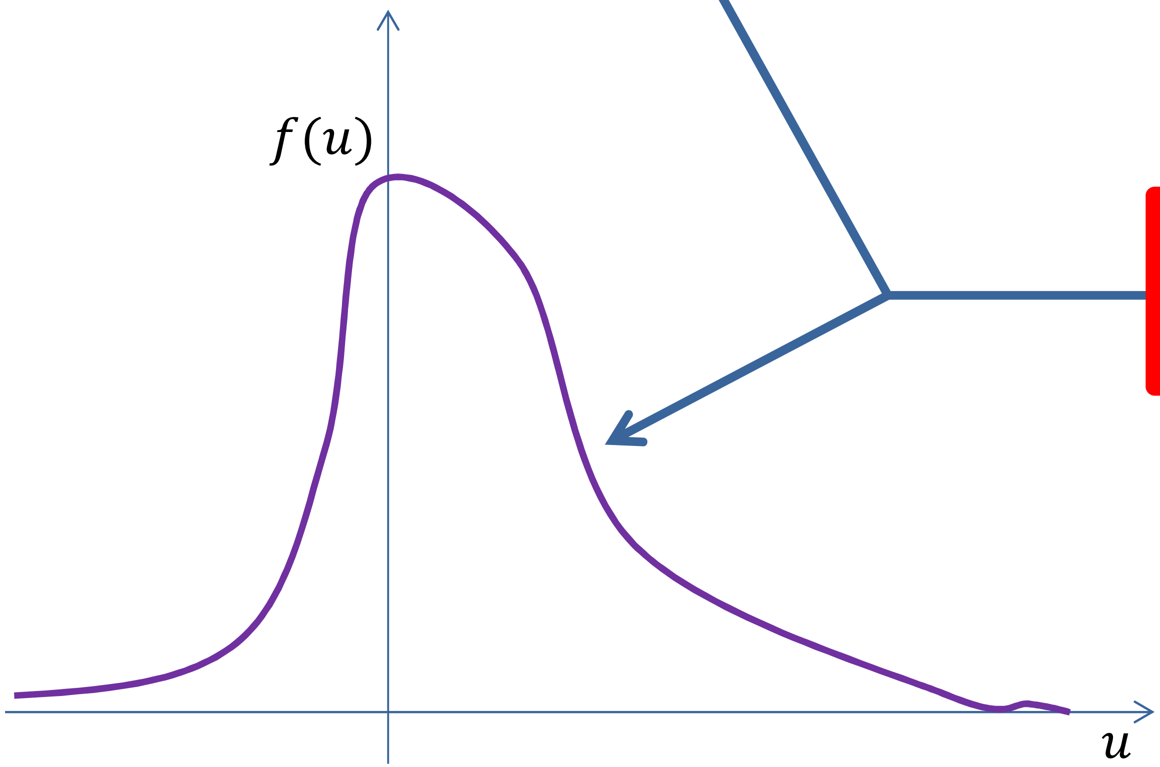
$$\boldsymbol{f(u) > 0}$$

area sotto la curva = 1

$$\boxed{P(X \leq x) = \int_{-\infty}^{x} f(u)du}$$

$$P(a < X \leq b) = \int_{a}^{b} f(u)du$$

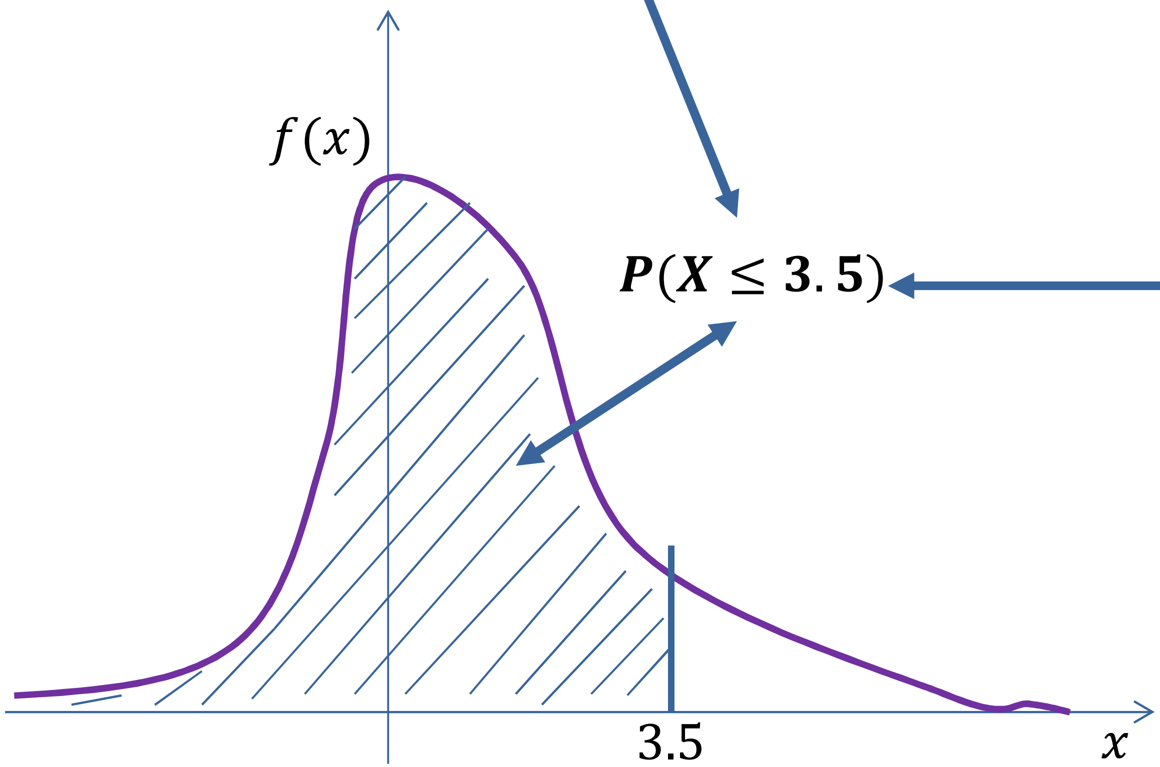
$$P(X = x) = 0$$



Variabili **continue**

$$\boldsymbol{X}: \quad (\cdots) \to \mathbb{R}$$

distribuzione: $\boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{x})$ **per ogni valore di** $\boldsymbol{x}$



$\boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{3.5})$

densità

$$f(u) > 0$$

area sotto la curva = 1

$$\boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{u})\boldsymbol{du}$$

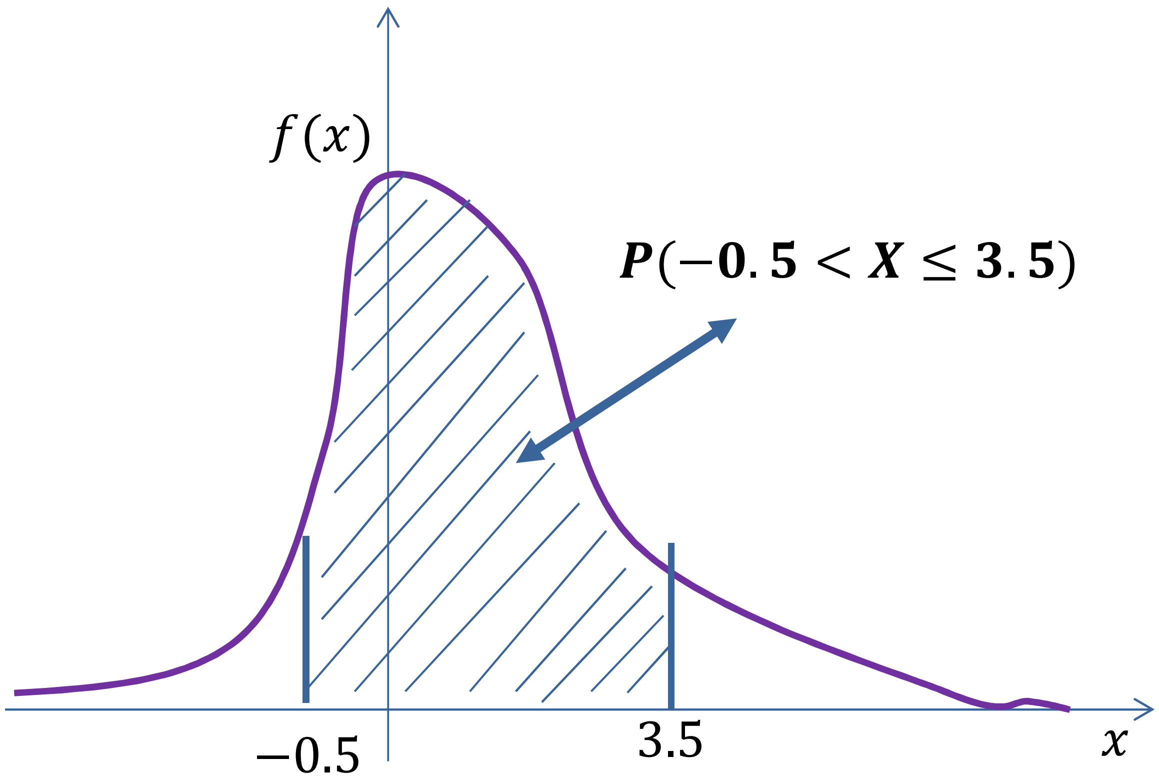
$$P(a < X \leq b) = \int_a^b f(u)du$$

$$P(X = x) = 0$$

Variabili **continue**

$$\boldsymbol{X} : \quad (\cdots) \to \mathbb{R}$$

distribuzione: $\boldsymbol{P(X \leq x)}$ **per ogni valore di $\boldsymbol{x}$**



