

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

Una domanda che viene posta di frequente dagli studenti che affrontano il corso di statistica usando R e' la seguente: "qual'e' l'equivalente in R delle tabelle di probabilita' e dei corrispondenti z-score che si trovano nell'appendice dei libri di statistica?"

Un esempio di tipico di tali tabelle e' disponibile a questo indirizzo:

<http://home.ubalt.edu/ntsbarsh/Business-stat/StatisticalTables.pdf>

Introduzione alle funzioni $dnorm$, $pnorm$, $qnorm$ e $rnorm$ in R

In R esistono funzioni specifiche per molte distribuzioni (normale, t, Fisher, ...). In questi esempi ci concentreremo sulla distribuzione normale e cercheremo di descrivere il set di funzioni che R mette a disposizione per ogni distribuzione ed il loro utilizzo.

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

dnorm()

La funzione di densita' di probabilita' per la distribuzione normale e' la seguente:

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

La funzione `dnorm(x, mean, sd)` restituisce il valore di questa funzione dati in input i valori di x , μ e σ .

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`dnorm()`

```
> dnorm(0, mean = 0, sd = 1)  
[1] 0.3989423
```

Lo stesso risultato si ottiene mediante:

```
> dnorm(0)  
[1] 0.3989423
```

Dato che 0 e 1 sono valori di default per gli argomenti `mean` e `sd`

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

dnorm()

Nonostante il fatto che x rappresenta la variabile indipendente della funzione di densità di probabilità della normale e' utile pensare ad essa come ad uno Z-score . Vediamo meglio questo concetto plottando la funzione di densità di probabilità (pdf) della normale ...

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

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

dnorm()

```
# Costruisco un vettore di Z-score
```

```
> z_scores <- seq(-3, 3, by = .1)
```

```
# Stampo il vettore
```

```
> z_scores
```

```
[1] -3.0 -2.9 -2.8 -2.7 -2.6 -2.5 -2.4 -2.3 -2.2 -2.1 -2.0 -1.9 -1.8 -1.7  
[15] -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3  
[29] -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1.0  1.1  
[43]  1.2  1.3  1.4  1.5  1.6  1.7  1.8  1.9  2.0  2.1  2.2  2.3  2.4  2.5  
[57]  2.6  2.7  2.8  2.9  3.0
```

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R dnorm()

```
# Costruisco un vettore dei valori che pdf assume in  
# corrispondenza degli Z-score  
> dvalues <- dnorm(z_scores)  
# Stampo il vettore  
> dvalues
```

```
[1] 0.004431848 0.005952532 0.007915452 0.010420935 0.013582969  
[6] 0.017528300 0.022394530 0.028327038 0.035474593 0.043983596  
[11] 0.053990967 0.065615815 0.078950158 0.094049077 0.110920835  
[16] 0.129517596 0.149727466 0.171368592 0.194186055 0.217852177  
[21] 0.241970725 0.266085250 0.289691553 0.312253933 0.333224603  
[26] 0.352065327 0.368270140 0.381387815 0.391042694 0.396952547  
[31] 0.398942280 0.396952547 0.391042694 0.381387815 0.368270140  
[36] 0.352065327 0.333224603 0.312253933 0.289691553 0.266085250  
[41] 0.241970725 0.217852177 0.194186055 0.171368592 0.149727466  
[46] 0.129517596 0.110920835 0.094049077 0.078950158 0.065615815  
[51] 0.053990967 0.043983596 0.035474593 0.028327038 0.022394530  
[56] 0.017528300 0.013582969 0.010420935 0.007915452 0.005952532  
[61] 0.004431848
```

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`dnorm()`

Ora stampiamo i valori

```
> plot(dvalues, # plotta (valore, indice di valore nel vettore)
      xaxt = "n", # Non etichettare asse x
      type = "l", # Line plot
      main = "pdf of the Standard Normal",
      xlab = "Z-score")
```

