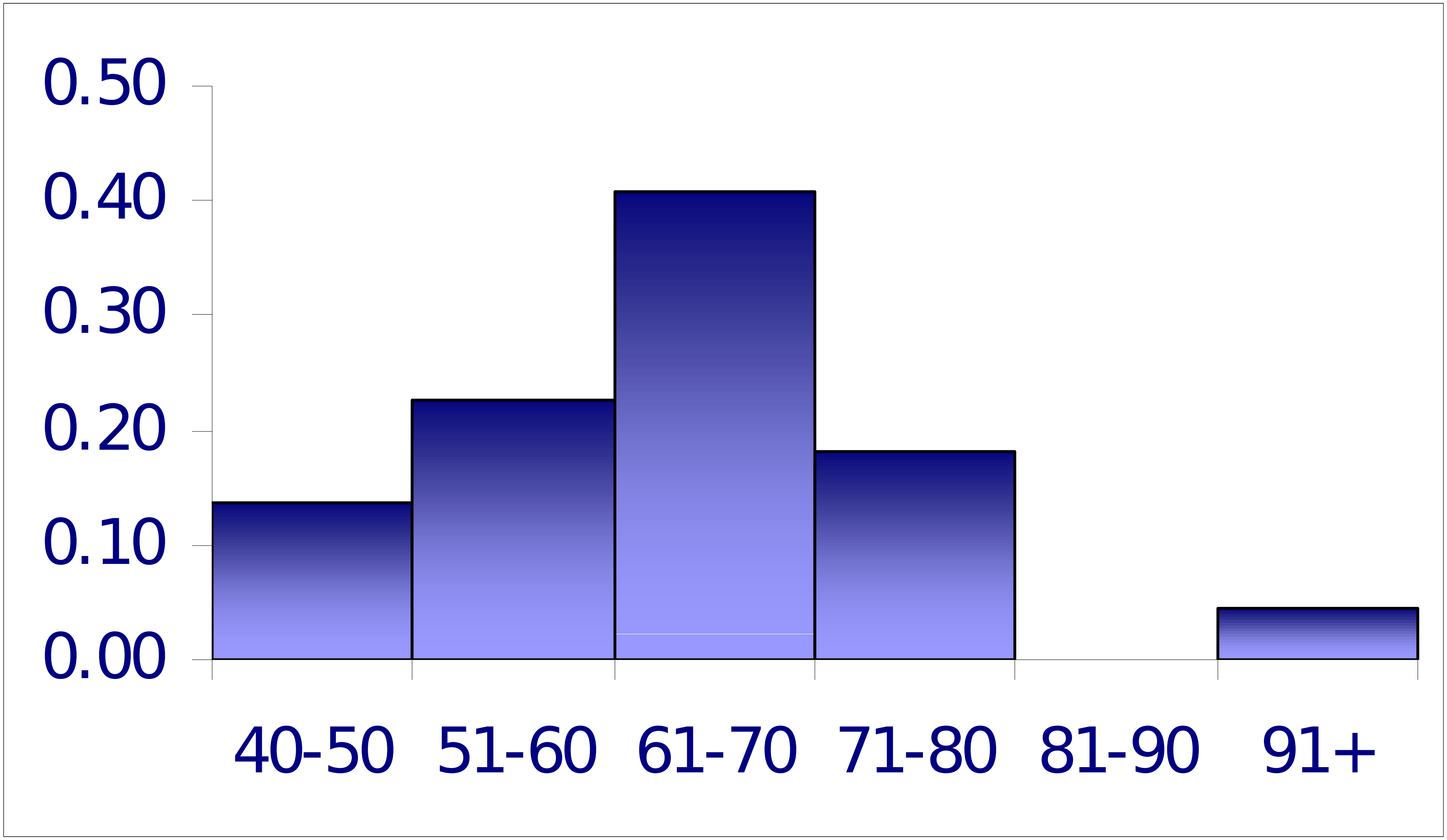
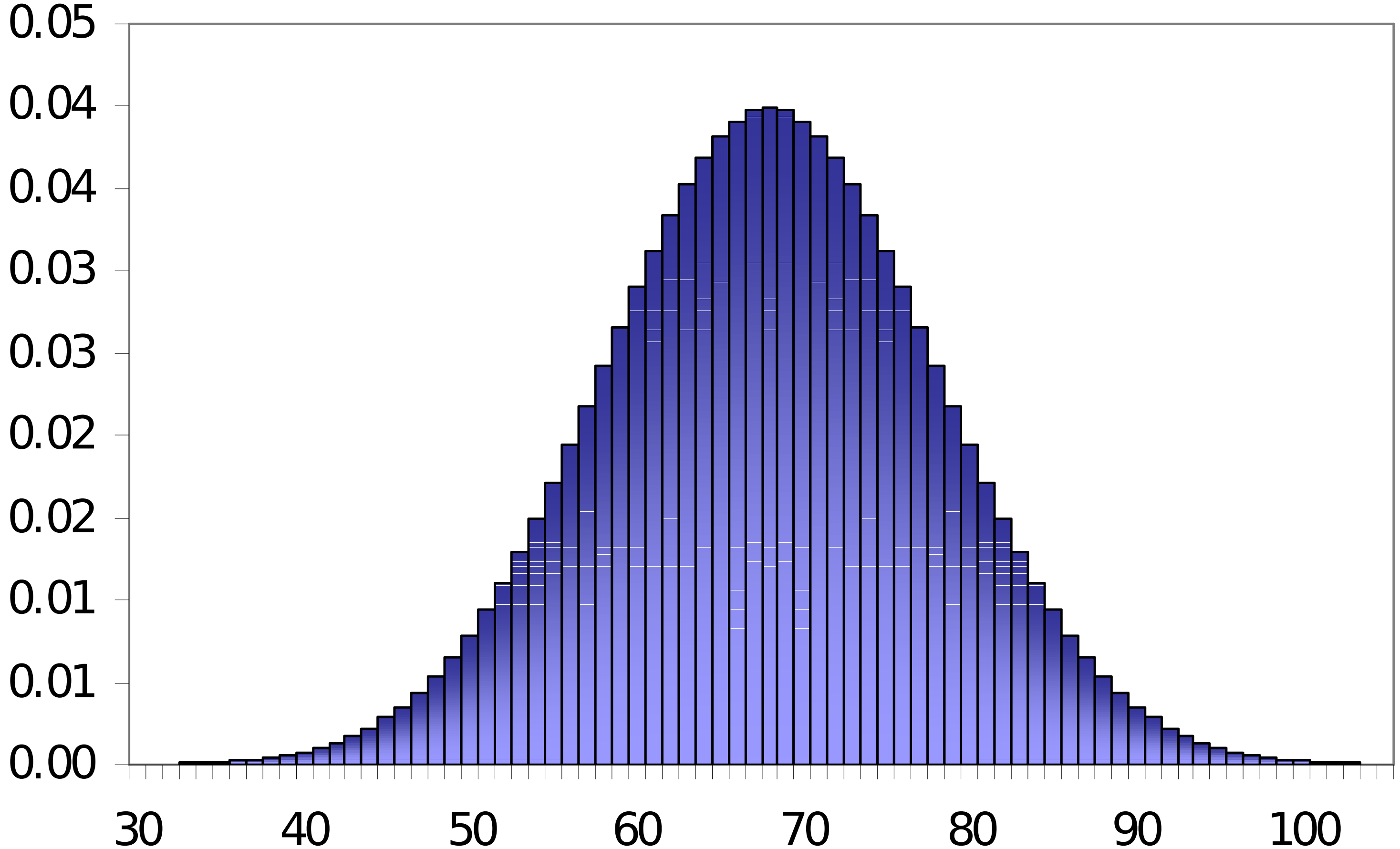
Variabili aleatorie continue: L’ISTOGRAMMA