densità

$$f(u) > 0$$

area sotto la curva = 1

$$P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

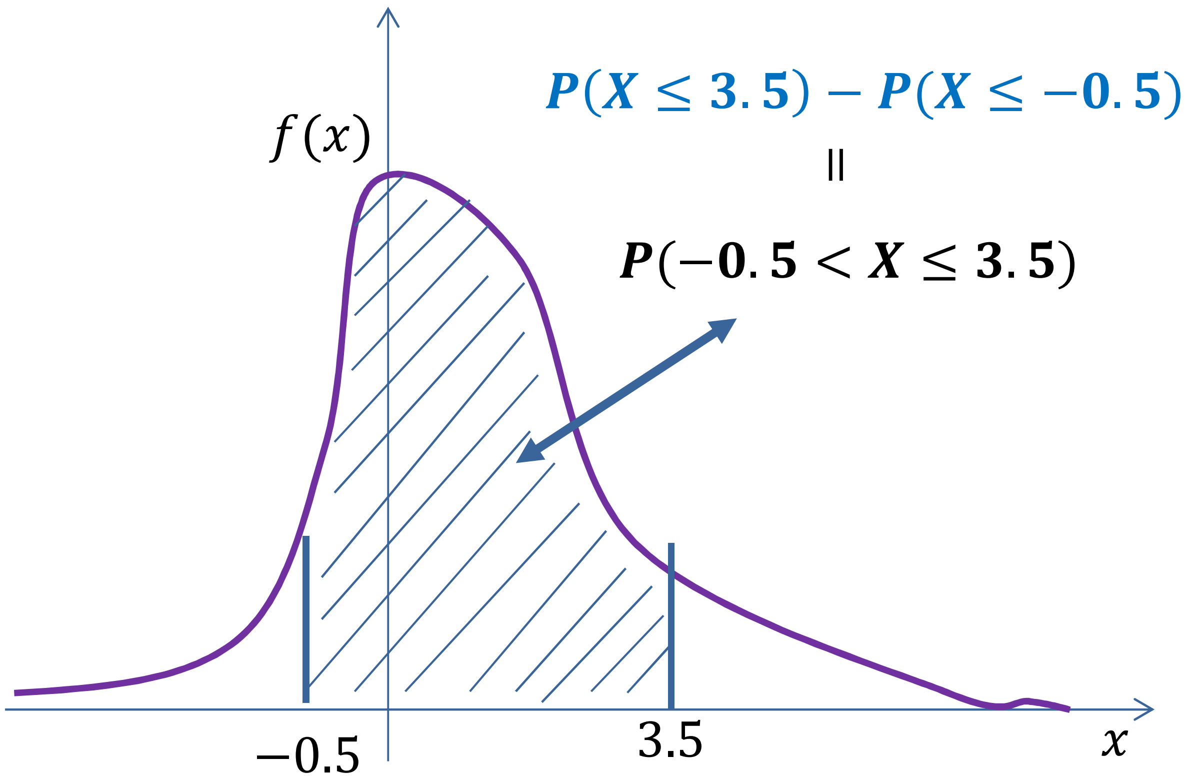
$$\boldsymbol{P(a < X \leq b) = \int_{a}^{b} f(u)du}$$

$$P(X = x) = 0$$

Variabili **continue**

$$\boldsymbol{X}: \quad (\cdots) \to \mathbb{R}$$

distribuzione: $\boldsymbol{P(X \leq x)}$ **per ogni valore di x**



densità

$$f(u) > 0$$

area sotto la curva = 1

$$P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

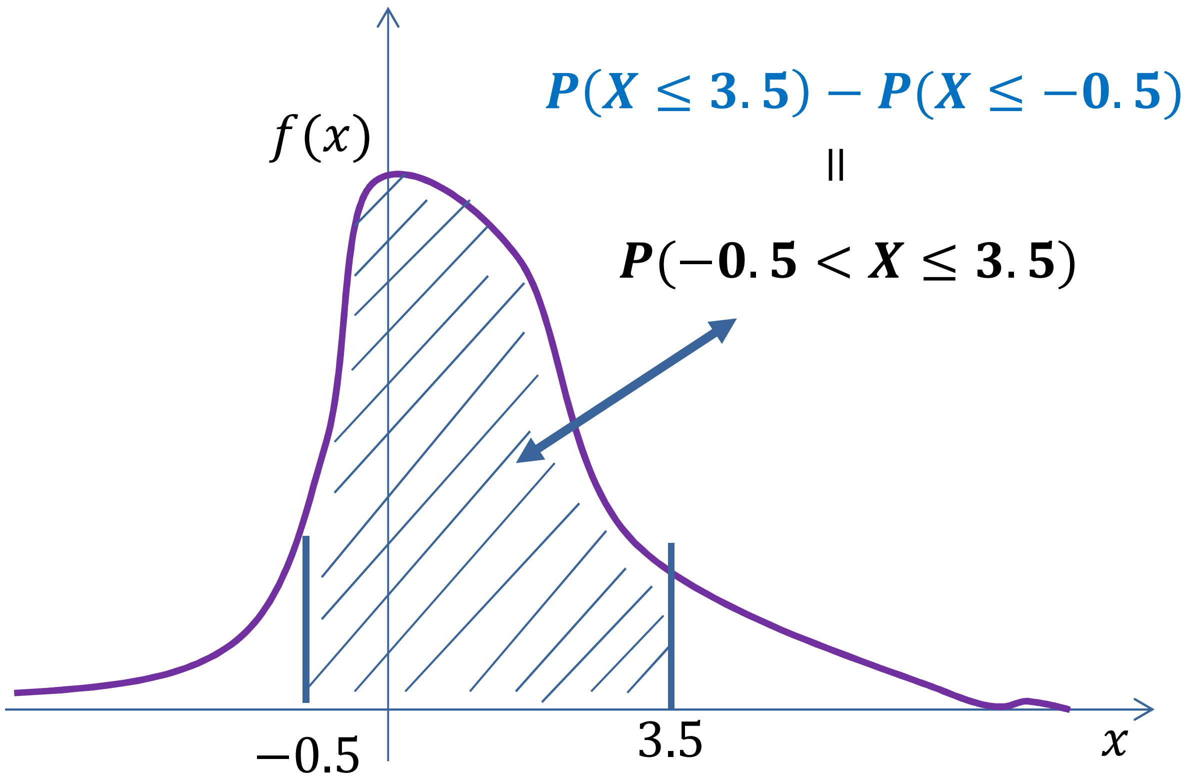
$$\boldsymbol{P(a < X \leq b) = \int_a^b f(u)du}$$

$$P(X = x) = 0$$

Variabili **continue**

$$\boldsymbol{X}: \quad (\cdots) \rightarrow \mathbb{R}$$

distribuzione: $\boldsymbol{P(X \leq x)}$ **per ogni valore di** $\boldsymbol{x}$



densità

$$f(u) > 0$$

area sotto la curva = 1

$$P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

$$\boldsymbol{P(a < X \leq b) = \int_{a}^{b} f(u)du}$$

$$P(X = x) = 0$$

Valore atteso e varianza

$$E(X) = \sum_{i=1,\dots,n} x_i P(X = x_i)$$

$$Var(X) = \sum_{i=1,\dots,n} (x_i - E(X))^2 P(X = x_i)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

tanto per saperlo...

Valore atteso e varianza

$$E(X) = \sum_{i=1,\ldots,n} x_i P(X = x_i)$$

$$Var(X) = \sum_{i=1,\ldots,n} (x_i - E(X))^2 P(X = x_i)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X) = \int_{-\infty}^{+\infty} u f(u) du$$

