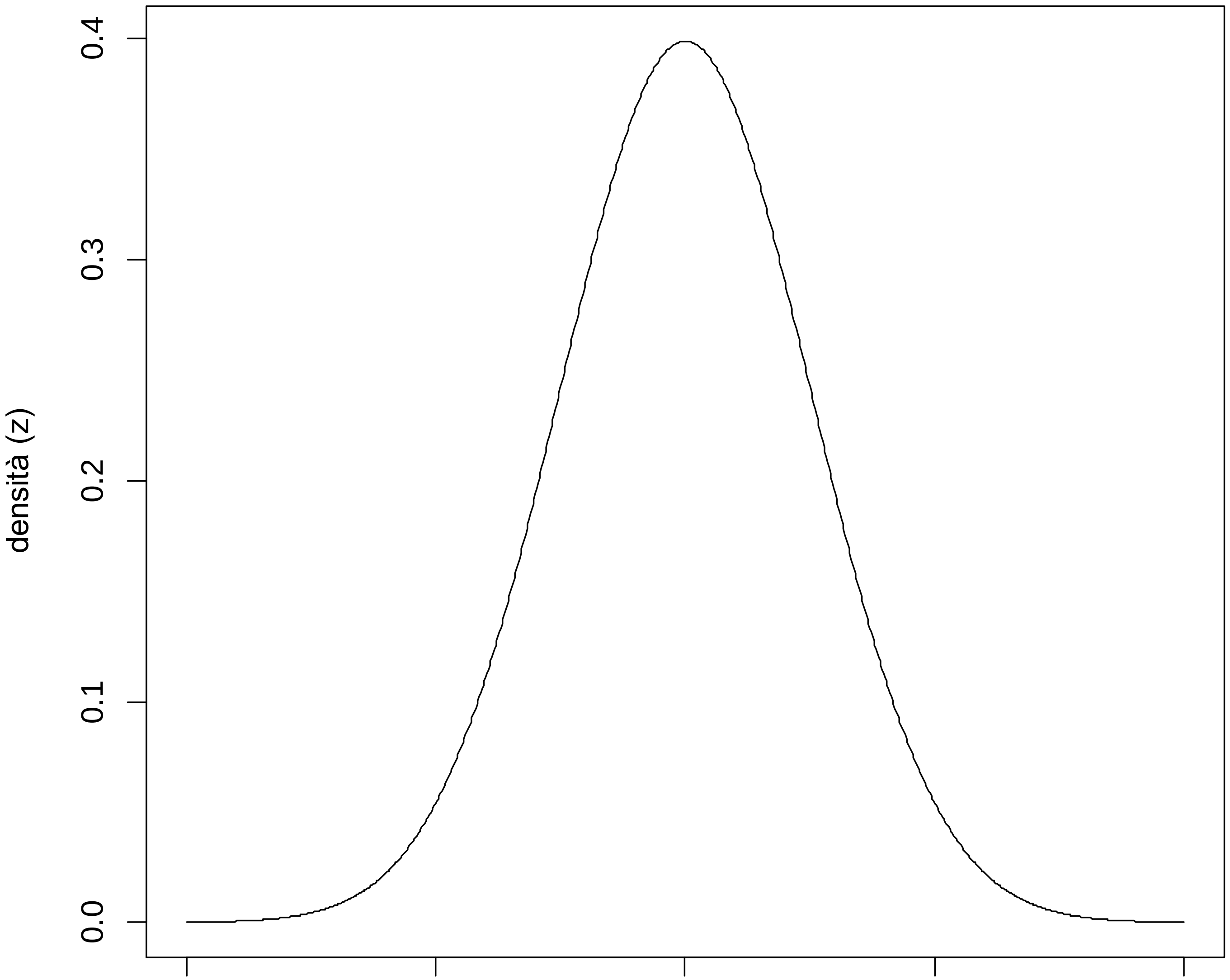
UNIVERSITA' DEGLI STUDI DI VERONA

LABORATORIO DI PROBABILITA' E STATISTICA

Docente: Bruno Gobbi

8 – VARIABILI CASUALI CONTINUE

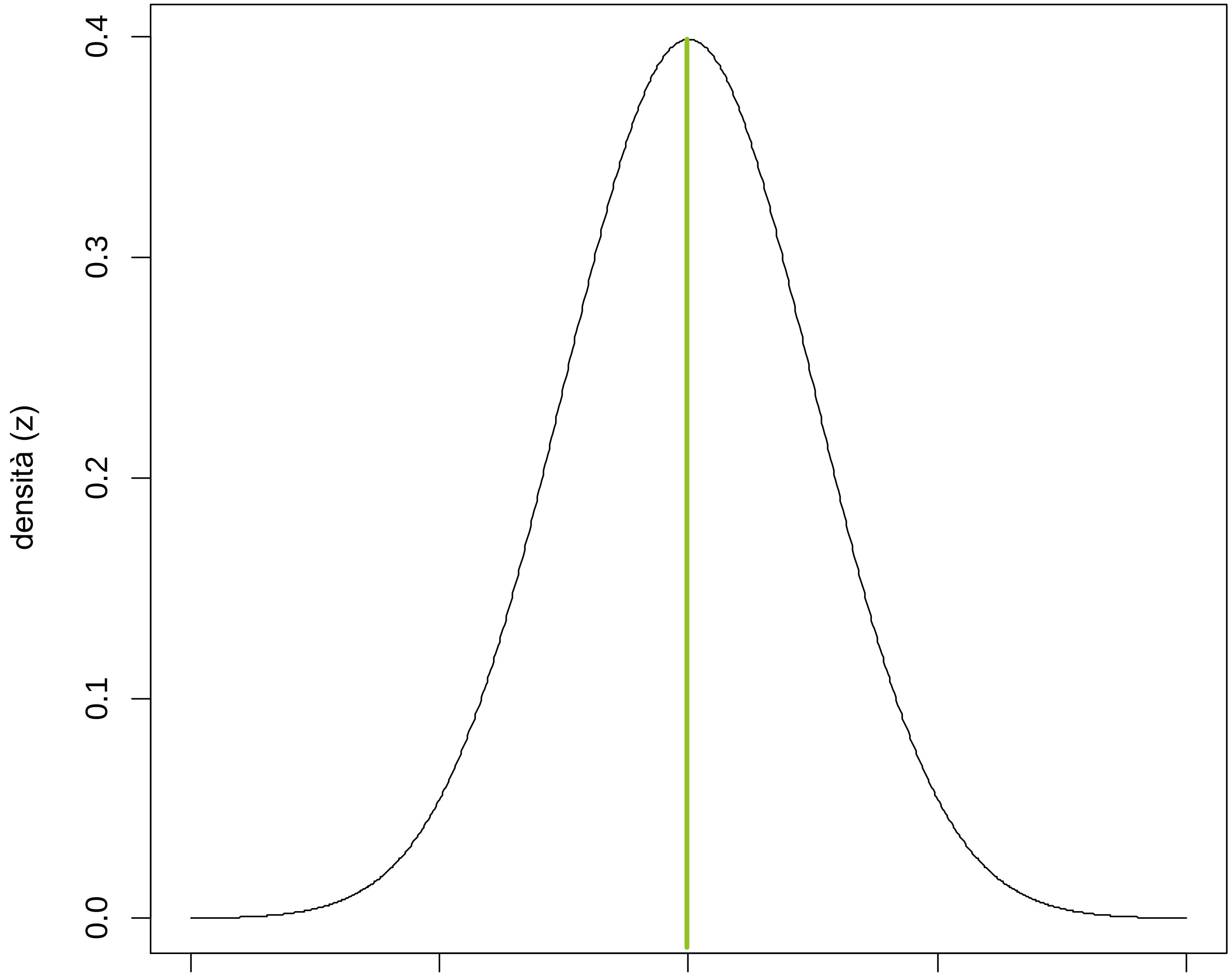
LA VARIABILE NORMALE



LA VARIABILE NORMALE

La variabile aleatoria **Normale** rappresenta la distribuzione di probabilità più usata in ambito statistico, perché molti fenomeni nella realtà si distribuiscono secondo una tipica forma a campana, con la maggioranza della popolazione concentrata nel mezzo con dei valori che scendono gradualmente man mano che andiamo verso gli estremi a sinistra e a destra.

LA VARIABILE NORMALE



Media = Mediana = Moda → E' simmetrica

LA VARIABILE NORMALE

Funzione di densità della v.a. Normale:

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

LA VARIABILE NORMALE

I momenti della variabile normale sono:

- Media: $\mu = M(x)$
- Varianza: $\sigma^2 = V(x)$
- Scarto quadratico medio: $\sigma = E(x)$

LA VARIABILE NORMALE

In R si definiscono quattro funzioni per la variabile normale:

- **dnorm()** calcola la densità di probabilità
- **pnorm()** è la funzione di probabilità cumulata
- **qnorm()** è l'inversa della probabilità cumulata
- **rnorm()** per creare dei valori random generati da una variabile aleatoria normale

ESEMPIO DI VARIABILE NORMALE

Proviamo a disegnare la distribuzione di probabilità dell’altezza media delle donne italiane. Sappiamo che l’altezza media delle donne intorno ai 20 anni del bel Paese è di 168 cm e che la variabilità ha uno scarto quadratico medio di 12 cm.

ESEMPIO DI VARIABILE NORMALE

- Media: $\mu = 168$
- Scarto quadratico medio: $\sigma = 12$

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

LA FUNZIONE dnorm(x, μ, σ)