Peso (kg) di 150 studenti tra i 12-18 anni



Variabili aleatorie continue: L'ISTOGRAMMA

Aumentiamo il n° delle osservazioni: Peso (Kg) di 150.000 studenti tra i 12-18 anni



E’ possibile aumentare il grado di precisione delle misurazioni in modo che le classi di frequenza siano a intervalli di 0.001 kg (1 g) invece che di 1 kg.

Se si considera un numero di osservazioni molto grande a un grado di precisione infinitamente elevato, i gradini dell’istogramma si trasformano in una curva continua simile a quella della distribuzione normale (con un andamento a campana).

La distribuzione normale

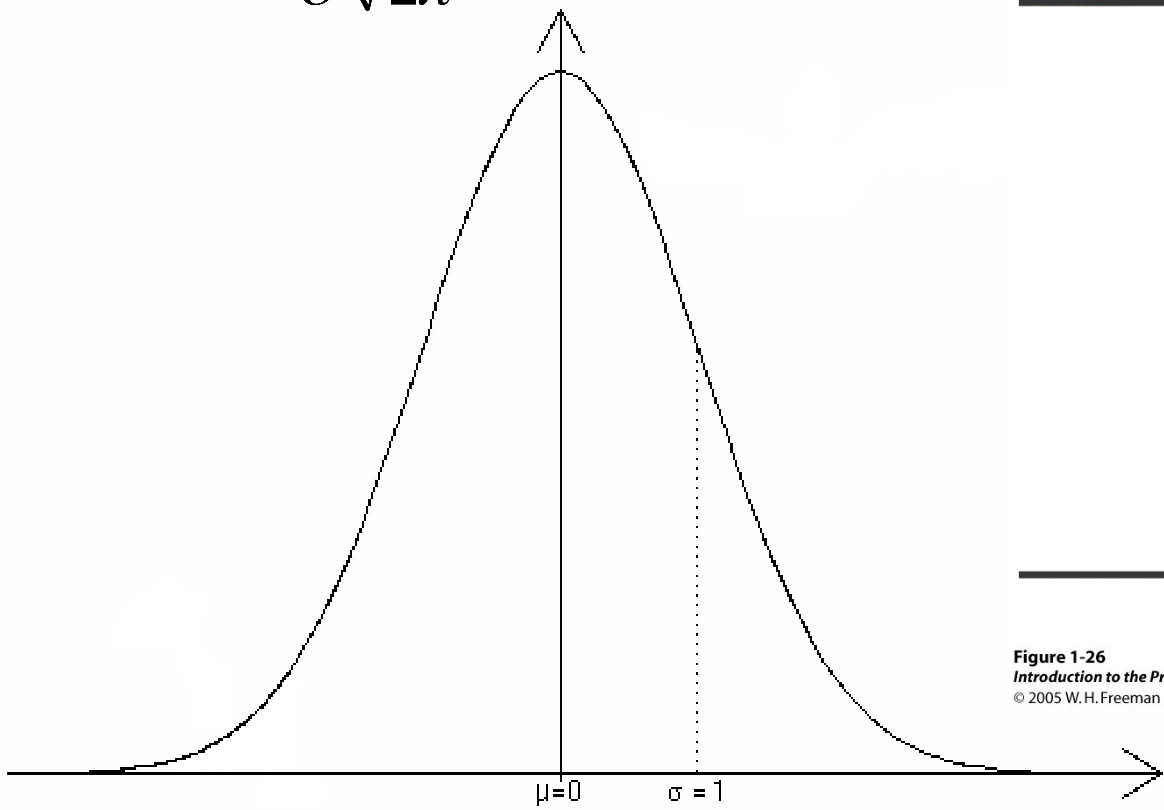
- Tutte le distribuzioni normali hanno la stessa forma generale. La curva di densità per una **particolare** distribuzione normale si ottiene specificando la sua media μ e la sua deviazione standard σ (o la sua varianza σ^2).