$$Var(X) = \int_{-\infty}^{+\infty} (u - E(X))^2 f(u) du$$

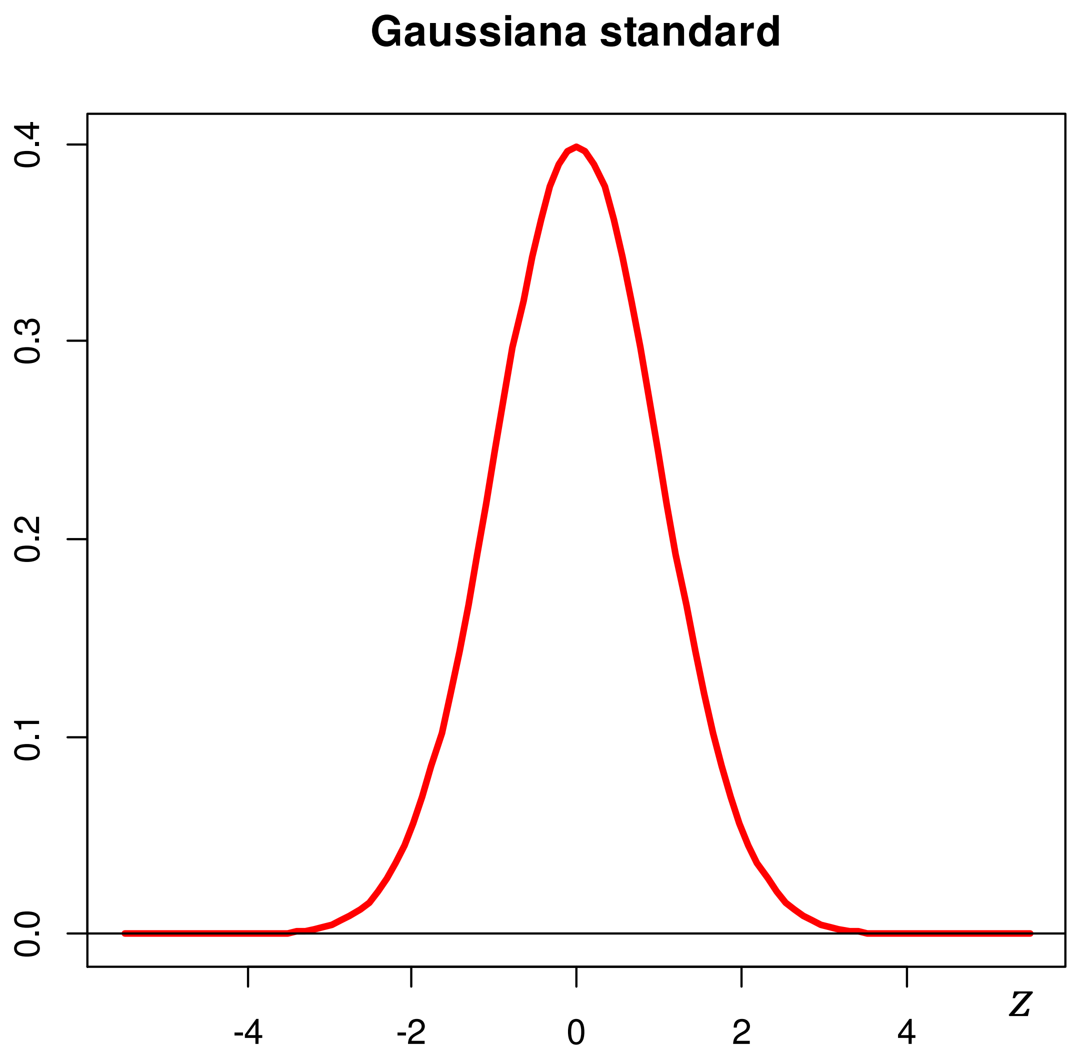
$$E(X^2) = \int_{-\infty}^{+\infty} u^2 f(u) du$$

$$\boxed{Var(X) = E(X^2) - [E(X)]^2}$$

La densità Gaussiana standard

Z ha densità Gaussiana (o Normale) standard, $N(0,1)$, se

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad , \quad z \in \mathbb{R}$$



La densità Gaussiana standard

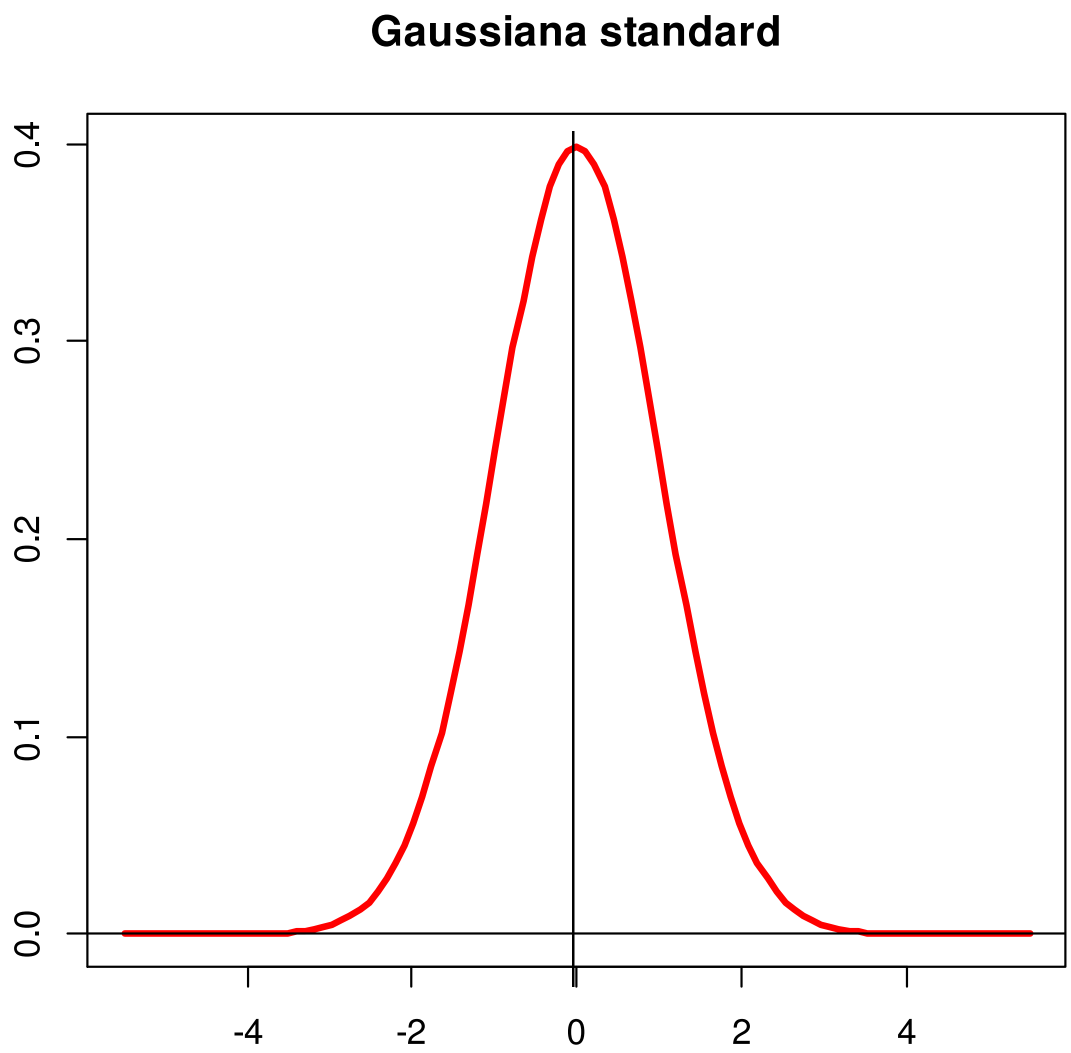
Z ha densità Gaussiana (o Normale) standard, $N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad , \quad z \in \mathbb{R}$$

Simmetrica attorno allo 0

$$E(Z) = 0$$

$$Var(Z) = 1$$



La densità Gaussiana standard

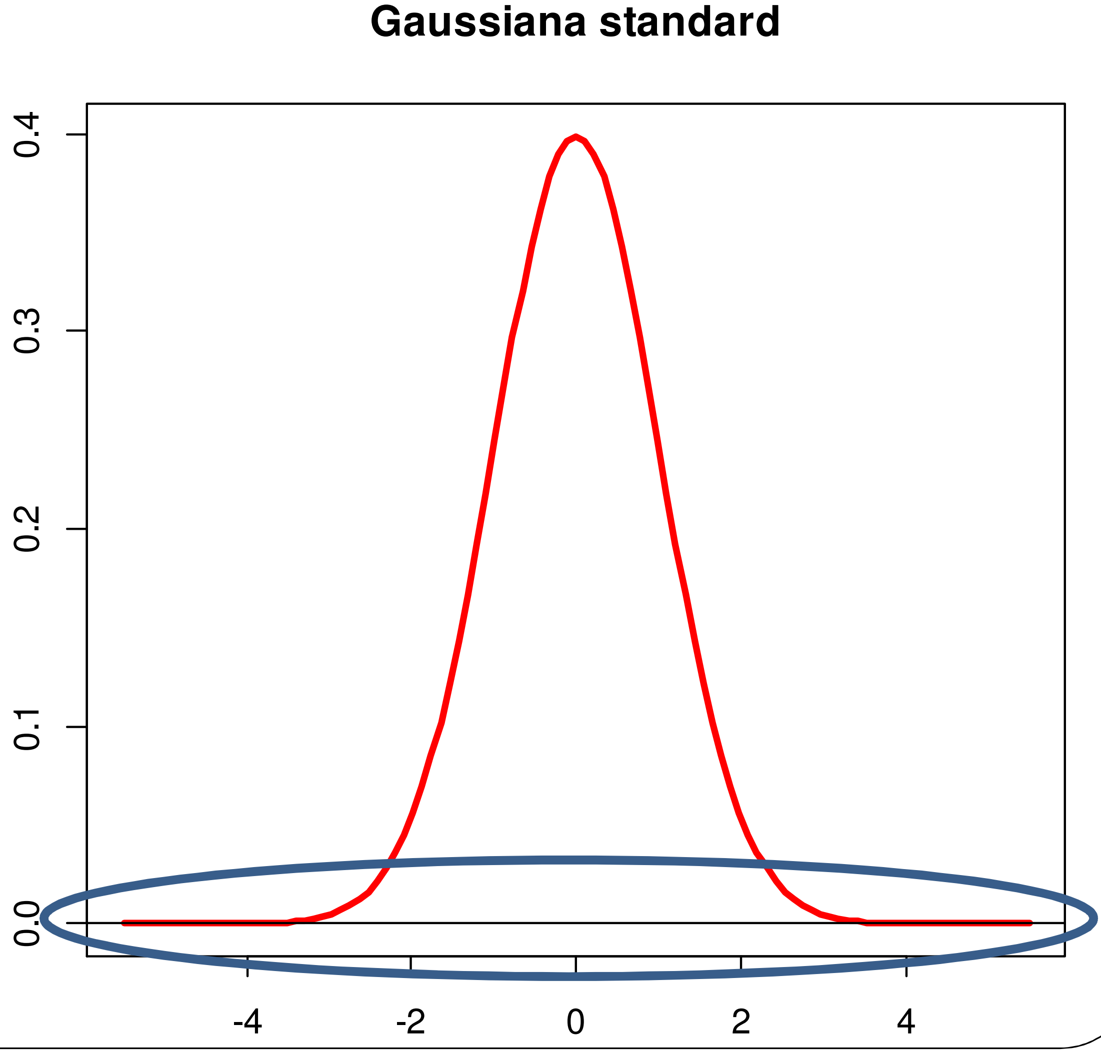
Z ha densità Gaussiana (o Normale) standard, $N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad , \quad z \in \mathbb{R}$$