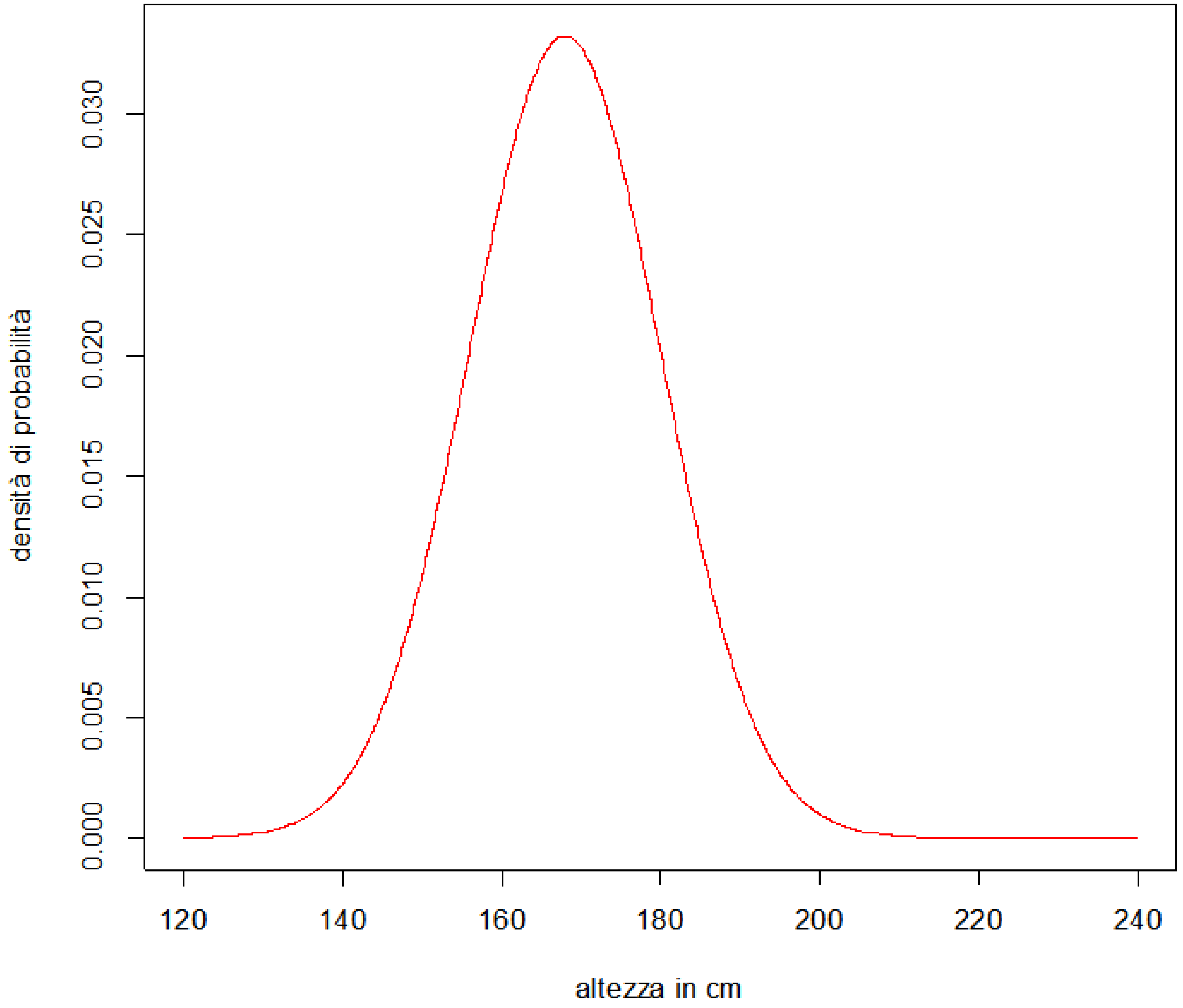
 # CREO INNANZITUTTO UN ASSE DELLE X SUFFICIENTEMENTE GRANDE PER CONTENERE TUTTE LE ALTEZZE, QUINDI DA 120 A 240 CM

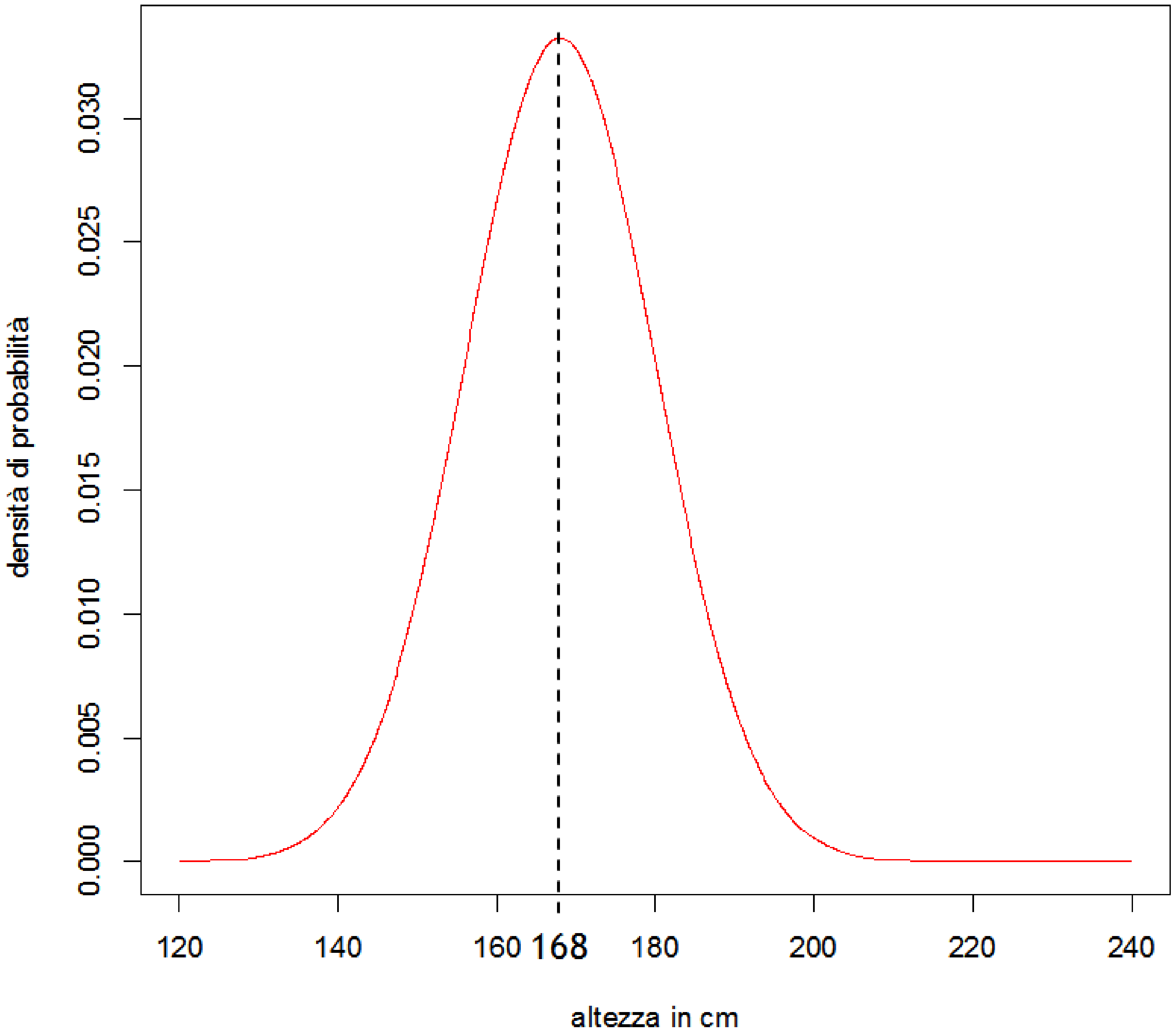
```
> x=seq(120, 240, by = 0.01)
```

 # CREO LA DISTRIBUZIONE NORMALE DELL'ALTEZZA DELLE DONNE ITALIANE CON LA FUNZIONE dnorm

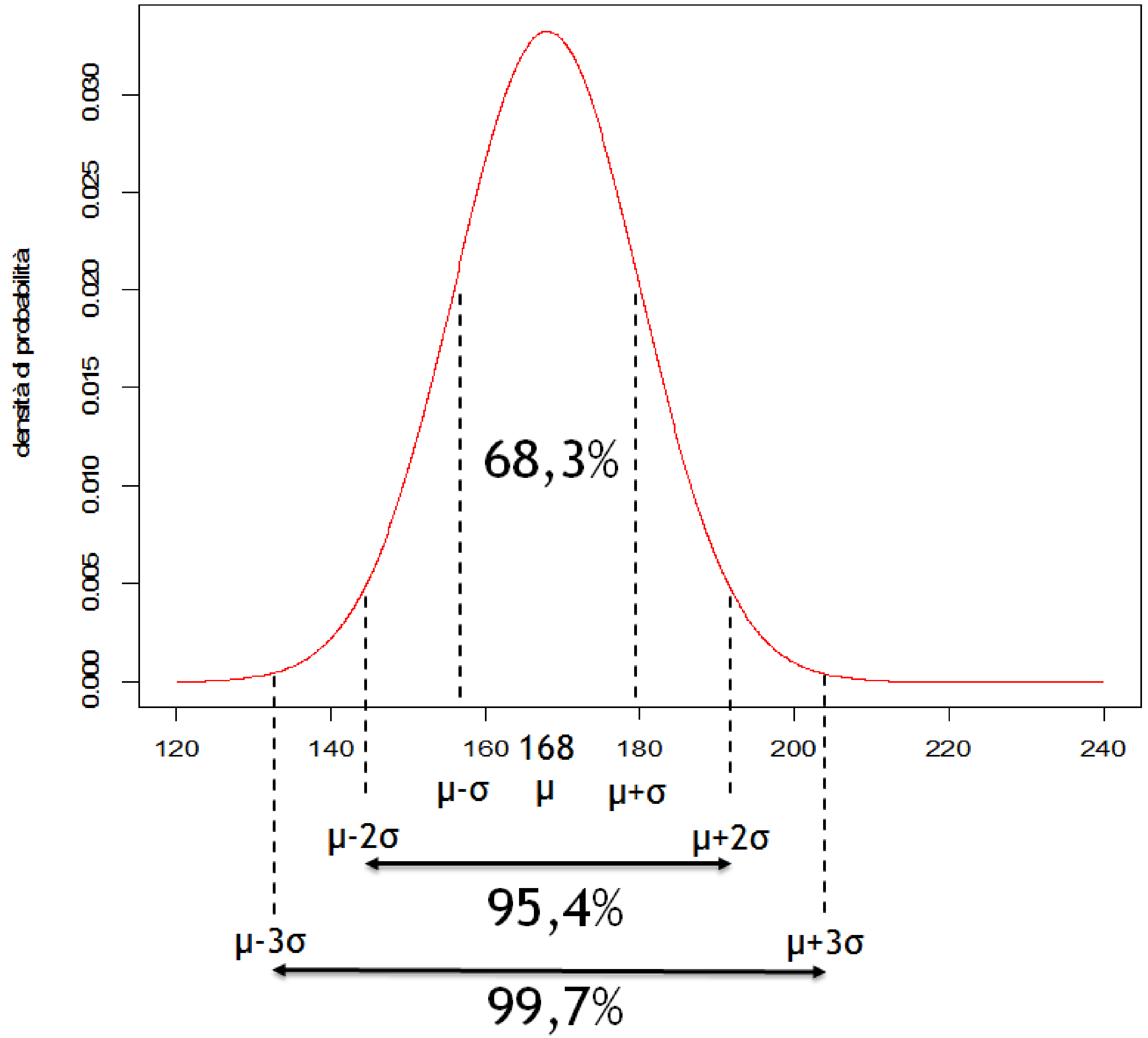
```
> donne=dnorm(x, 168, 12) # dnorm(x, μ, σ)
```

```
> plot(x, donne, type = "l", xlab="altezza in cm", ylab = "densità di probabilità", col="red")
```





NELLA V.A. NORMALE LA PUNTA MASSIMA SI RAGGIUNGE IN CORRISPONDENZA DELLA MEDIA, IN QUESTO CASO 168 CM. QUESTO SIGNIFICA CHE PIU' DEL 3% DELLE DONNE ITALIANE HA UN'ALTEZZA DI 168,00 CM. MAN MANO CHE CI SPOSTIAMO A DESTRA O SINISTRA CI SONO I CASI DI DONNE PIU' O MENO ALTE, FINO AD ARRIVARE A 140 E 200 CM. OLTRE QUESTI ESTREMI CI SONO POCHISSIMI CASI.