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

$$(-\infty < X < +\infty)$$

La Distribuzione Normale

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



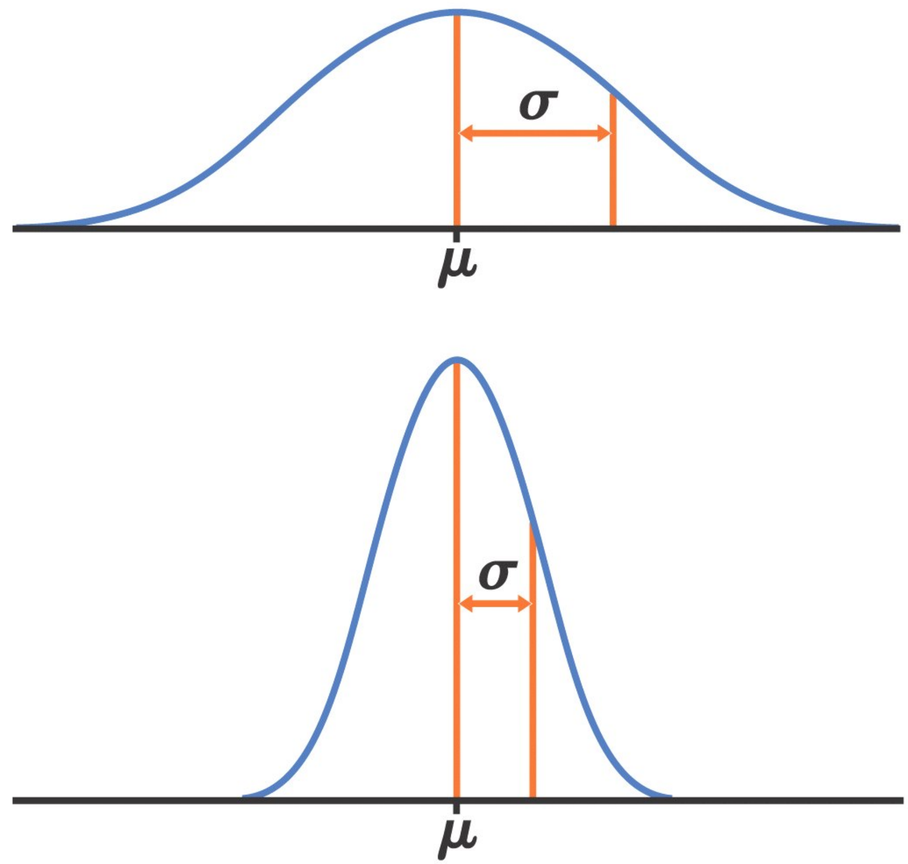
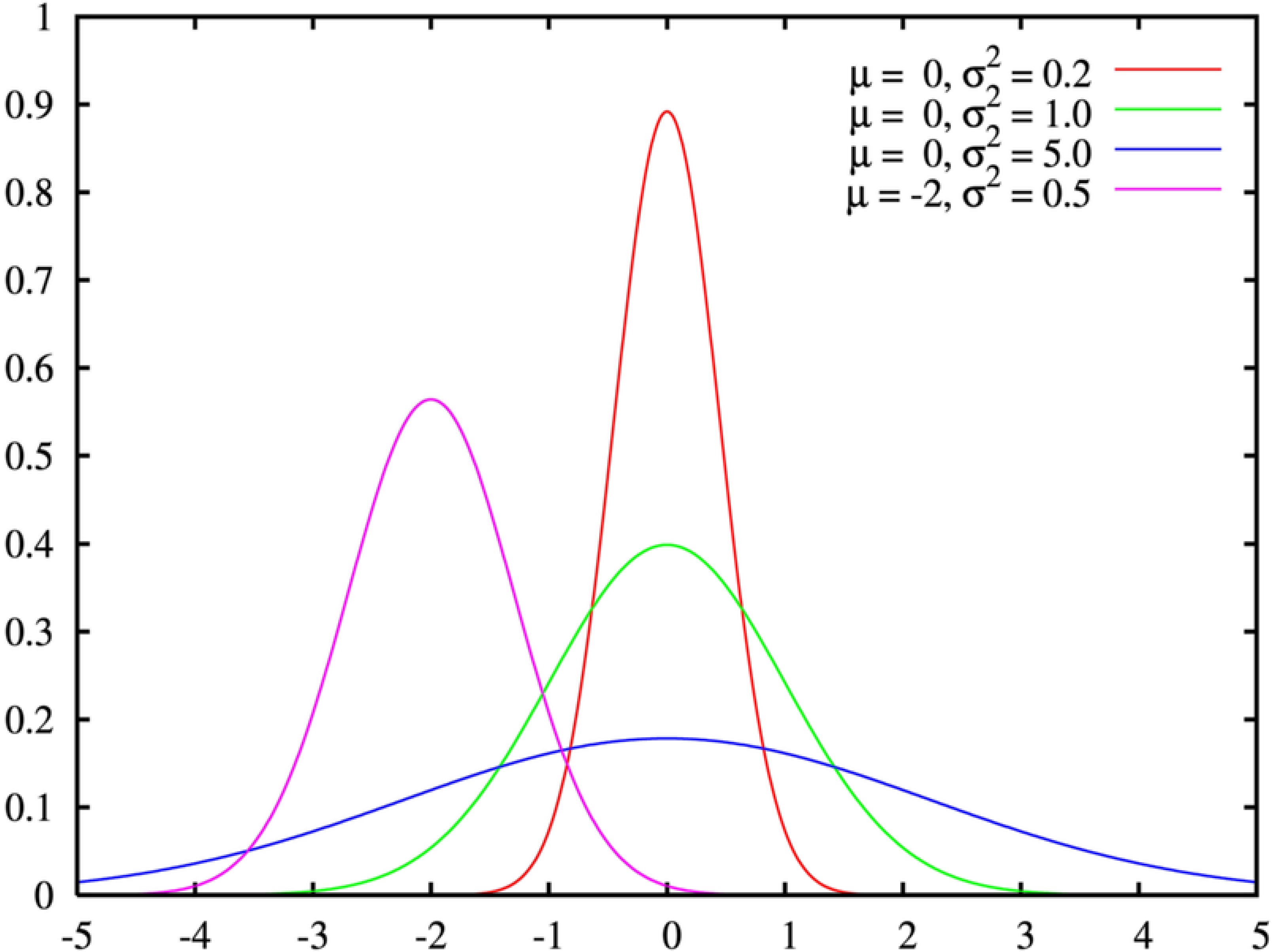