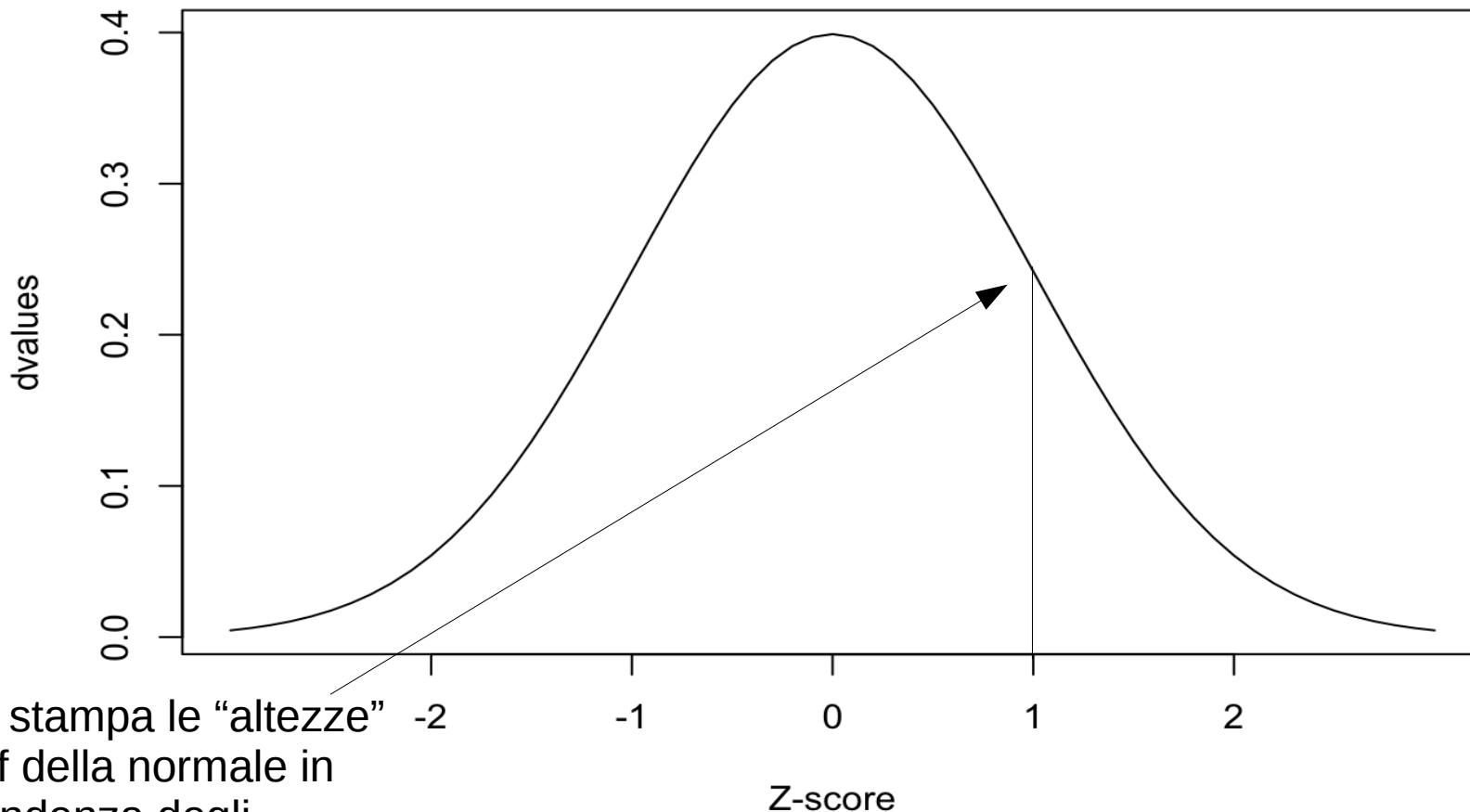
Questi comandi etichettano l'asse x

```
> axis(1, at=which(dvalues == dnorm(0)), labels=c(0))
> axis(1, at=which(dvalues == dnorm(1)), labels=c(-1, 1))
> axis(1, at=which(dvalues == dnorm(2)), labels=c(-2, 2))
```


Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`dnorm()`

pdf of the Standard Normal



`dnorm()` stampa le "altezze" della pdf della normale in corrispondenza degli Z-score forniti in input