Simmetrica attorno allo 0

$$E(Z) = 0$$

$$Var(Z) = 1$$



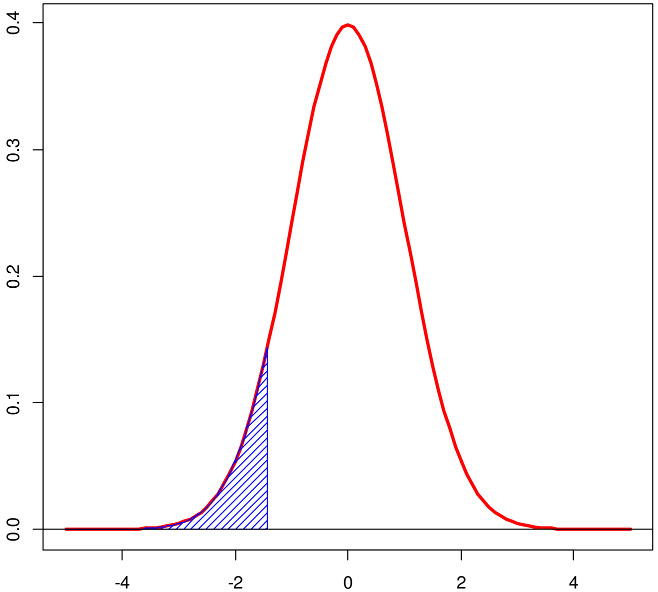
8.1 R as a set of statistical tables

One convenient use of R is to provide a comprehensive set of statistical tables. Functions are provided to evaluate the cumulative distribution function $P(X \leq x)$, the probability density function and the quantile function (given q, the smallest x such that $P(X \leq x) > q$), and to simulate from the distribution.

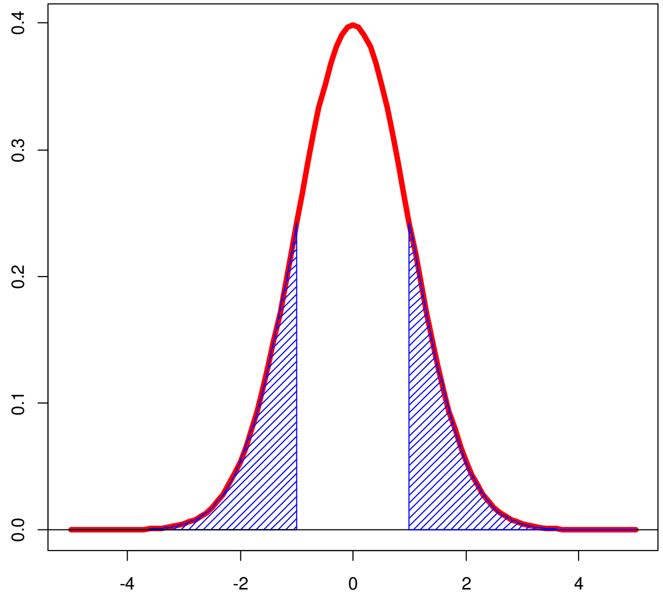
Distribution	**R name**	**additional arguments**
beta	`beta`	`shape1, shape2, ncp`
binomial	`binom`	`size, prob`
Cauchy	`cauchy`	`location, scale`
chi-squared	`chisq`	`df, ncp`
exponential	`exp`	`rate`
F	`f`	`df1, df2, ncp`
gamma	`gamma`	`shape, scale`
geometric	`geom`	`prob`
hypergeometric	`hyper`	`m, n, k`
log-normal	`lnorm`	`meanlog, sdlog`
logistic	`logis`	`location, scale`
negative binomial	`nbinom`	`size, prob`
normal	`norm`	`mean, sd`
Poisson	`pois`	`lambda`
signed rank	`signrank`	`n`
Student's t	`t`	`df, ncp`
uniform	`unif`	`min, max`
Weibull	`weibull`	`shape, scale`
Wilcoxon	`wilcox`	`m, n`

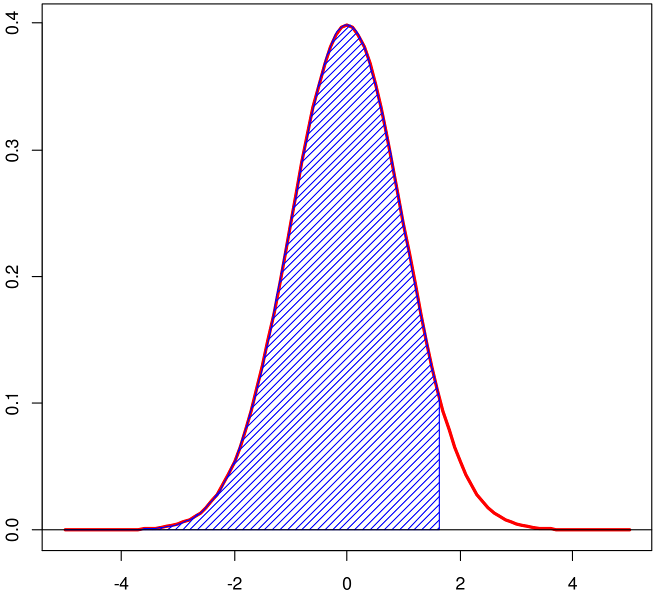
Prefix the name given here by 'd' for the density, 'p' for the CDF, 'q' for the quantile function and 'r' for simulation (*r*andom deviates). The first argument is `x` for `dxxx`, `q` for `pxxx`, `p` for `qxxx` and `n` for `rxxx` (except for `rhyper`, `rsignrank` and `rwilcox`, for which it is `nn`). In not quite all cases is the non-centrality parameter `ncp` currently available: see the on-line help for details.

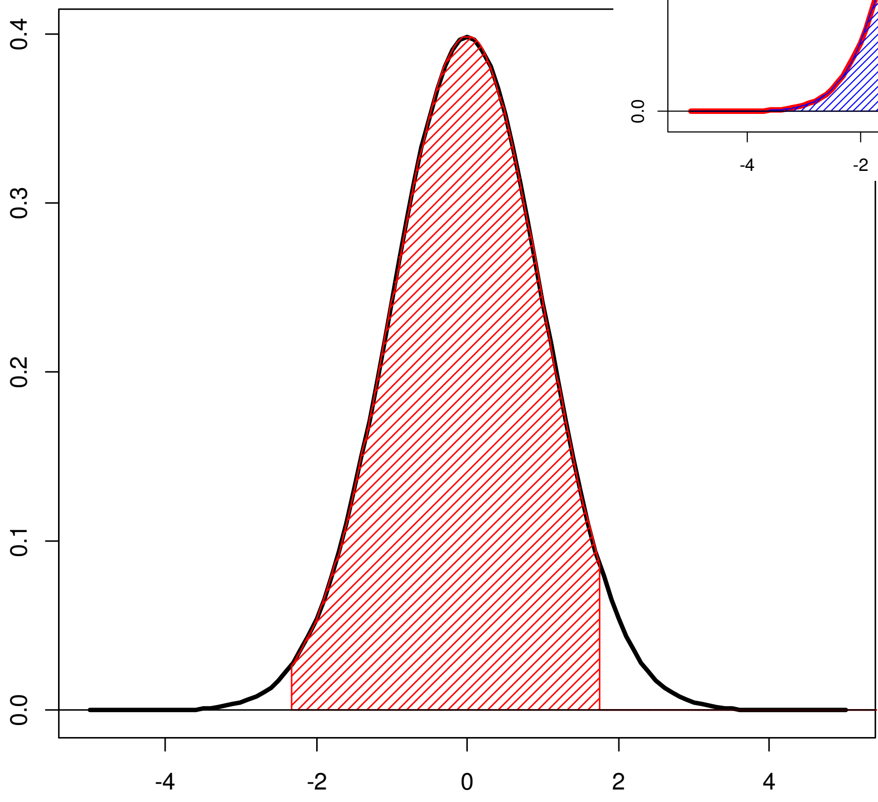
Facciamo un salto in R



Script2.R







La densità Gaussiana standard

$Z \sim N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad , \quad z \in \mathbb{R}$$