Figure 1-26
Introduction to the Practice of Statistics, Fifth Edition

Rappresentazione grafica di una distribuzione normale

Distribuzione normale N (μ, σ²) al variare di μ e σ

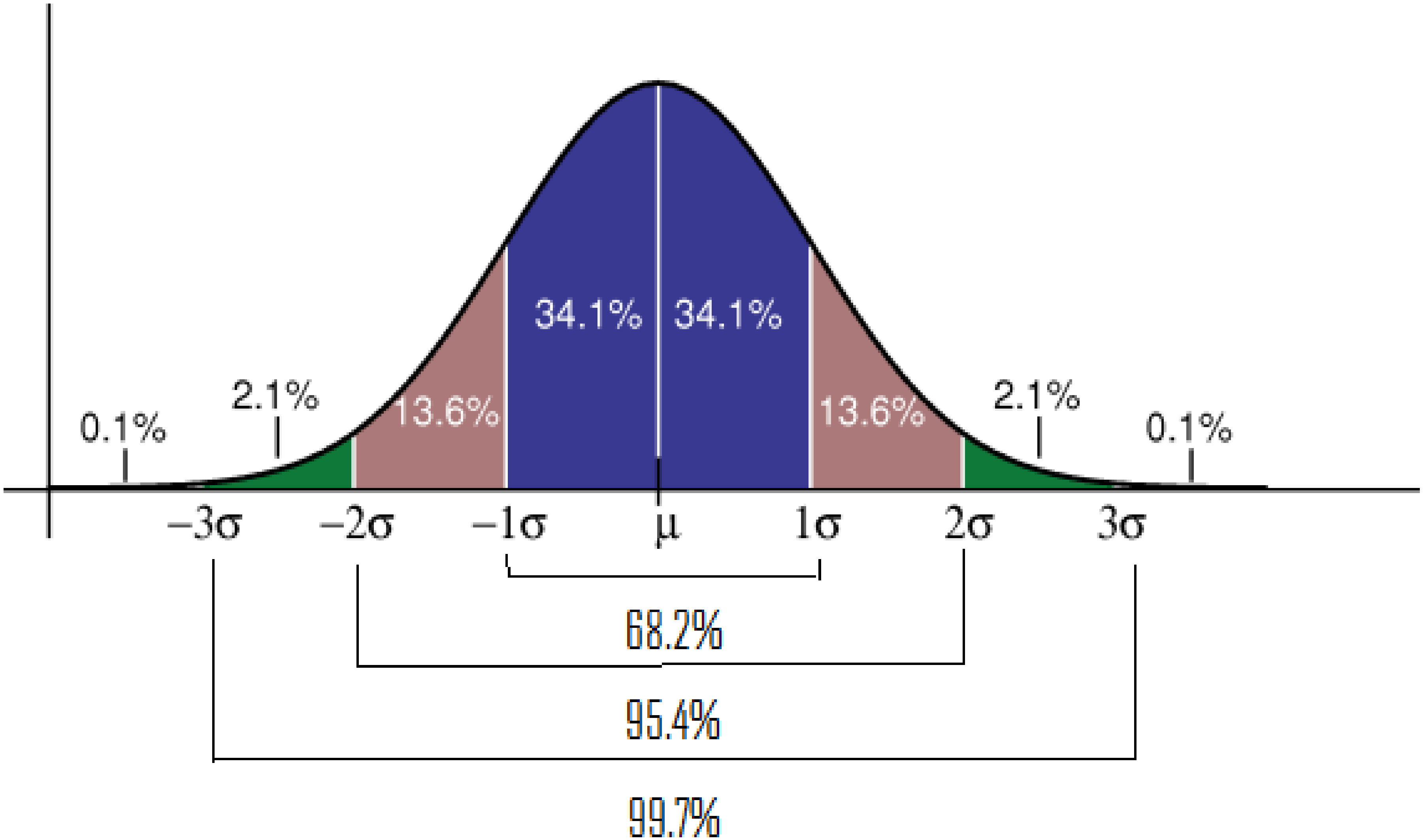


Distribuzione normale

Nella distribuzione normale con media μ e deviazione standard σ:

- il **68.3%** delle osservazioni è compreso nell'intervallo
  $[\mu - \sigma, \mu + \sigma]$
- il **95.5%** delle osservazioni è compreso nell'intervallo
  $[\mu - 2\sigma, \mu + 2\sigma]$
- il **99.7%** delle osservazioni è compreso nell'intervallo
  $[\mu - 3\sigma, \mu + 3\sigma]$