IL 68,3% DEL TOTALE DI VALORI SI TROVA NELL'INTERVALLO:

$$\mu-\sigma \leq x \leq \mu+\sigma$$

$$(168-12) \leq x \leq (168+12)$$

IL 95,4% DEL TOTALE DI VALORI SI TROVA NELL'INTERVALLO:

$$\mu-2\sigma \leq x \leq \mu+2\sigma$$

$$(168-2*12) \leq x \leq (168+2*12)$$

IL 99,7% DEL TOTALE DI VALORI SI TROVA NELL'INTERVALLO:

$$\mu-3\sigma \leq x \leq \mu+3\sigma$$

$$(168-3*12) \leq x \leq (168+3*12)$$

ESEMPIO DI VARIABILE NORMALE

Se volessimo aggiungere il dato sull'altezza relativo agli uomini, possiamo farlo sapendo che l'altezza media è di 178 cm con una variabilità di 15.

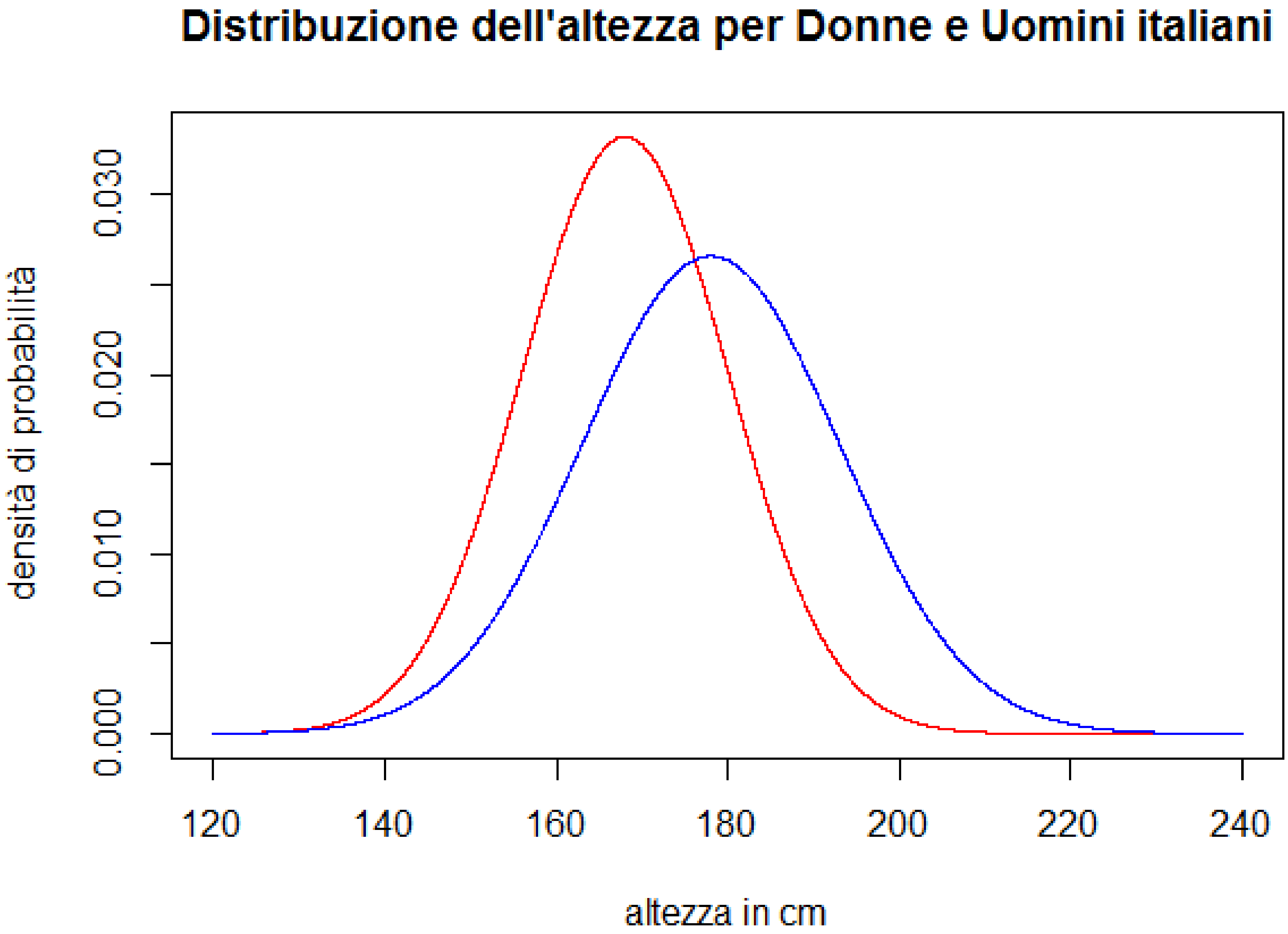
LA FUNZIONE dnorm(x, μ, σ)

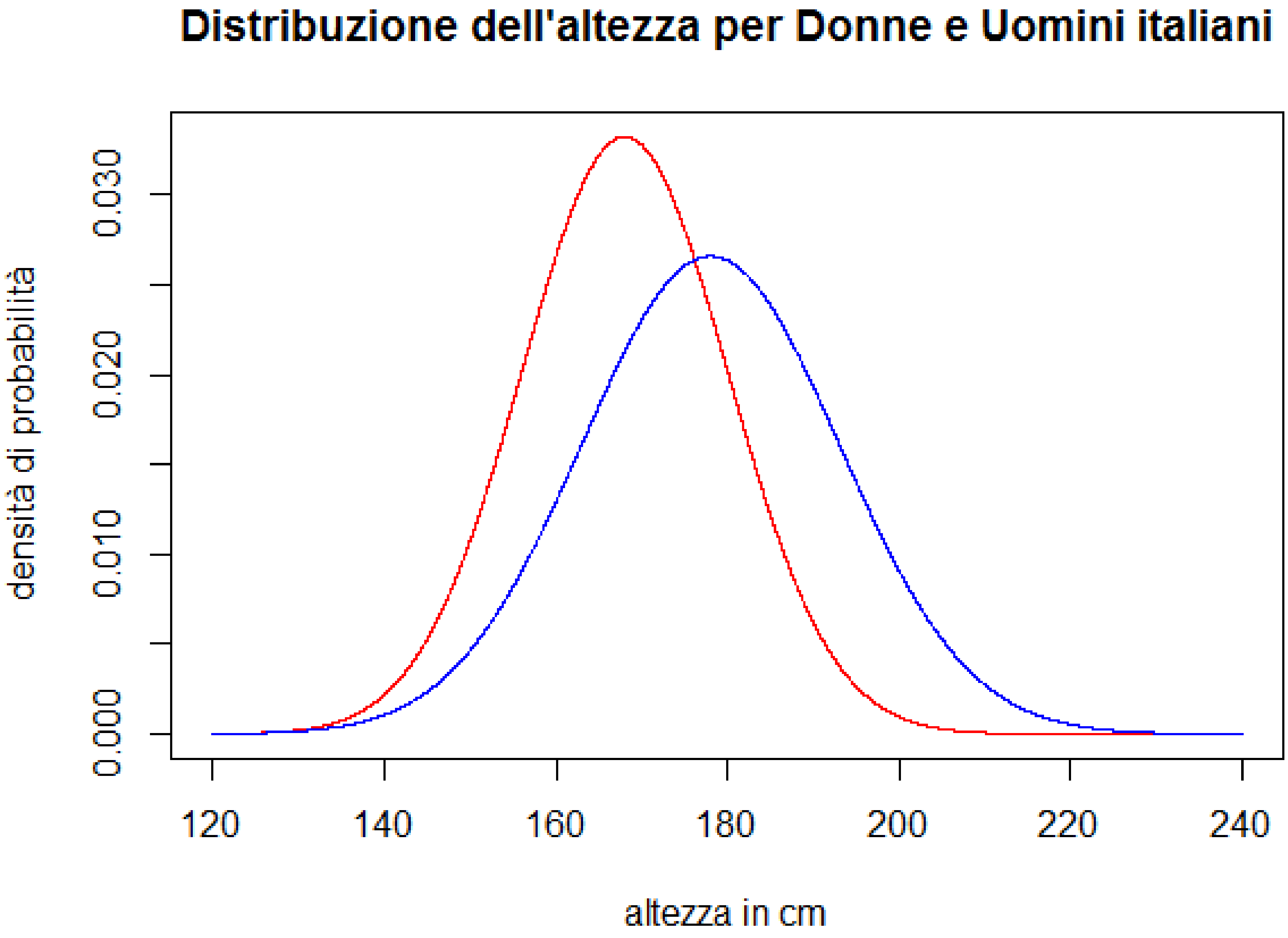
 # DISTRIBUZIONE NORMALE DELL'ALTEZZA DEGLI UOMINI ITALIANI

```
> uomini=dnorm(x, 178, 15)
```

 # CREO IL GRAFICO E AGGIUNGO UN TITOLO

```
>lines(x, uomini, col = "blue")
> title(main="Distribuzione dell'altezza per Donne e Uomini italiani")
```


Distribuzione dell'altezza per Donne e Uomini italiani
densità di probabilità
0.000
0.010
0.020
0.030
120
140
160
180
200
220
240
altezza in cm