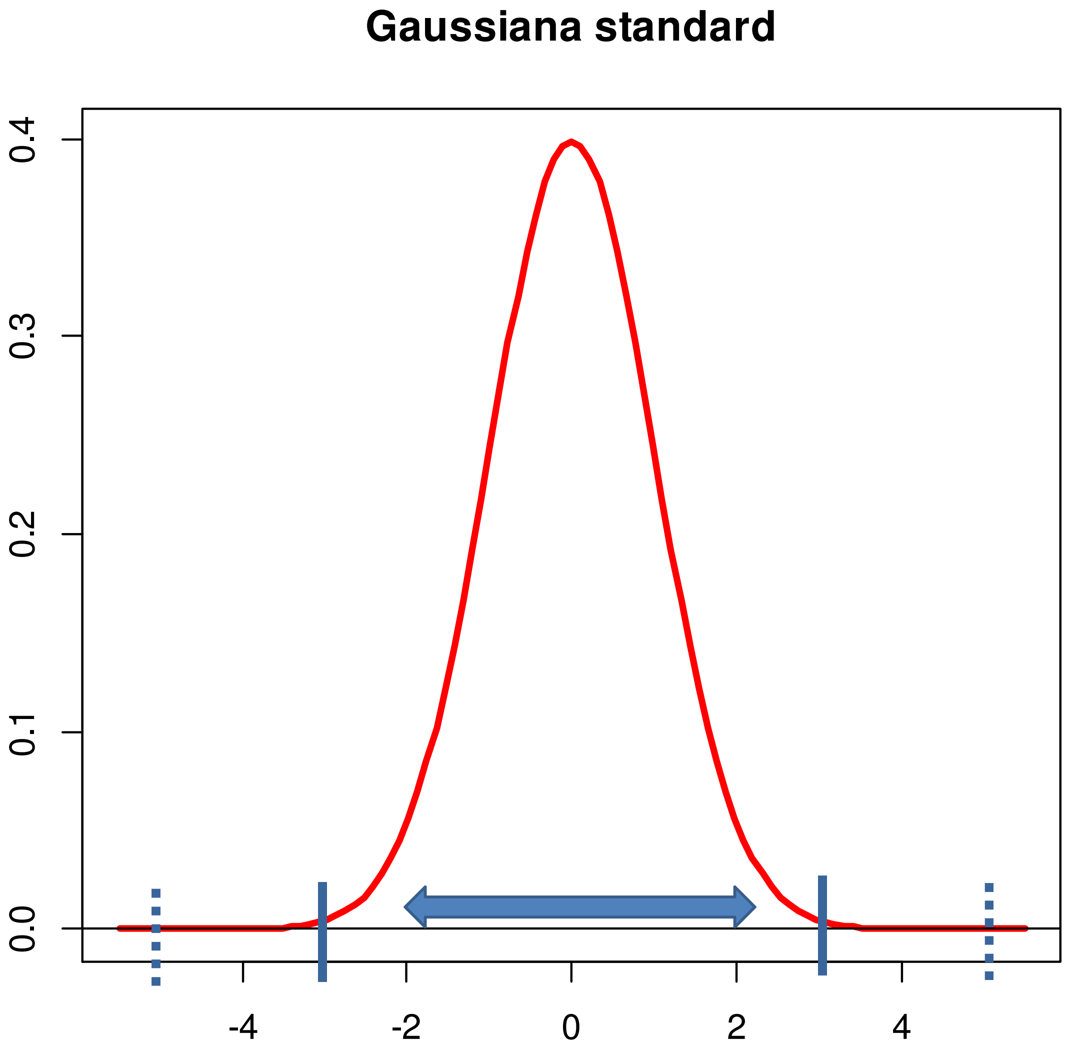
$E(Z) = \mu = 0$

$Var(Z) = \sigma^2 = 1$

$\boldsymbol{P(-2 < Z < 2) =}$

$\boldsymbol{P(-3 < Z < 3) =}$

$\boldsymbol{P(-5 < Z < 5) =}$



La densità Gaussiana

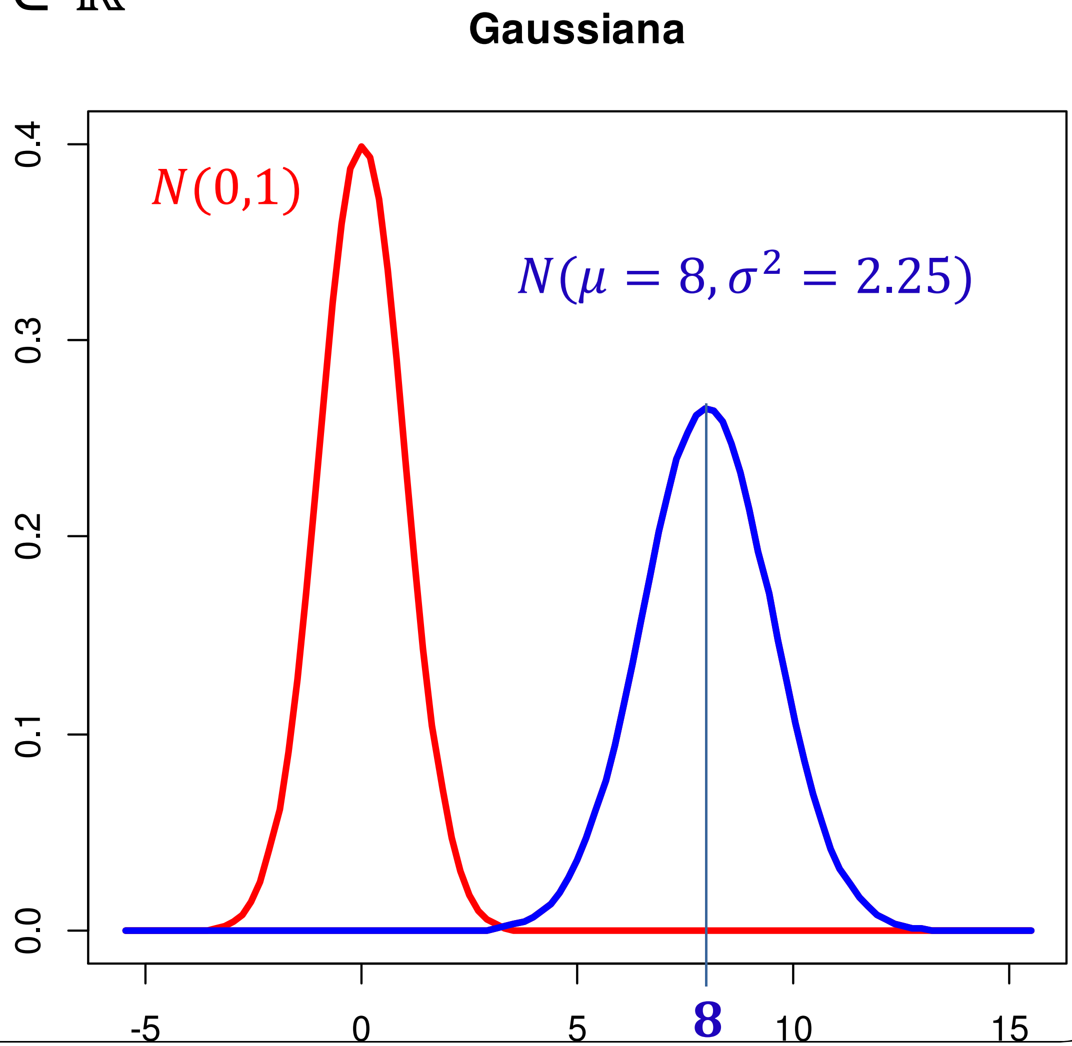
$\boxed{X \sim N(\mu, \sigma^2)}$

$$f(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} \quad , \quad u \in \mathbb{R}$$