Distribuzione Normale



Insieme di dati normali

Un insieme di dati si dice normale se il rispettivo istogramma

1. Ha il punto di massimo in corrispondenza dell’intervallo centrale.

2. E’ a forma di campana.

3. E’ simmetrico rispetto all’intervallo centrale.

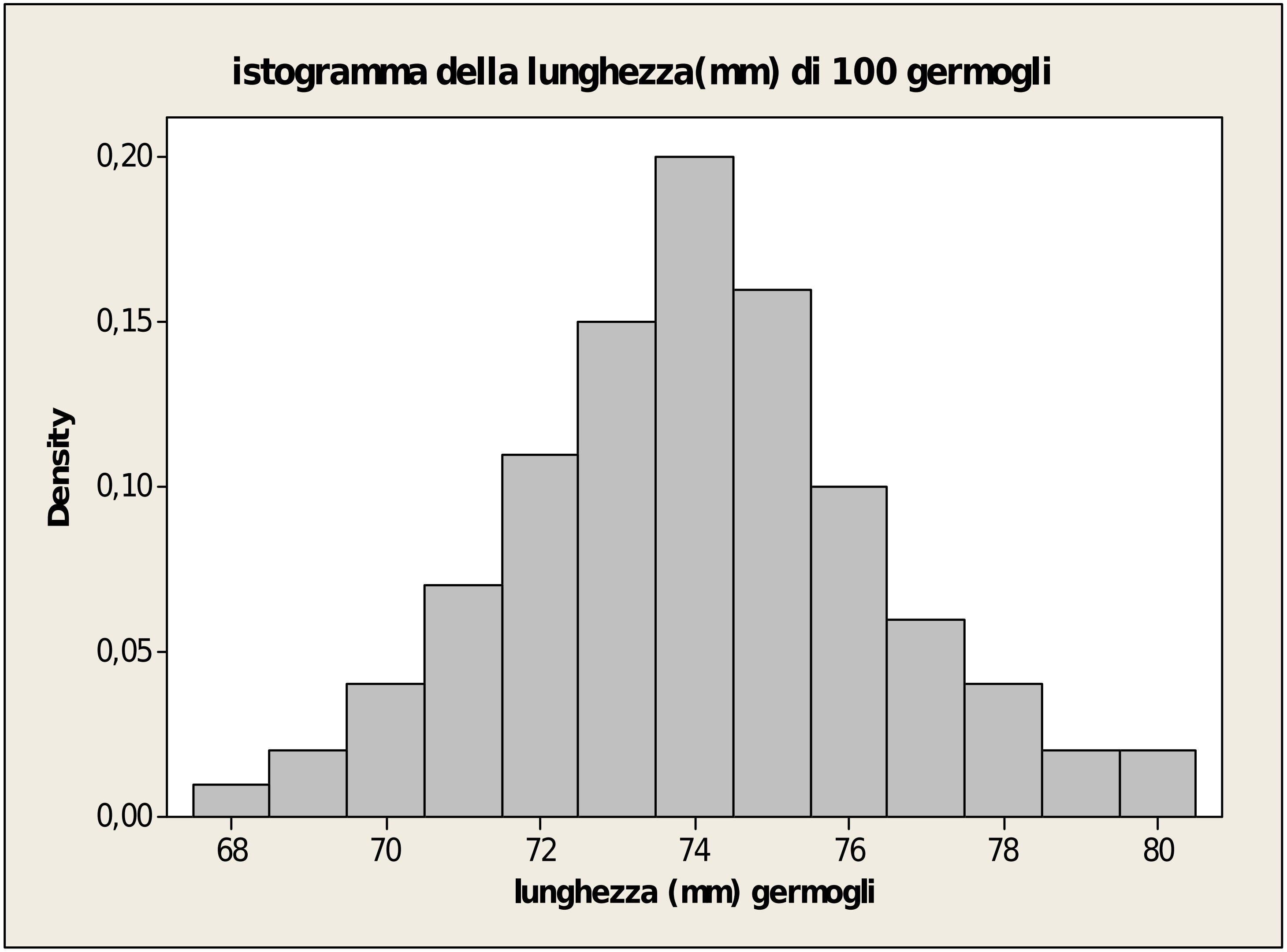
Esempio

- Lunghezze (in mm) di 100 germogli nati da semi piantati allo stesso tempo.