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

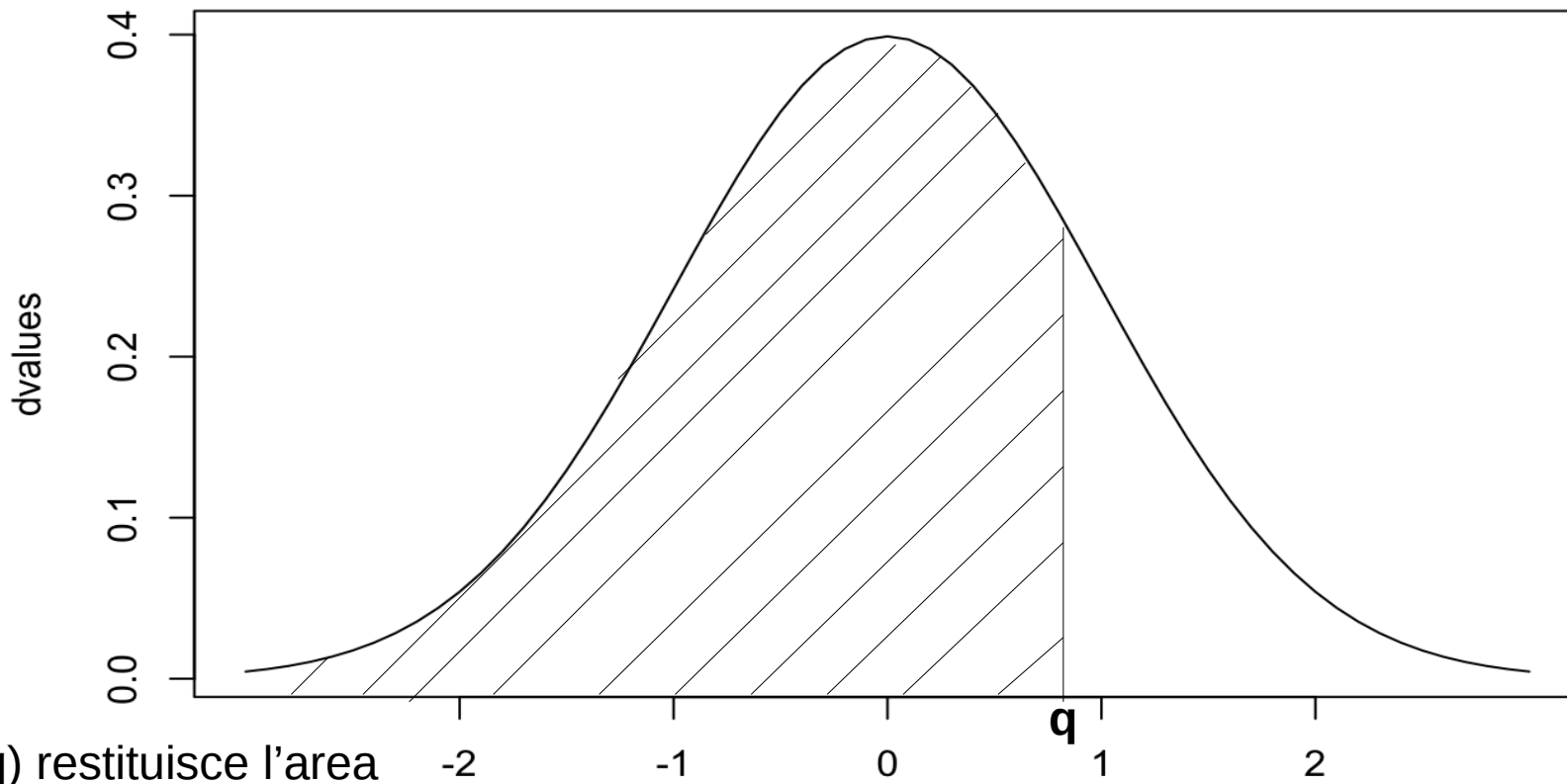
`pnorm()`

La funzione `pnorm(q)` restituisce l'integrale da $-\infty$ a q della pdf della normale, quando q è uno Z-score. Provate ad indovinare il valore di `pnorm(0)`. `Pnorm()` ha gli stessi valori di default di `mu` e `sigma` (mean e sd) della funzione `dnorm()`.