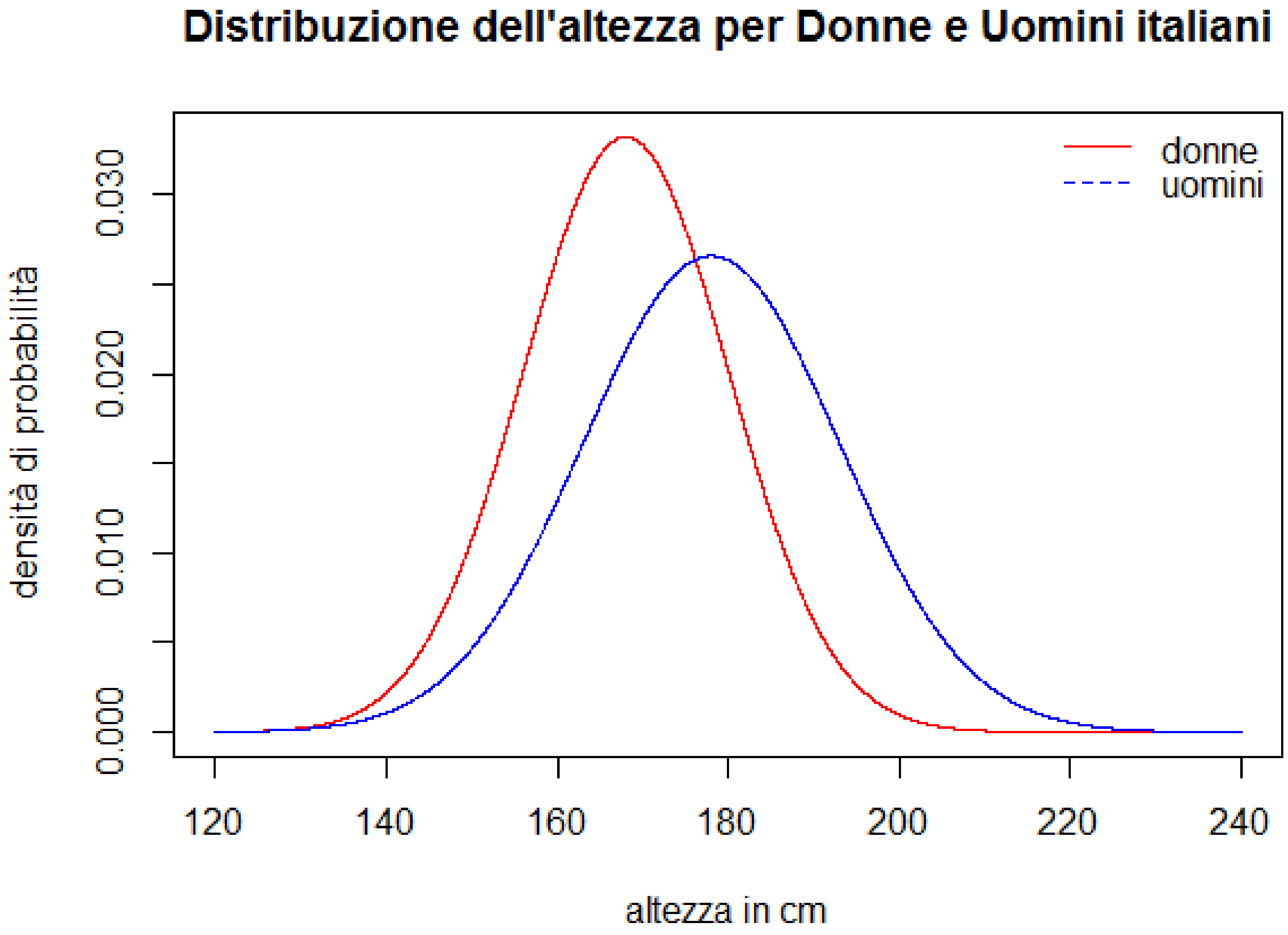


CONFRONTANDO FRA DI LORO LE DISTRIBUZIONI NORMALI DI DONNE E UOMINI, SI PUO' NOTARE COME LA CURVA DEGLI UOMINI SIA PIU' BASSA. QUESTO SUCCEDE PERCHE' HA UNA MAGGIORE VARIABILITA' (INFATTI HA σ DI 15 INVECE CHE DI 12) E CIO' COMPORTA CHE PER GLI UOMINI CI SIANO PIU' CASI ESTREMI, SIA VERSO IL BASSO CHE VERSO L'ALTO. LE FEMMINE CIOE' HANNO UN'ALTEZZA PIU' "REGOLARE", PIU' INTORNO ALLA MEDIA, MENTRE GLI UOMINI HANNO UN'ALTEZZA MEDIA SUPERIORE, MA CI SONO PIU' CASI DI UOMINI MOLTO BASSI O MOLTO ALTI RISPETTO ALLE DONNE.

PER AGGIUNGERE UNA LEGENDA

```
> legend("topright",  c("donne", "uomini"), cex = 1,
bty = "n", col=c("red", "blue"), lty=1:2)
```

 # legend(posizione, etichette, font, bordo, colore, stile linee)



Distribuzione dell'altezza per Donne e Uomini italiani
donne
uomini
densità di probabilità
0.000
0.010
0.020
0.030
120
140
160
180
200
220
240
altezza in cm

ESEMPIO LIVELLO DI GLUCOSIO

Si supponga che il livello di glucosio nella popolazione italiana sia in media di 100 mg/ml con una deviazione standard di 20 mg/ml. (si consiglia di usare un asse delle x da 0 a 200)

ESEMPIO LIVELLO DI GLUCOSIO

 # CREO INNANZITUTTO L'ASSE DELLE X

```
> x=seq(0, 200, by = 0.01)
```

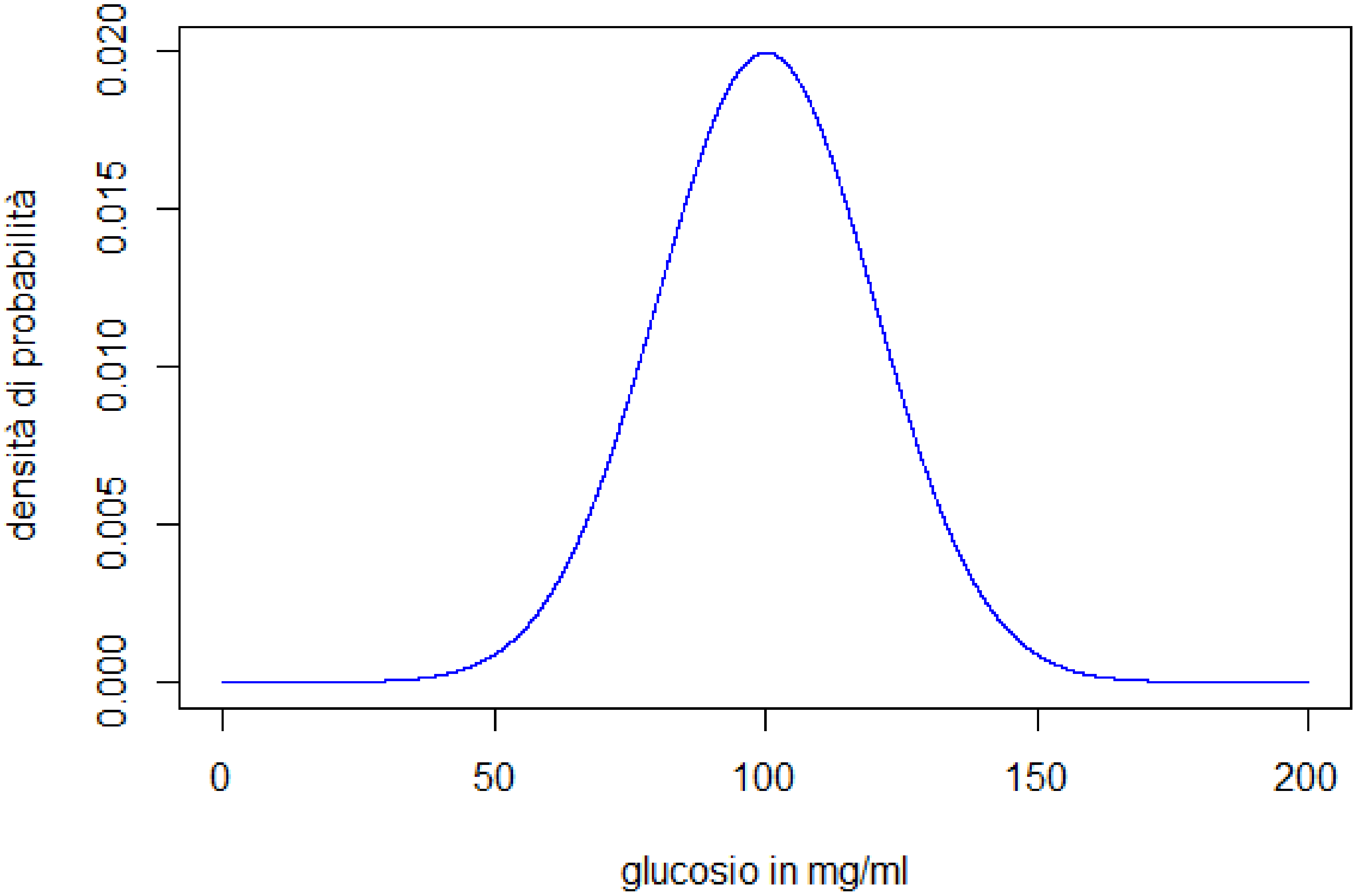
 # CREO LA DISTRIBUZIONE NORMALE

```
> glucosio=dnorm(x, 100, 20)
```

 # CREO IL GRAFICO

```
> plot(x, glucosio, type = "l", xlab="glucosio in mg/ml", ylab = "densità di probabilità", col="blue")
```

ESEMPIO LIVELLO DI GLUCOSIO



ESEMPIO LIVELLO DI GLUCOSIO

Proviamo ad aggiungere il livello di glucosio dei tedeschi, che si distribuisce normalmente con media 110 e deviazione standard 25.

Aggiungiamo anche una legenda per distinguere le due nazionalità.

ESEMPIO LIVELLO DI GLUCOSIO

 # CREO LA DISTRIBUZIONE NORMALE PER I TEDESCHI

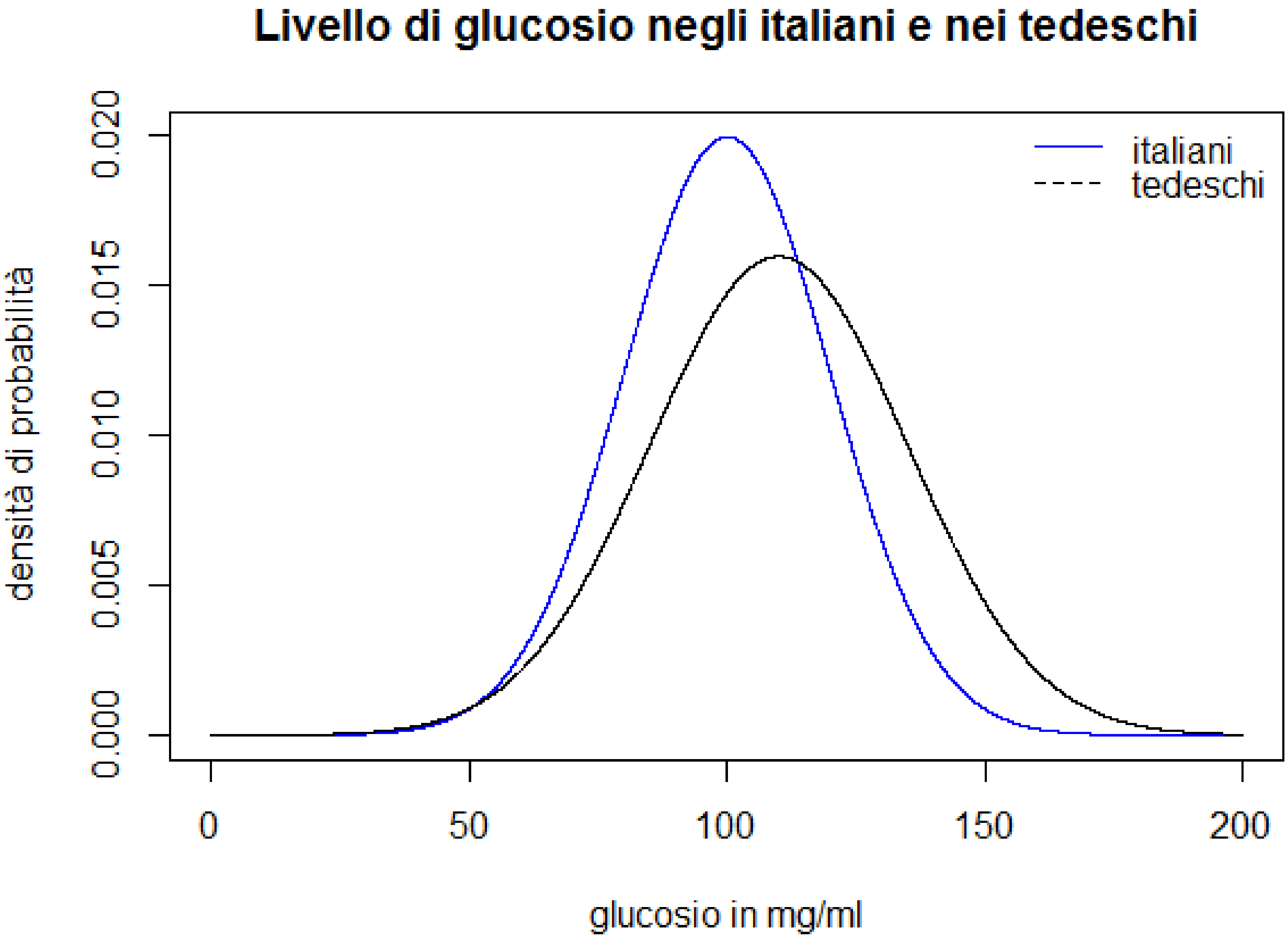
> glucosioted=dnorm(x, 110, 25)