$$E(X) = \mu$$

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

Simmetrica attorno a μ



La densità Gaussiana

$$X \sim N(\mu, \sigma^2)$$

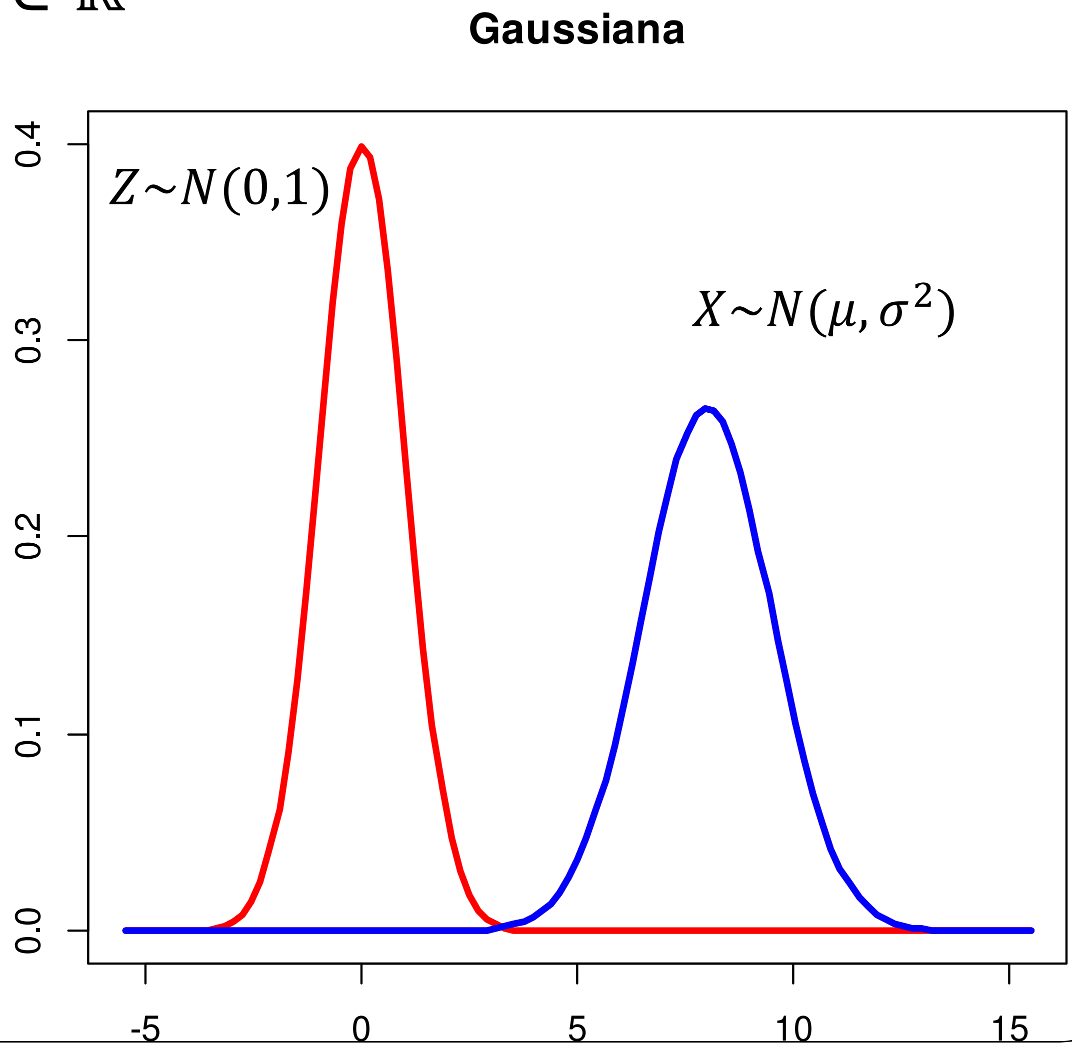
$$f(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} \quad , \quad u \in \mathbb{R}$$