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`pnorm()`

pdf of the Standard Normal



`pnorm(q)` restituisce l'area sotto la curva della pdf della normale che va da $-\infty$ a q

Z-score

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

```
# Valore integrale sotto pdf della normale da -Inf a 0 (con mean=0  
# e sd=1, valori di default)  
> pnorm(0)  
[1] 0.5
```

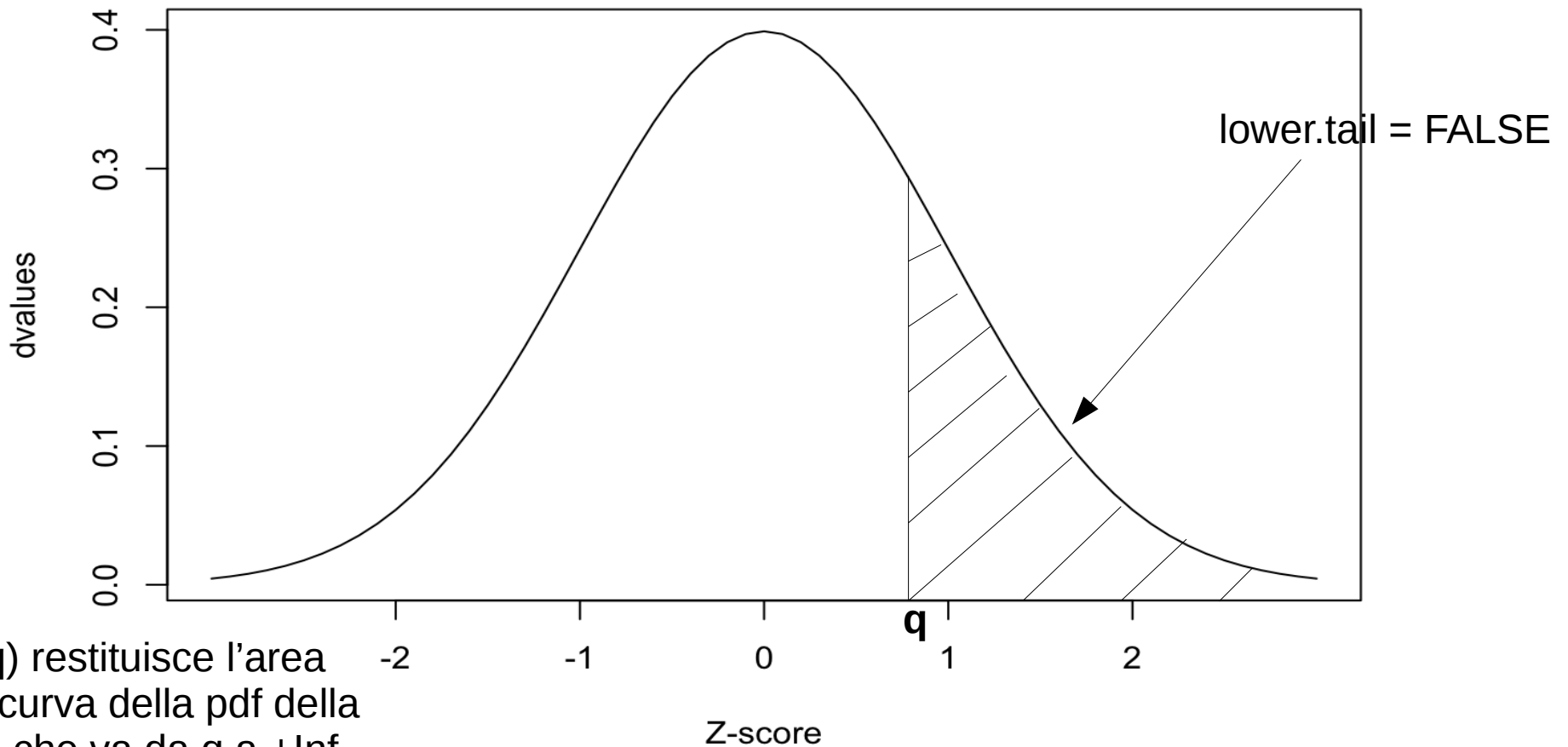
Pnorm() ha anche un argomento **lower.tail**. Se lower.tail e' importato a FALSE pnorm() restituisce l'integrale della pdf della normale da q a +Inf

```
# Questi comandi etichettano l'asse x  
# notare che pnorm(q) = 1 - pnorm(q, lower.tail=FALSE)  
> pnorm(2, mean = 5, sd = 3)  
[1] 0.1586553  
> pnorm(2, mean = 5, sd = 3, lower.tail = FALSE)  
[1] 0.8413447  
1 - pnorm(2, mean = 5, sd = 3, lower.tail = FALSE)  
[1] 0.1586553
```