 # CREO IL GRAFICO CON LA LEGENDA

> lines(x, glucosioted, col = "black")

> title(main="Livello di glucosio negli italiani e nei tedeschi")

> legend("topright", c("italiani", "tedeschi"), cex = 1, bty = "n", col=c("blue", "black"), lty=1:2)


Livello di glucosio negli italiani e nei tedeschi
italiani
tedeschi
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
50
100
150
200
glucosio in mg/ml

LA FUNZIONE pnorm

Con la funzione pnorm() otteniamo la curva della probabilità cumulativa della normale. La sintassi è uguale a quella di dnorm(), ma invece dell'altezza della curva ora calcoliamo l'area relativa (l'area totale = 1) sotto la curva dal valore dato di X fino a + infinito o - infinito. Di default R si basa sulla "coda" inferiore, ossia l'area da meno infinito a X. Settando lower.tail = FALSE si usa in la coda superiore. (N.B.: quando X è la media la probabilità di osservare un valore pari o minore a X è sempre 0.5 perché la distribuzione è simmetrica.)

ESEMPIO TARTARUGA GALAPAGOS

La lunghezza di una specie di tartaruga gigante delle Galapagos è risultata distribuirsi come una normale con media pari a 60 cm e deviazione standard di 20 cm. Costruire il grafico e calcolare:

- probabilità fino a 50 cm
- probabilità di tartarughe > 50 cm
- probabilità fra 30 e 90 cm
- quale valore include il 70% delle tartarughe?

ESEMPIO TARTARUGA GALAPAGOS

 # CREO INNANZITUTTO L'ASSE DELLE X

> x=seq(0, 120, 0.01)

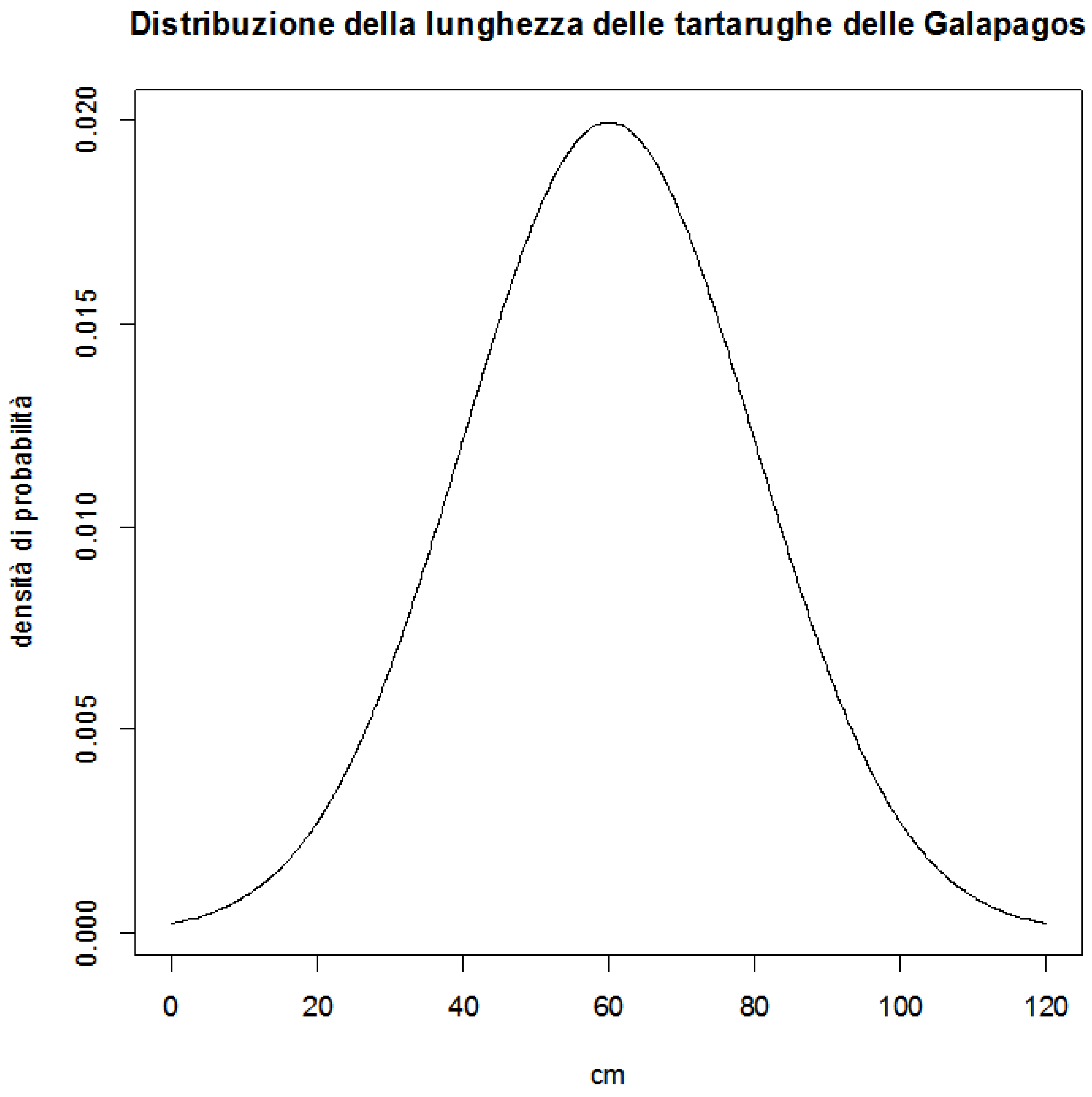
 # CREO LA DISTRIBUZIONE NORMALE

> tartarughe=dnorm(x, 60, 20)

 # CREO IL GRAFICO

> plot(x, tartarughe, type = "l", xlab="cm", ylab = "densità di probabilità")

title(main="Distribuzione della lunghezza delle tartarughe delle Galapagos")


Distribuzione della lunghezza delle tartarughe delle Galapagos
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
20
40
60
80
100
120
cm

PROBABILITA’ < 50

 # CALCOLIAMO LA PROBABILITA’ DI AVERE UNA LUNGHEZZA FINO A 50 CM.
PER CONOSCERE TUTTI I VALORI A SINISTRA DI UN CERTO PUNTO DI CUTOFF (OVVERO FINO A 5) SI SCRIVE:

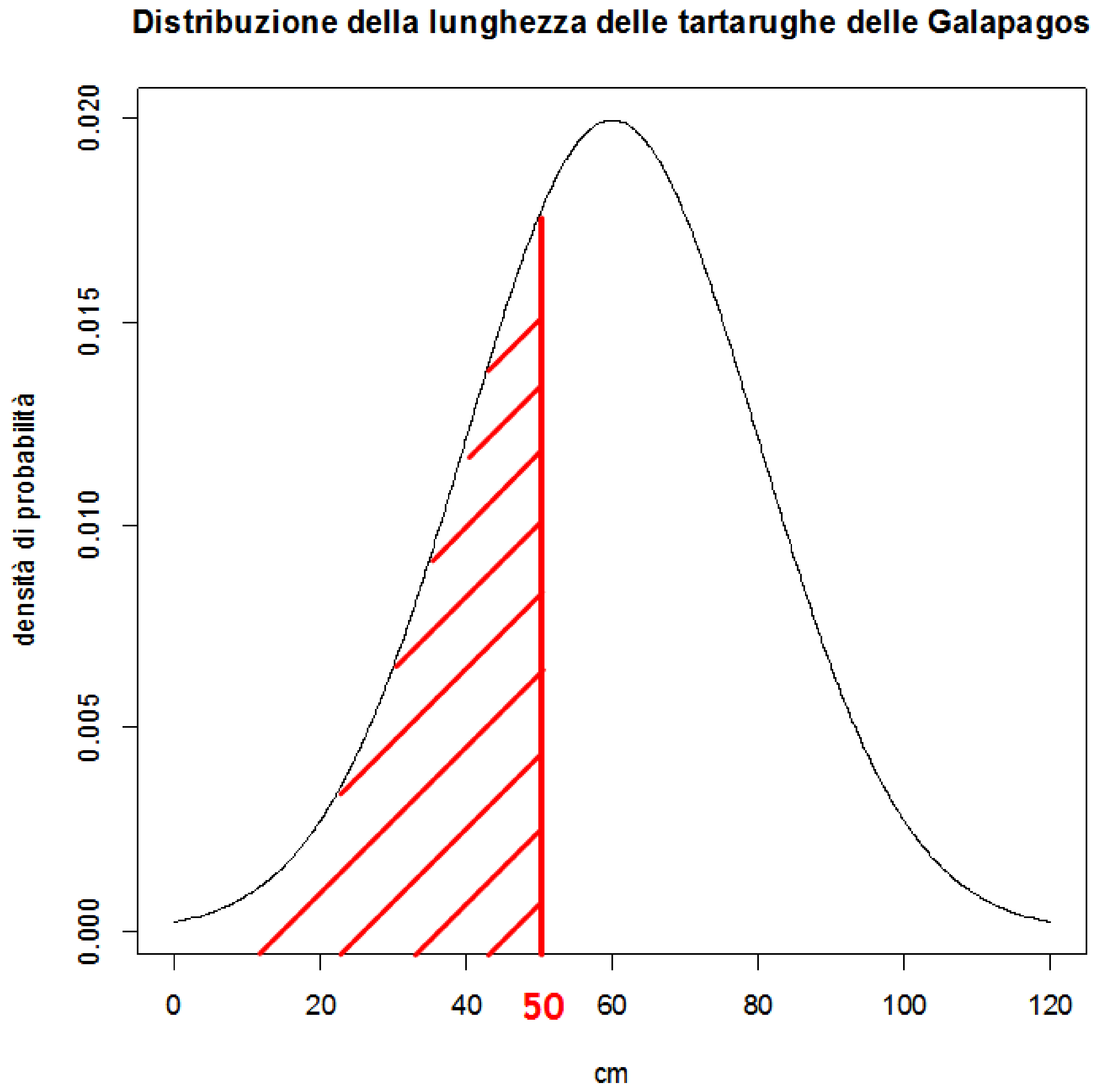
```
> pnorm(50, mean=60, sd=20, lower.tail=TRUE)
[1] 0.3085375
```