$$E(X) = \mu$$

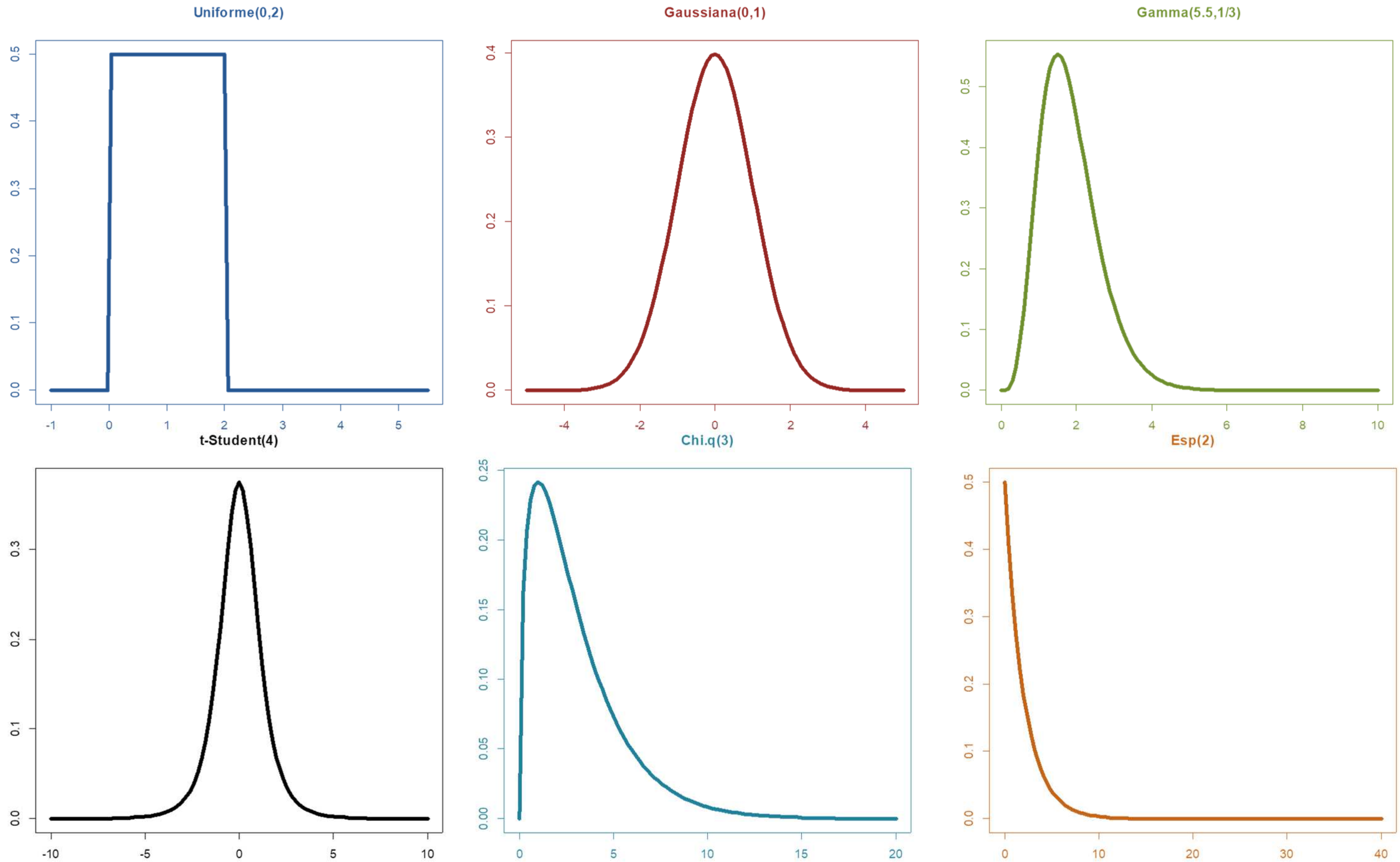
$$Var(X) = \sigma^2$$

$$Z = \frac{\boldsymbol{X - \mu}}{\boldsymbol{\sigma}} \sim N(0,1)$$

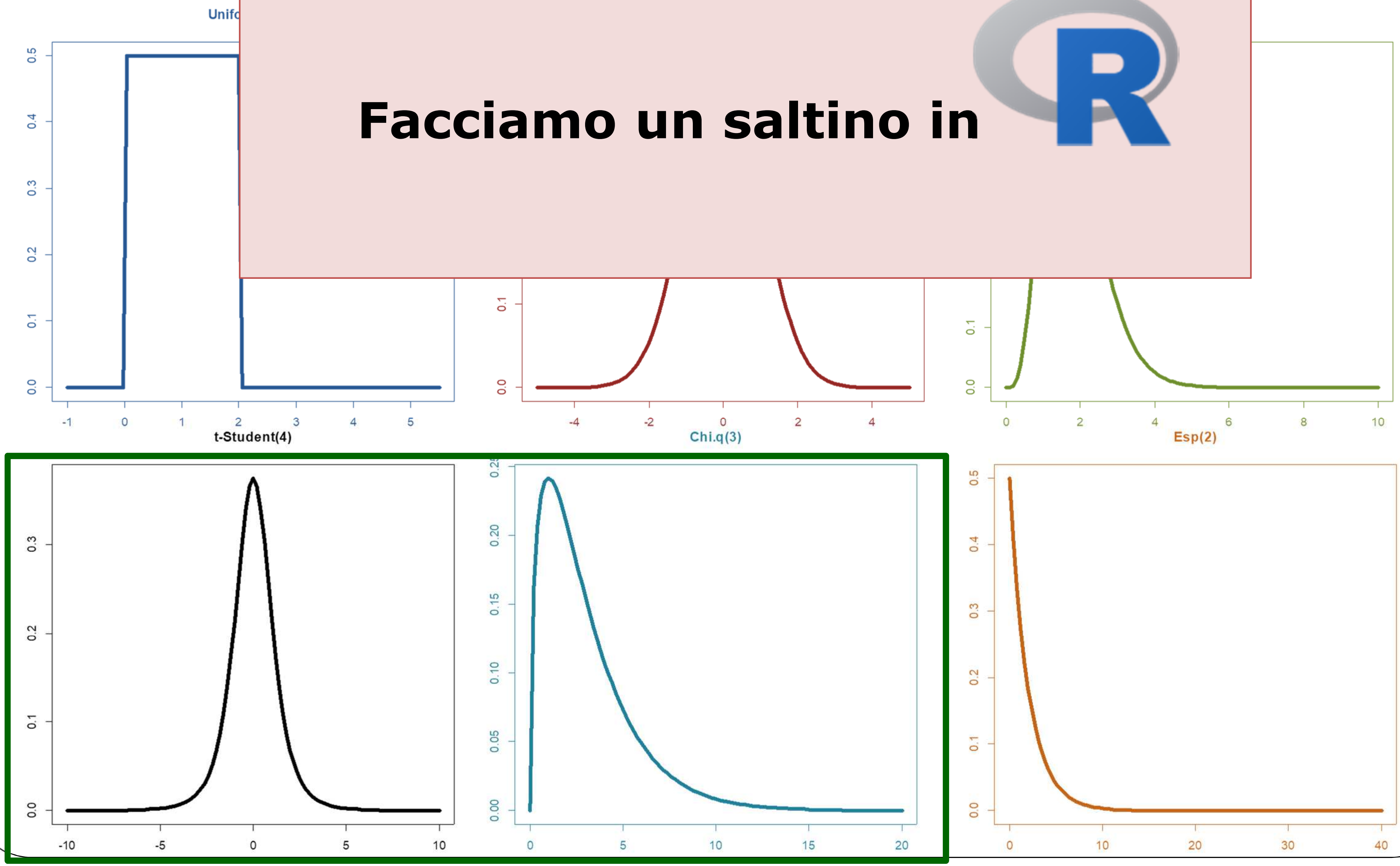
standardizzazione



NON E’ CHE ESISTA SOLO LA GAUSSIANA!!!

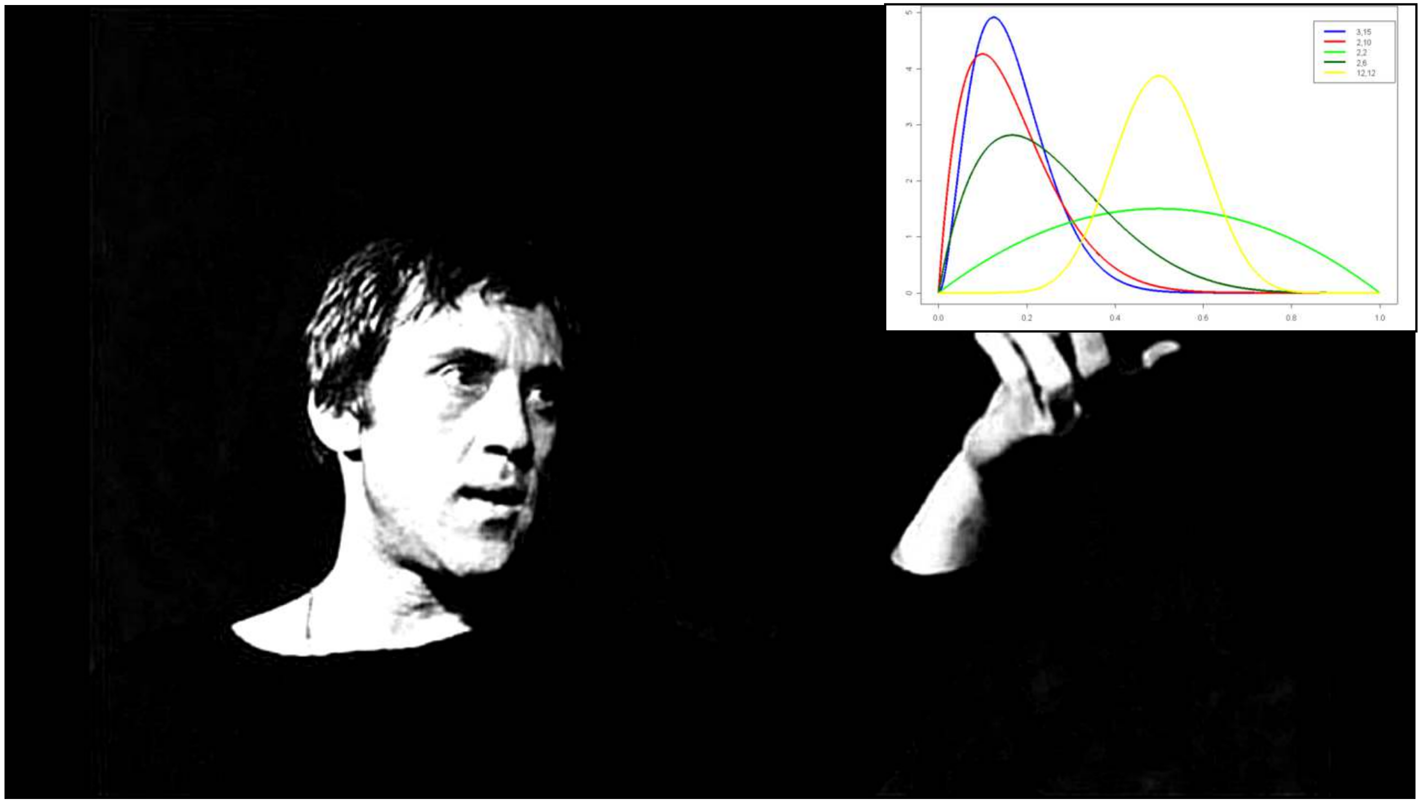


NON E' CHE ESISTA SOLO LA GAUSSIANA!!!



Facciamo un saltino in R

Gaussian or not gaussian?

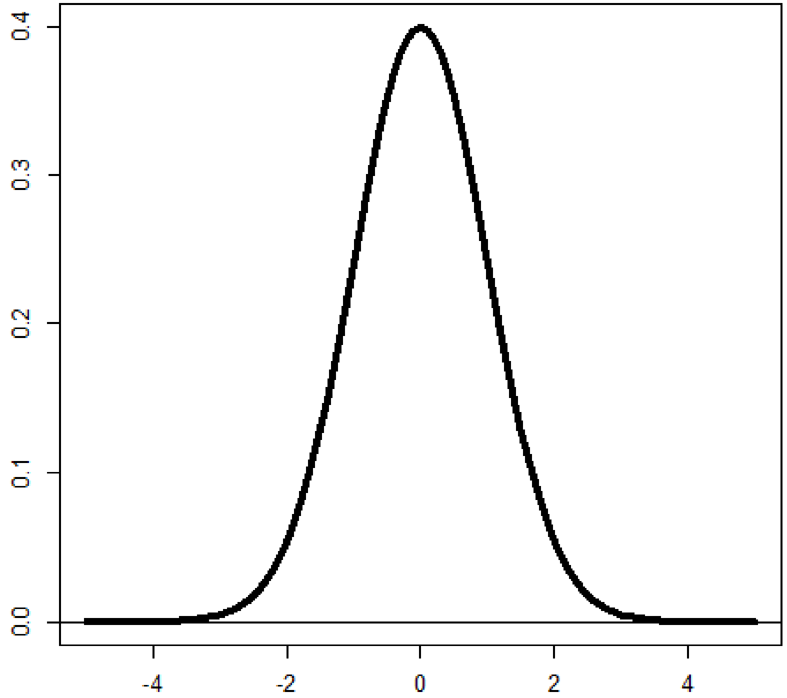


ci torneremo…

Dal nostro *test*

Sezione 2

1. La curva nella figura sottostante rappresenta la densità Gaussiana (o Normale) standard.



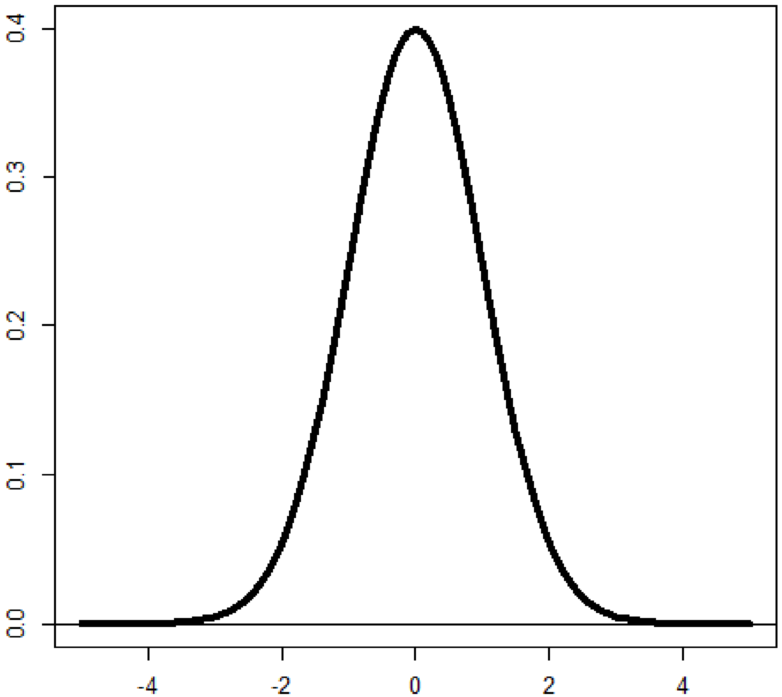
Per *densità* si intende:

Una densità di probabilità
Una frequenza di valori osservati (dati)
Non conosco la densità Gaussiana

Dal nostro *test*

Sezione 2

1. La curva nella figura sottostante rappresenta la densità Gaussiana (o Normale) standard.



Per *densità* si intende:

Una densità di probabilità
Una frequenza di valori osservati (dati)
Non conosco la densità Gaussiana

Dal nostro *test*

Le domande da 2 a 4 sono riservate a chi NON abbia dichiarato di non conoscere la densità Gaussiana.