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`pnorm()`

pdf of the Standard Normal



`pnorm(q)` restituisce l'area
sotto la curva della pdf della
normale che va da q a $+\text{Inf}$

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

`pnorm()`

`Pnorm()` e' la funzione che rimpiazza le tavole degli Z-score della distribuzione normale che si trovano nelle appendici dei libri di statistica. Riprendiamo il vettore degli z-score utilizzato prima e calcoliamo un nuovo vettore con le masse di probabilita' basate sulla funzione `pnorm()`

```
> pvalues <- pnorm(z_scores)
```

Poi plottiamo il vettore `pvalues`. Riuscite ad indovinare la forma della curva? (motivate la risposta)

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

```
# plottiamo il vettore pvalues
```

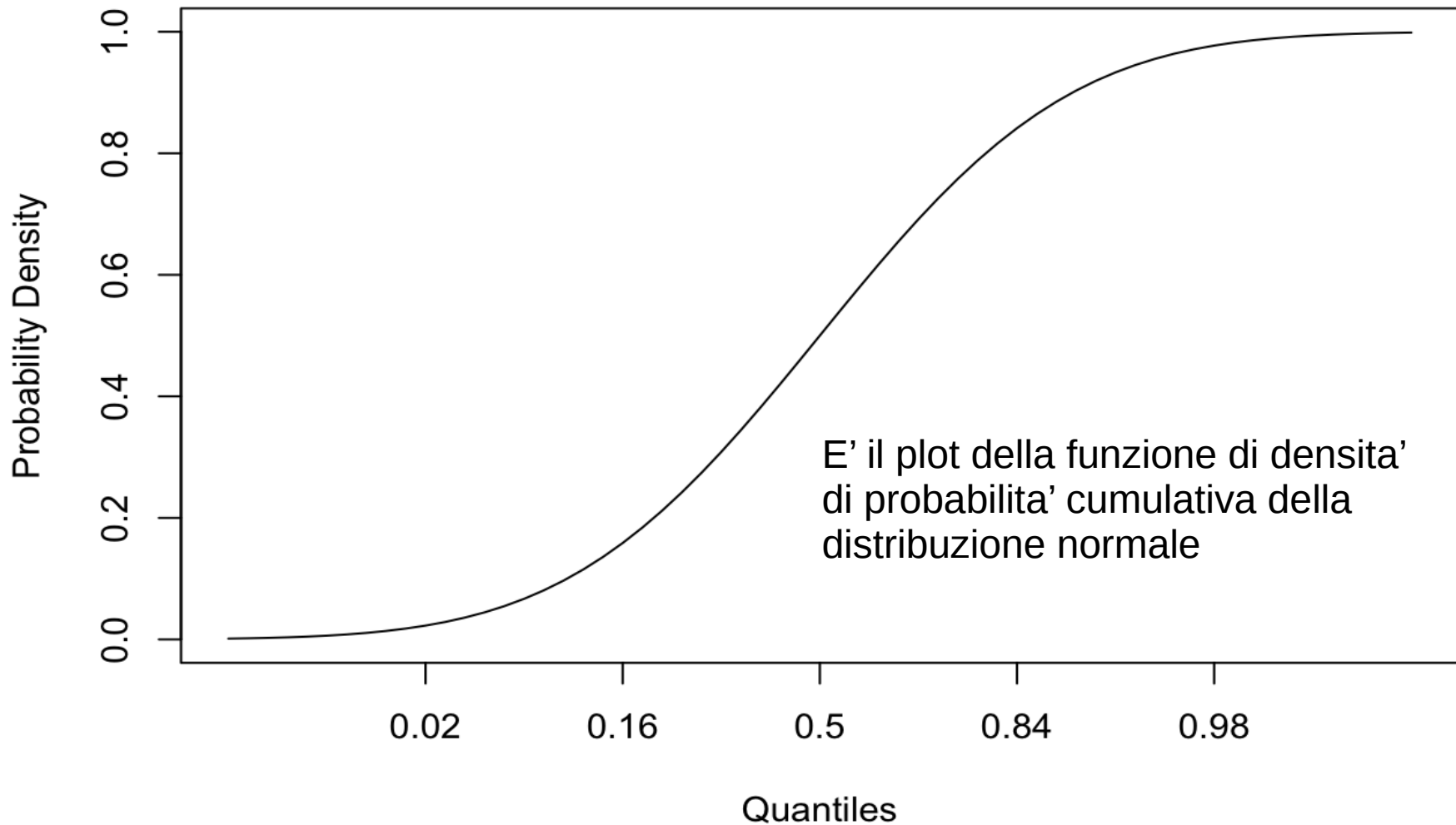
```
plot(pvalues, # plot (valore, indice di valore in pvalues)  
     xaxt = "n", # Non etichettare asse x  
     type = "l", # line plot  
     main = "cdf of the Standard Normal",  
     xlab = "Quantiles",  
     ylab = "Probability Density")
```

```
# Comandi per etichettare asse x
```

```
axis(1, at=which(pvalues == pnorm(-2)), labels=round(pnorm(-2), 2))  
axis(1, at=which(pvalues == pnorm(-1)), labels=round(pnorm(-1), 2))  
axis(1, at=which(pvalues == pnorm(0)), labels=c(.5))  
axis(1, at=which(pvalues == pnorm(1)), labels=round(pnorm(1), 2))  
axis(1, at=which(pvalues == pnorm(2)), labels=round(pnorm(2), 2))
```