 # DOVE:

lower.tail=TRUE INDICA CHE VOGLIO SAPERE I VALORI A SINISTRA DI UN DETERMINATO PUNTO

lower.tail STA PER CODA PIU’ BASSA, OSSIA A SINISTRA

lower.tail=FALSE INDICA INVECE CHE VOGLIO SAPERE I VALORI A DESTRA


Distribuzione della lunghezza delle tartarughe delle Galapagos
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
20
40
50
60
80
100
120
cm

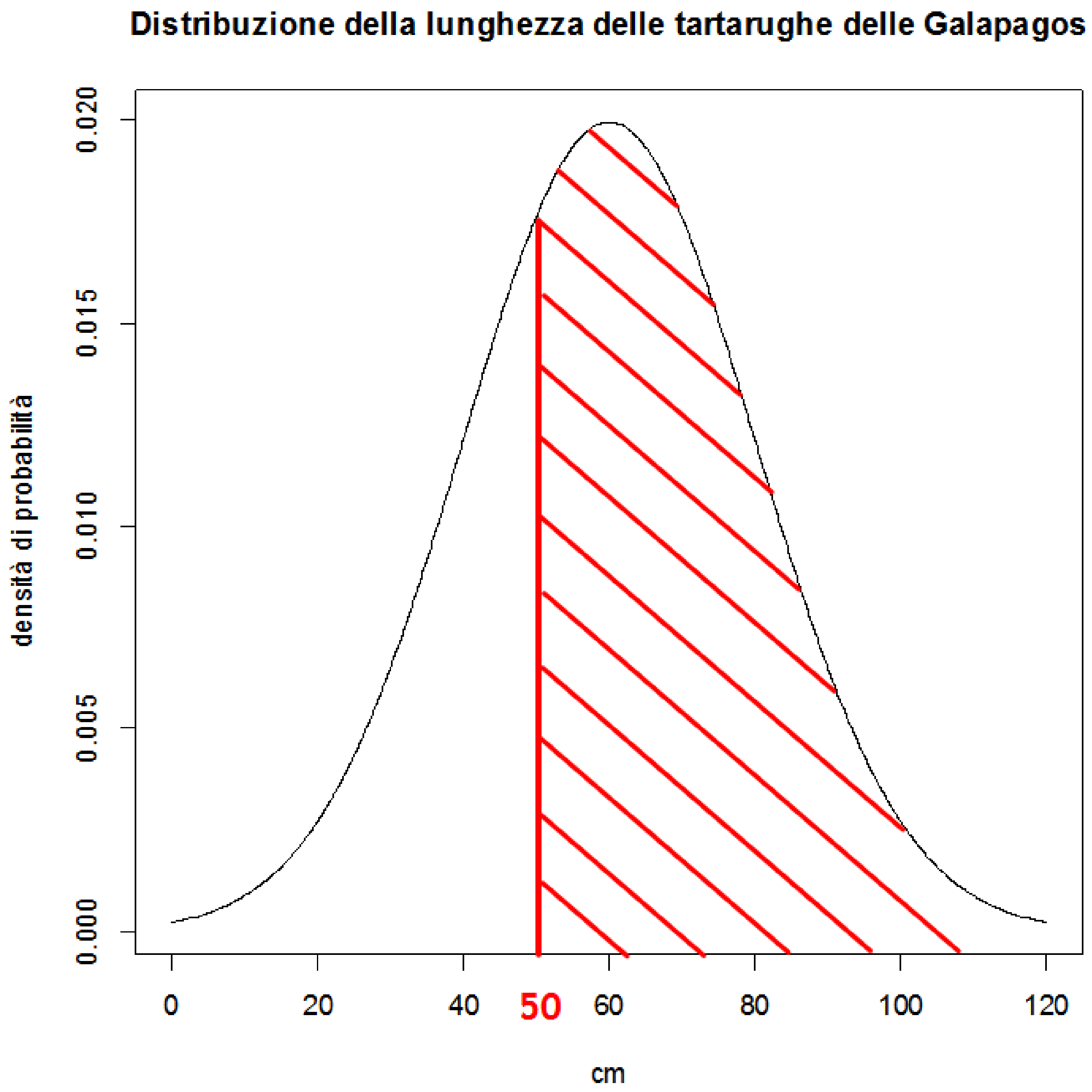
PROBABILITA' > 50

 # CALCOLIAMO LA PROBABILITA' DI AVERE UNA LUNGHEZZA MAGGIORE DI 50 CM.
PER CONOSCERE TUTTI I VALORI A SINISTRA DI UN CERTO PUNTO DI CUTOFF (OVVERO FINO A 5) SI SCRIVE:

```
> pnorm(50, mean=60, sd=20, lower.tail=FALSE)
[1] 0.6914625
```

 # OPPURE:

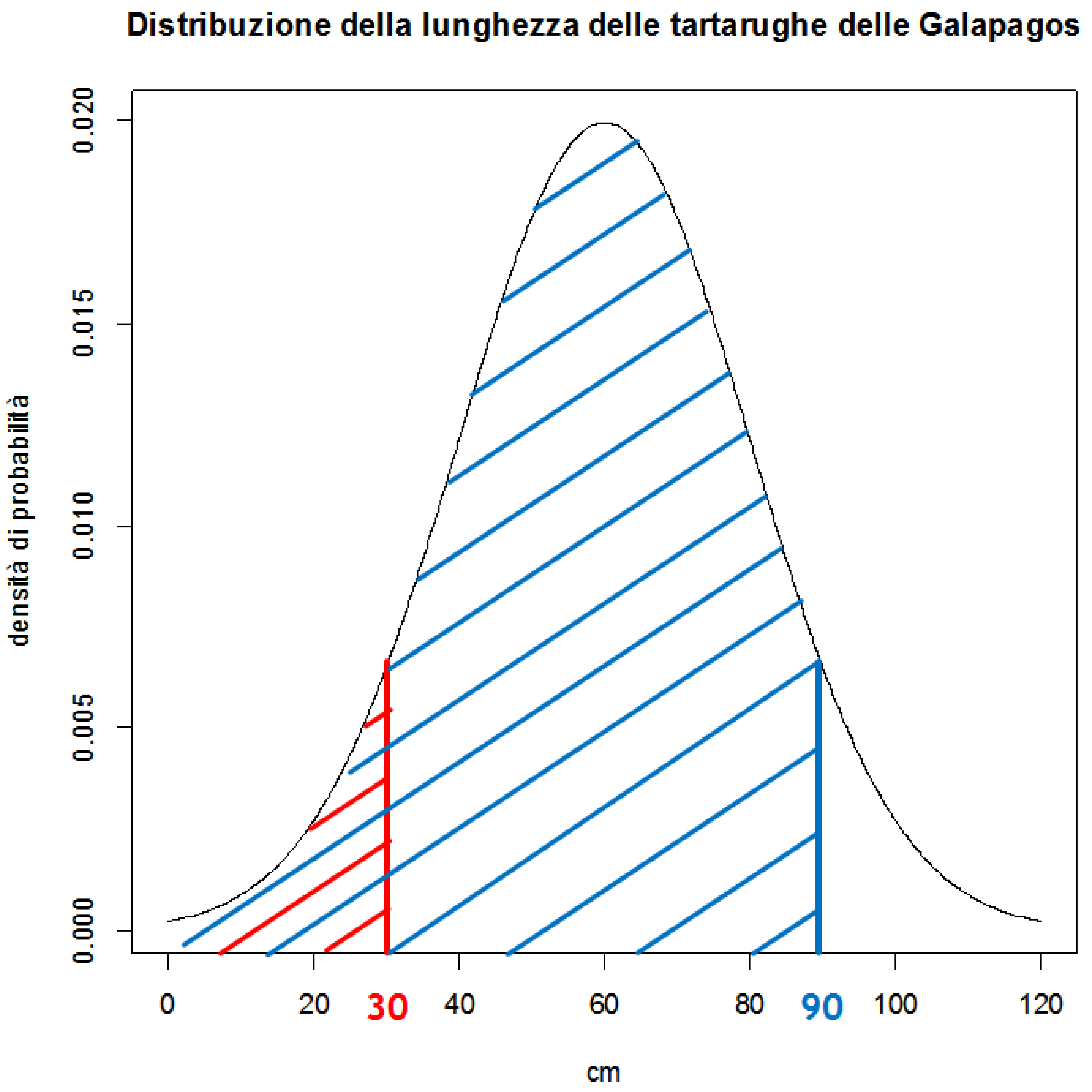
```
> pnorm(50, 60, 20, lower.tail=FALSE)
[1] 0.6914625
```


Distribuzione della lunghezza delle tartarughe delle Galapagos
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
20
40
50
60
80
100
120
cm

PROBABILITA' FRA 30 E 90

 # CALCOLIAMO LA PROBABILITA' DI AVERE UNA X FRA I 30 E I 90 CM

```
> pnorm(90, 60, 20, lower.tail=TRUE) - pnorm(30, 60, 20, lower.tail=TRUE)
[1] 0.8663856
```



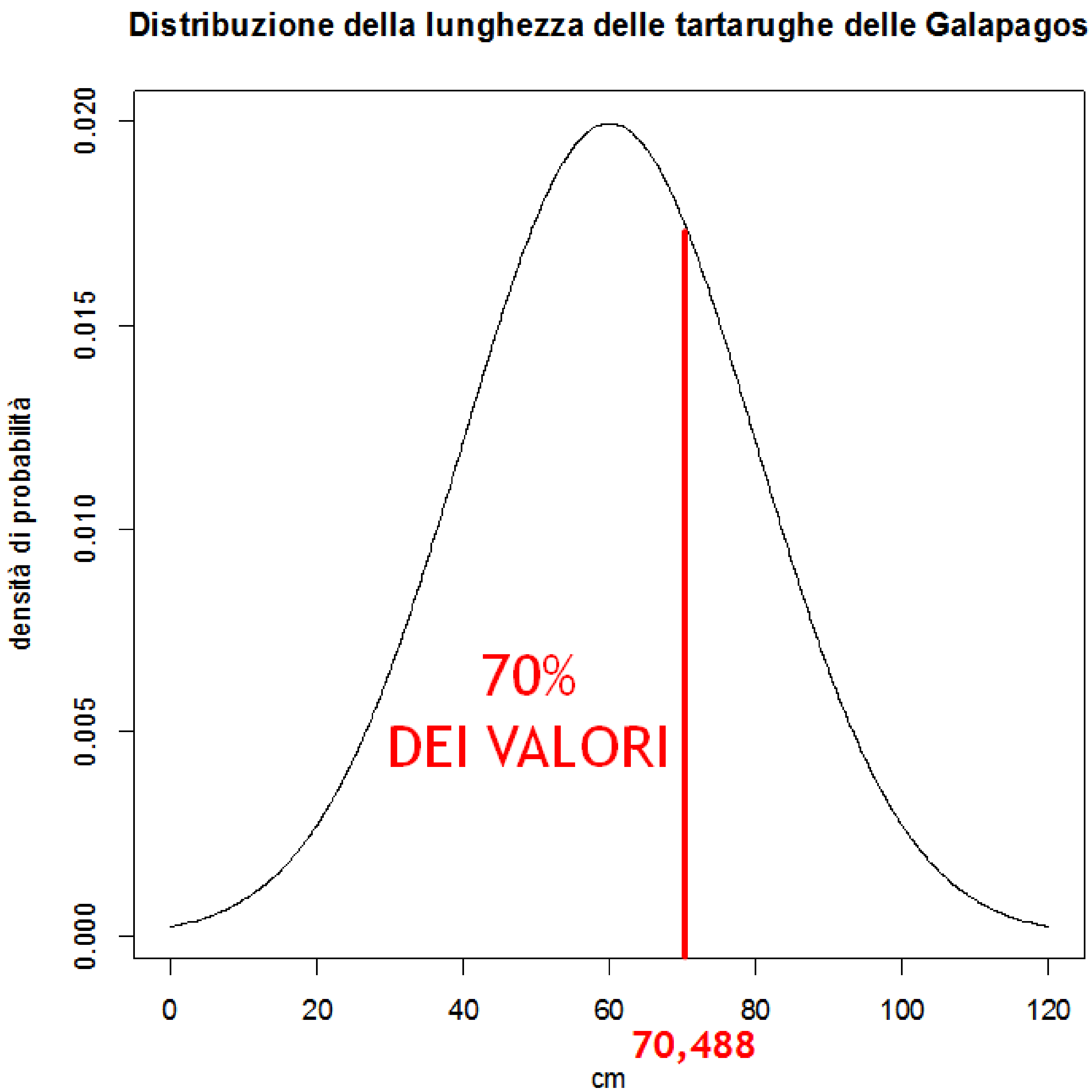
Distribuzione della lunghezza delle tartarughe delle Galapagos
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
20
30
40
60
80
90
100
120
cm