2. Con riferimento alla figura della domanda 1, l'area sotto la curva vale:

100	1	NON SI PUO' SAPERE SENZA FARE CALCOLI

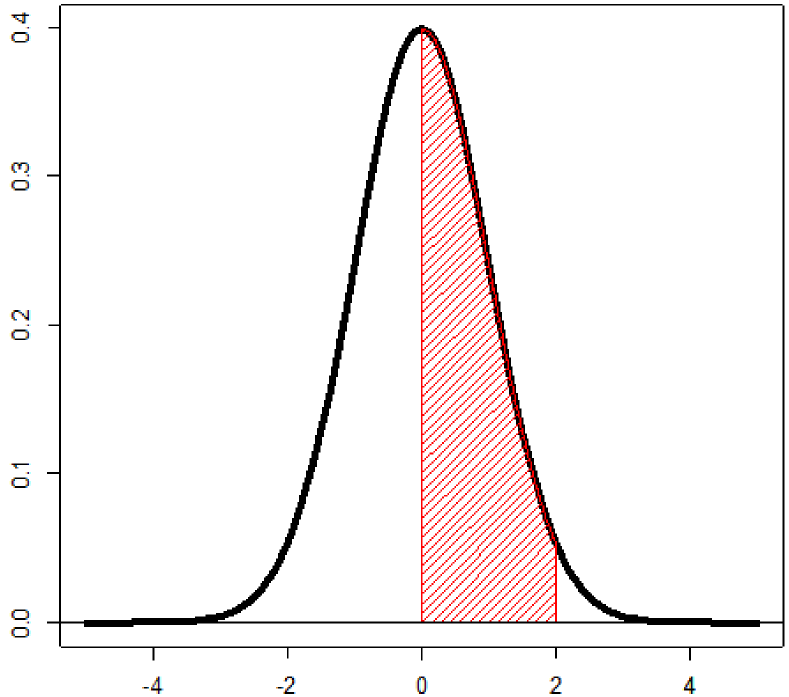
Dal nostro *test*

Le domande da 2 a 4 sono riservate a chi NON abbia dichiarato di non conoscere la densità Gaussiana.

2. Con riferimento alla figura della domanda 1, l'area sotto la curva vale:

100	1	NON SI PUO' SAPERE SENZA FARE CALCOLI

3. Con riferimento alla figura sottostante, la zona rossa rappresenta:



Un'area
Una probabilità
Una regione critica

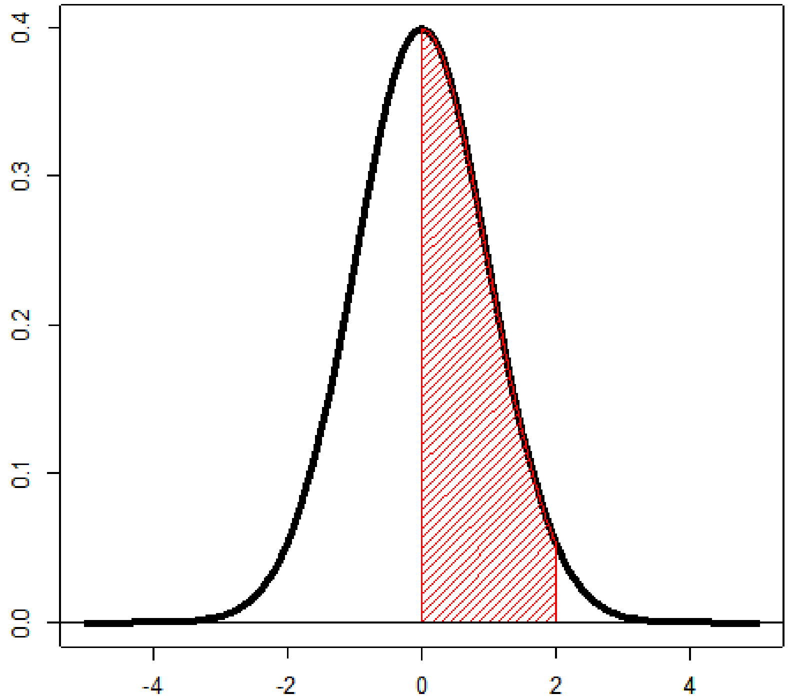
Dal nostro *test*

Le domande da 2 a 4 sono riservate a chi NON abbia dichiarato di non conoscere la densità Gaussiana.

2. Con riferimento alla figura della domanda 1, l'area sotto la curva vale:

100	1	NON SI PUO' SAPERE SENZA FARE CALCOLI

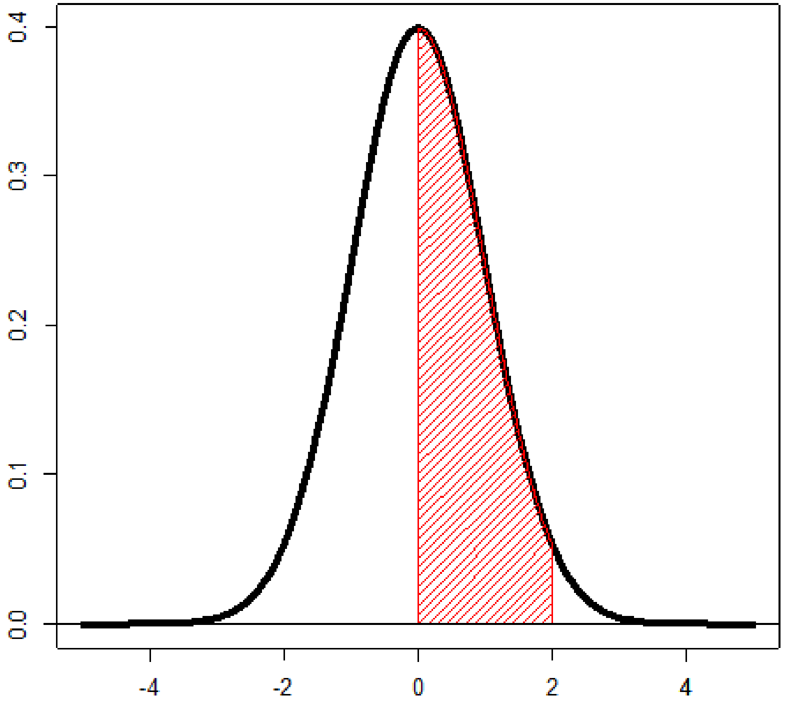
3. Con riferimento alla figura sottostante, la zona rossa rappresenta:



Un'area
Una probabilità
Una regione critica

Dal nostro *test*

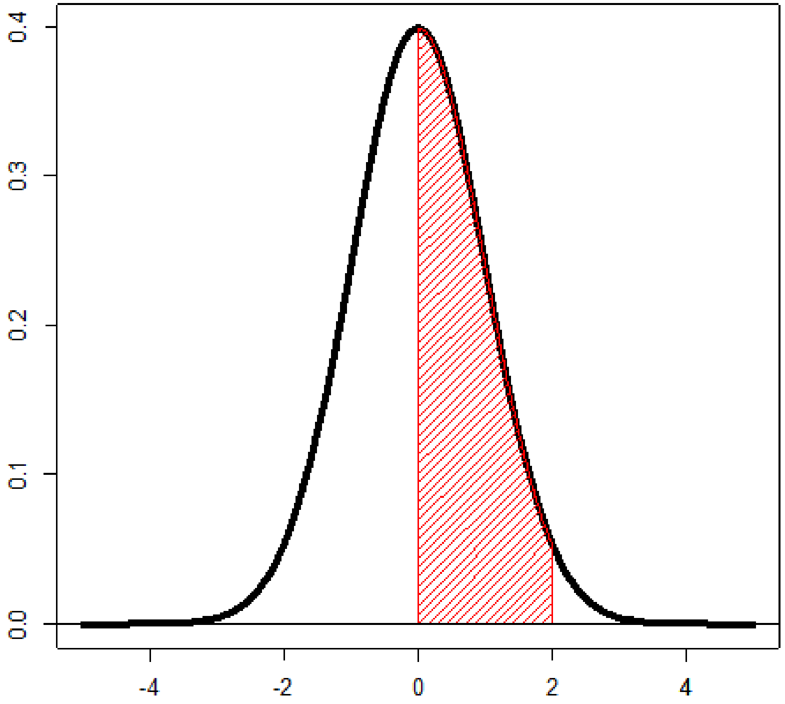
4. Con riferimento alla figura sottostante, l'area rossa vale:



Meno di 0.5
Più di 0.5
Impossibile dirlo senza fare calcoli

Dal nostro *test*

4. Con riferimento alla figura sottostante, l'area rossa vale:



Meno di 0.5
Più di 0.5
Impossibile dirlo senza fare calcoli