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

cdf of the Standard Normal



Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

Esempio di calcolo di p-value in R : calcolo di un singolo p-value dalla normale

Per rendere le cose piu' semplici consideriamo un test a "due code"

$H_0 : \mu = a$ (ipotesi nulla)

$H_a : \mu \neq a$

Il p-value e' calcolato per un particolare valore della media campionaria. Qui assumiamo di aver ottenuto un particolare valore di media campionaria (\bar{x}) e di voler trovare il suo p-value.

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

Esempio di calcolo di p-value in R : calcolo di un singolo p-value dalla normale

Per rendere le cose piu' semplici consideriamo un test a "due code"

$H_0 : \mu = a$ (ipotesi nulla)

$H_a : \mu \neq a$

Nel caso particolare della distribuzione normale abbiamo bisogno anche della deviazione standard. Assumiamo che essa sia data (nell'esercizio).

Introduzione alle funzioni `dnorm`, `pnorm`, `qnorm` e `rnorm` in R

Z-score e' calcolato assumendo H_0 vera, sottraendo la media assunta e dividendo per la deviazione standard teorica (fornita). Una volta che abbiamo lo Z-score la probabilita' che il valore possa essere minore dello Z-score viene ottenuto utilizzando la funzione `pnorm()`.

Questo non e' sufficiente per calcolare sempre lo Z-score. Se lo Z-score calcolato e' POSITIVO allora dobbiamo calcolare $1 -$ la probabilita' associata. Inoltre, per un test a due code, dobbiamo moltiplicare il risultato per 2.

Per evitare questi problemi imponiamo uno Z-score negativo utilizzando il negativo del valore assoluto dello Z-score.

Introduzione alle funzioni dnorm, pnorm, qnorm e rnorm in R

$H_0 : \mu = a$ (ipotesi nulla)

$H_a : \mu \neq a$

Ora vediamo un esempio specifico in R . Nell'esempio che segue useremo un valore di a pari a 5, una deviazione standard pari a 2 ed una numerosita' del campione pari a 20. Troveremo il p-value per una media campionaria pari a 7.

```
> a <- 5
> s <- 2
> n <- 20
> xbar <- 7
> z <- (xbar-a)/(s/sqrt(n))      # calcolo Z-score
> z
[1] 4.472136
> 2*pnorm(-abs(z))              # calcolo p-value
[1] 7.744216e-06
```