QUALE VALORE INCLUDE IL 70%

 # IN QUESTO CASO DEVO USARE LA FUNZIONE DI RIPARTIZIONE INVERSA ‘’qnorm’’

 # qnorm(% desiderata, media, sd)

```
> qnorm(0.70, 60, 20)
[1] 70.48801
```


Distribuzione della lunghezza delle tartarughe delle Galapagos
densità di probabilità
0.000
0.005
0.010
0.015
0.020
0
20
40
60
80
100
120
cm
70%
DEI VALORI
70,488

ESEMPIO VARIABILE NORMALE

Ipotizziamo di avere dei dati distribuiti come una normale con media 100 cm e deviazione standard 30 cm (si consiglia asse delle X da 0 a 200).

Costruire il grafico e calcolare:

- probabilità > 80 cm
- probabilità fra 50 e 80 cm
- probabilità oltre 100 cm
- quale valore include il 95% della distribuzione?

ESEMPIO VARIABILE NORMALE

CREO INNANZITUTTO L'ASSE DELLE X

```
> x=seq(0, 200, 0.01)
```

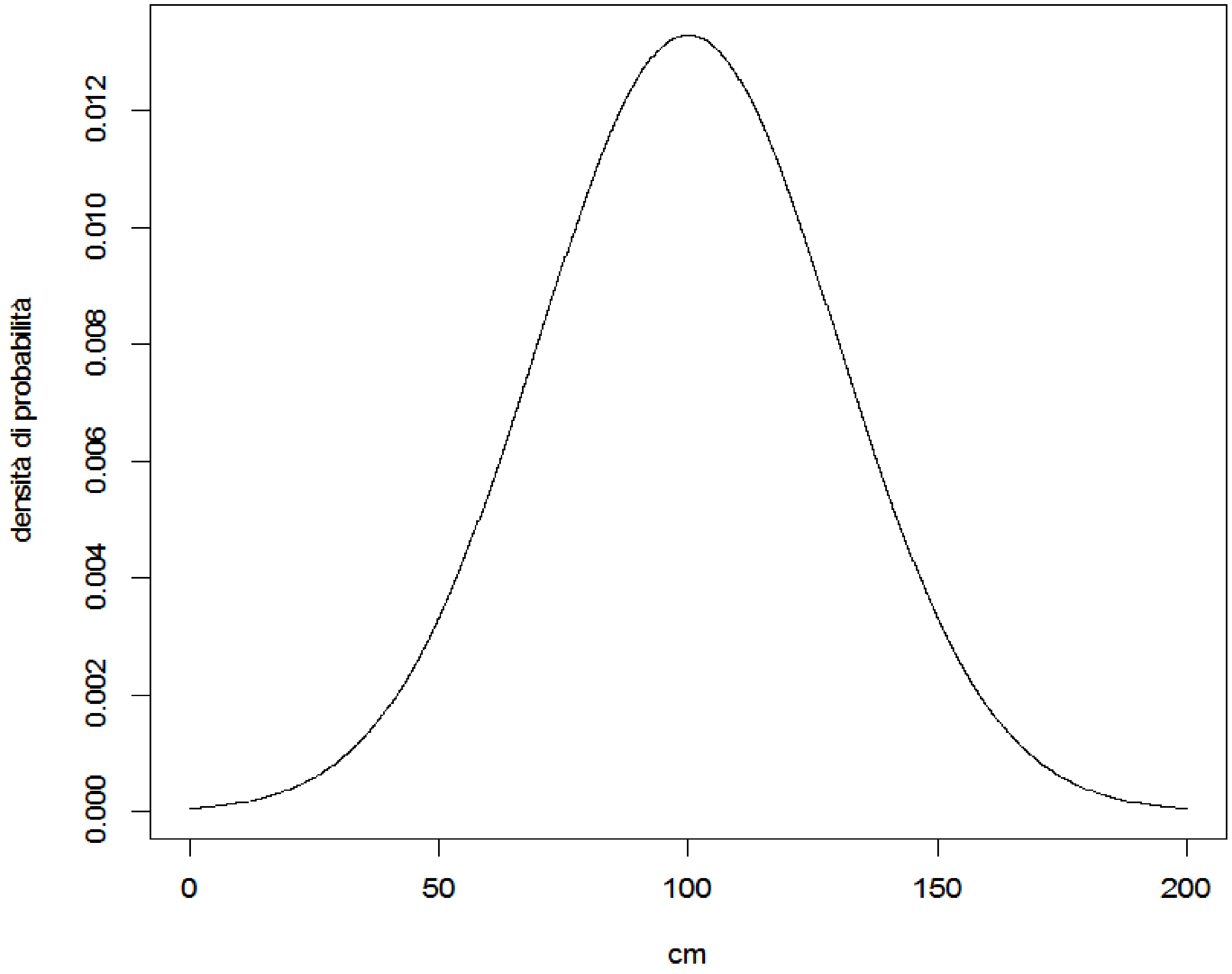
CREO LA DISTRIBUZIONE NORMALE

```
> normale=dnorm(x, 100, 30)
```

CREO IL GRAFICO

```
> plot(x, normale, type = "l", xlab="cm", ylab = "densità di probabilità")
```

ESEMPIO VARIABILE NORMALE



ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA’ DI UNA LUNGHEZZA > 80 CM:

```
> pnorm(80, 100, 30, lower.tail=FALSE)
[1] 0.7475075
```

 # OPPURE:

```
> 1 - pnorm(80, 100, 30, lower.tail=TRUE)
[1] 0.7475075
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' DI UNA LUNGHEZZA FRA 50 E 80 CM:

```
> pnorm(80, 100, 30, lower.tail=TRUE) - pnorm(50, 100, 30, lower.tail=TRUE)
[1] 0.2047022
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' DI UNA LUNGHEZZA MAGGIORE DI 100 CM:

```
> pnorm(100, 100, 30, lower.tail=FALSE)
[1] 0.5
```

ESEMPIO VARIABILE NORMALE

 # QUALE VALORE INCLUDE IL 95% DELLA DISTRIBUZIONE?

```
> qnorm(0.95, 100, 30)
[1] 149.3456
```

ESEMPIO VARIABILE NORMALE

Ipotizziamo di avere dei dati distribuiti come una normale con media 2500 cm e deviazione standard 400 (si consiglia asse delle X da 0 a 5000).

Costruire il grafico e calcolare:

- probabilità > 3000
- probabilità > 2000
- probabilità fra 1600 e 1800
- Probabilità fra 2600 e 2700
- quale valore include il 70% della distribuzione?

ESEMPIO VARIABILE NORMALE

 # CREO INNANZITUTTO L'ASSE DELLE X

```
> x=seq(0, 5000, 0.01)
```

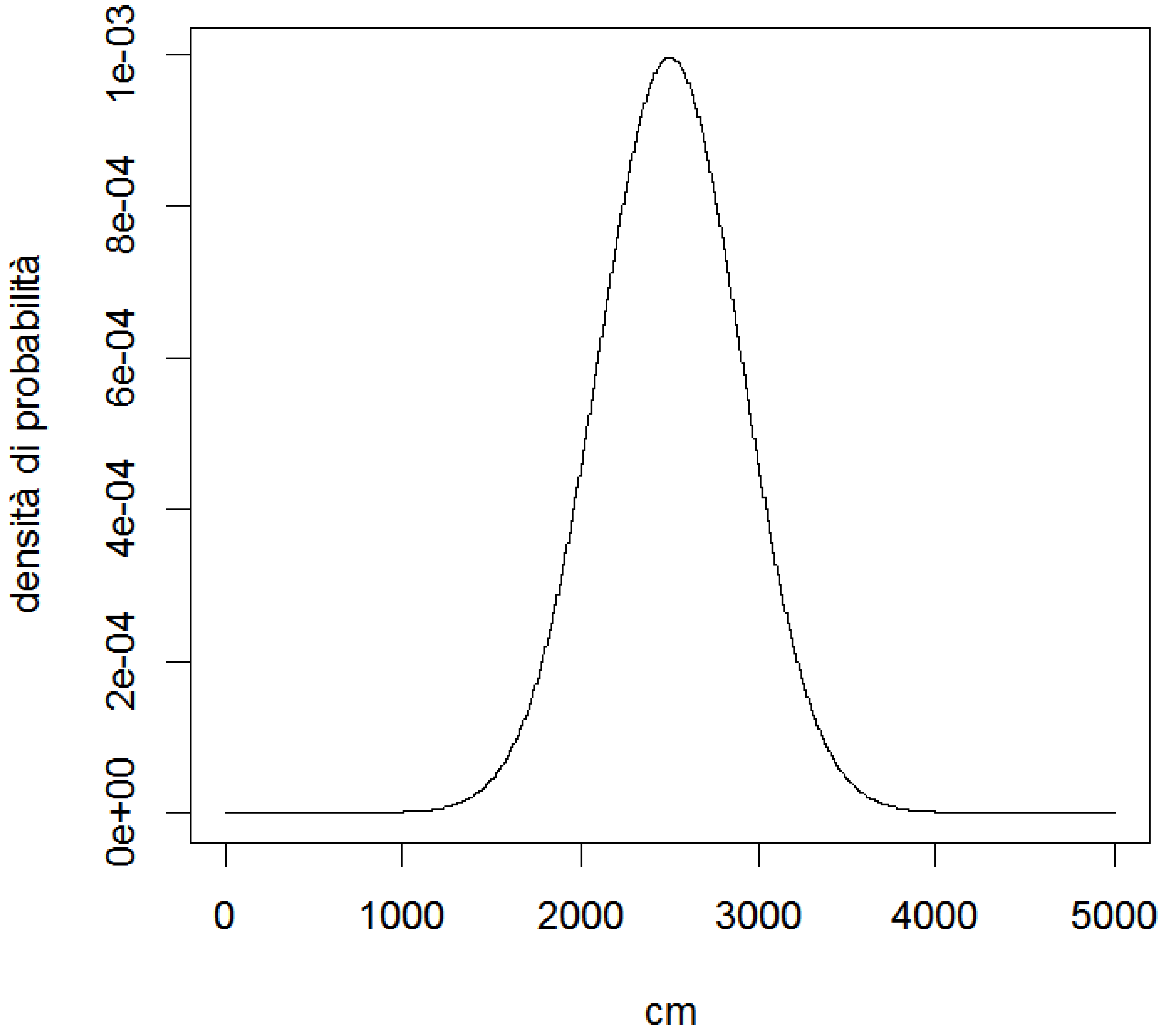
 # CREO LA DISTRIBUZIONE NORMALE

```
> normale=dnorm(x, 2500, 400)
```

 # CREO IL GRAFICO

```
> plot(x, normale, type = "l", xlab="cm", ylab =
"densità di probabilità")
```

ESEMPIO VARIABILE NORMALE



ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' > 3000:

```
> pnorm(3000, 2500, 400, lower.tail=FALSE)
[1] 0.1056498
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' > 2000:

```
> pnorm(2000, 2500, 400, lower.tail=FALSE)
[1] 0.8943502
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' FRA 1600 E 1800:

```
> pnorm(1800, 2500, 400, lower.tail=TRUE)
- pnorm(1600, 2500, 400, lower.tail=TRUE)
[1] 0.02783468
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' FRA 2600 E 2700:

```
> pnorm(2600, 2500, 400, lower.tail=FALSE)
- pnorm(2700, 2500, 400, lower.tail=FALSE)
[1] 0.09275614
```

 # OPPURE:

```
> pnorm(2700, 2500, 400, lower.tail=TRUE)-
pnorm(2600, 2500, 400, lower.tail=TRUE)
[1] 0.09275614
```

ESEMPIO VARIABILE NORMALE

 # QUALE VALORE INCLUDE IL 70% DELLA DISTRIBUZIONE?

```
> qnorm(0.70, 2500, 400)
[1] 2709.76
```

ESEMPIO VARIABILE NORMALE

Ipotizziamo di avere dei dati distribuiti come una normale con media 50 cm e deviazione standard 12 (si consiglia asse delle X da 0 a 100).

Disegnare il grafico e calcolare:

- probabilità = 42
- probabilità < 42
- probabilità > 42
- quale valore include il 50% della distribuzione?

ESEMPIO VARIABILE NORMALE

 # CREO INNANZITUTTO L'ASSE DELLE X

> x=seq(0, 100, 0.01)

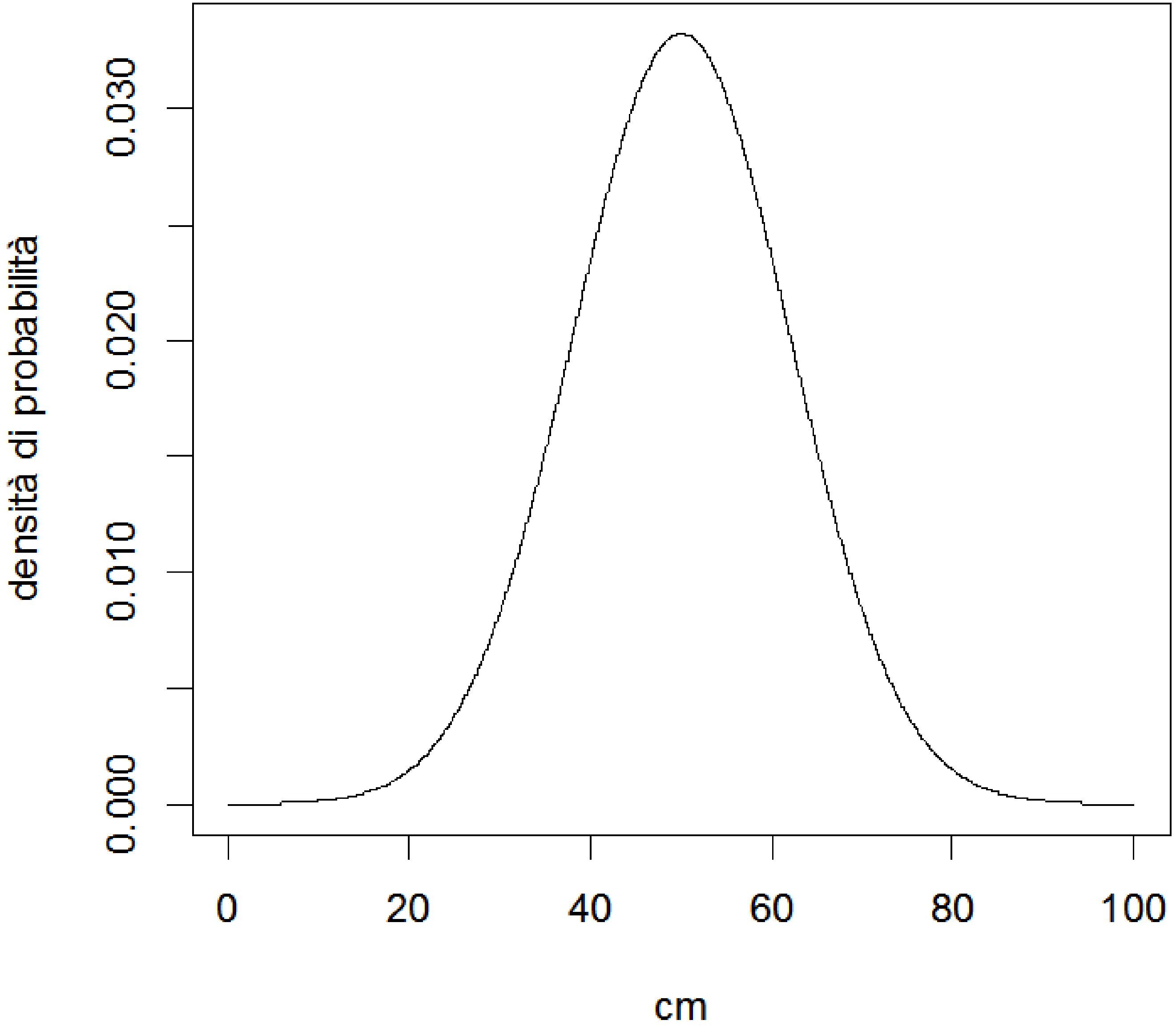
 # CREO LA DISTRIBUZIONE NORMALE

> normale=dnorm(x, 50, 12)

 # CREO IL GRAFICO

> plot(x, normale, type = "l", xlab="cm", ylab = "densità di probabilità")

ESEMPIO VARIABILE NORMALE



ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' = 42:

```
> dnorm(42, 50, 12)
[1] 0.02662067
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA’ < 42:

```
> pnorm(42, 50, 12, lower.tail=TRUE)
[1] 0.2524925
```

ESEMPIO VARIABILE NORMALE

 # PER CONOSCERE LA PROBABILITA' > 42:

```
> pnorm(42, 50, 12, lower.tail=FALSE)
[1] 0.7475075
```

ESEMPIO VARIABILE NORMALE

 # QUALE VALORE INCLUDE IL 50% DELLA DISTRIBUZIONE?

```
> qnorm(0.50, 50, 12)
[1] 50
```

LA FUNZIONE rnorm

 # UN’ALTRA FUNZIONE MOLTO UTILE CON LA NORMALE E’ rnorm, CHE GENERA NUMERI CASUALI DISTRIBUITI SECONDO UNA FORMA NORMALE

rnorm(n° di valori, media, sd)

LA FUNZIONE rnorm

 # AD ESEMPIO SE VOLESSI 100 NUMERI DISTRIBUITI SECONDO UNA NORMALE CON MEDIA=10 E SD=3 SCRIVO:

```
> normali = rnorm(100, 10, 3)
> normali
```

LA FUNZIONE rnorm

```
[1]  9.101775 10.178300 16.361825  7.658928 12.808514 18.971137 11.145115
[8]  9.216454  6.454659  9.538805 12.869454 12.090412  9.803894  6.655278
[15]  6.241585 11.760953  3.903822 10.905144  9.278810  9.317943  9.890864
[22]  8.133759 10.047835 10.688680 12.494748  6.799698 12.760327 12.629900
[29]  6.360088 10.734453  7.334972 14.342733 10.741129 10.917598  9.164149
[36]  7.129749 11.388610 12.465963 14.450955 12.031190 10.452938  4.128776
[43] 10.863655 11.484951  4.080159  7.375252  9.613027 12.114334  6.187797
[50] 12.523779 13.031138 14.380559  8.751982  5.522560  9.084749  8.999432
[57]  9.742766  7.739000 14.934484 11.810485  9.318454  6.191593 11.875519
[64]  3.671111 12.249314 12.643422 13.872522 12.695860 13.146455 12.868675
[71]  7.621365  7.849916  8.684181  5.223391 10.792060  8.265917 19.709426
[78]  6.745397 12.657500  8.805269 14.345293 16.228451  9.210950 10.345743
[85]  6.121875  9.143650  9.788000 11.109219 10.491586 11.527055 13.094629
[92] 11.959499 10.730225  8.009685 12.722475  8.566456  9.739095  6.389590
[99]  3.503712 14.445061
```

LA FUNZIONE rnorm

 # ESSENDO UNA GENERAZIONE RANDOM, IL RISULTATO CAMBIA OGNI VOLTA CHE FACCIO RIGIRARE LA rnorm

```
> normali = rnorm(100, 10, 3)
> normali
```

LA FUNZIONE rnorm

```
  [1] 12.219959 10.010144  8.274351  6.737842 10.371871  6.825607  9.889556  4.174221  9.470322
 [10] 12.881229 13.006578 11.313214  6.536770  5.896736 10.084139 12.763642 10.679511 13.773569
 [19] 13.908794  5.799971  9.052886  7.448082 10.945705 13.555606  9.036975  9.498445  6.063145
 [28]  7.727221  6.881729 14.124523  5.968233  9.656225  3.736503 11.442649  6.349359 14.573650
 [37]  7.206319  7.583469 11.563384  6.551514 10.817677  8.499822 14.120180  9.751970 14.583080
 [46] 14.519570  8.468743  4.712776 11.886612  7.754050 13.718112  9.552893  7.401971  8.256023
 [55]  4.543036  8.404632 12.043188 14.782622 13.542304  9.434687 11.787563 15.168621  4.977038
 [64] 15.617698  8.032629  8.898195 10.386072 11.733109 11.724578 11.545083  4.353410 12.377442
 [73] 11.092246  6.910395 13.094385 11.075365  7.386838  9.820823 14.202449 10.938280  7.615347
 [82]  9.331125 10.602865 10.778740 10.300893 14.017932  8.462295  8.478741  5.712669  9.851226
 [91]  5.271629  9.890182  9.674378  8.536093 13.979726 10.607006 10.832335 11.892957 11.651635
[100]  8.814592
```