- Classi di ampiezza 1 mm

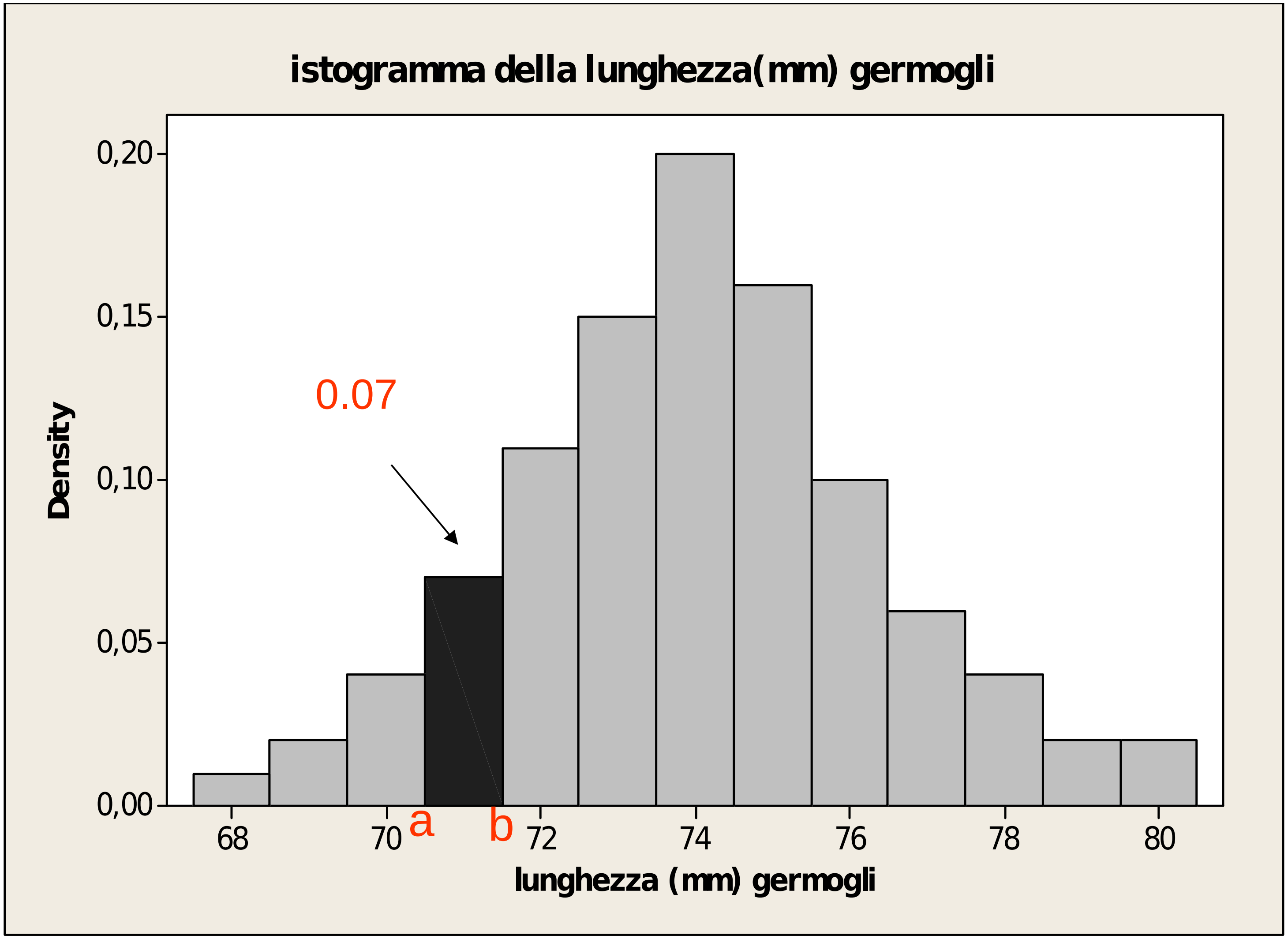
- Tabella delle frequenze, che riporta anche le frequenze relative e delle frequenze cumulate dei valori della variabile “lunghezza dei germogli”.

Intervallo classe	Lunghezza (mm)	Frequenza	Frequenza relativa	Frequenza rel. cumulata
67.5-68.5	68	1	0.01	0.01
68.5-69.5	69	2	0.02	0.03
69.5-70.5	70	4	0.04	0.07
70.5-71.5	71	7	0.07	0.14
71.5-72.5	72	11	0.11	0.25
72.5-73.5	73	15	0.15	0.40
73.5-74.5	74	20	0.20	0.60
74.5-75.5	75	16	0.16	0.76
75.5-76.5	76	10	0.10	0.86
76.5-77.5	77	6	0.06	0.92
77.5-78.5	78	4	0.04	0.96
78.5-79.5	79	2	0.02	0.98
79.5-80.5	80	2	0.02	1.00

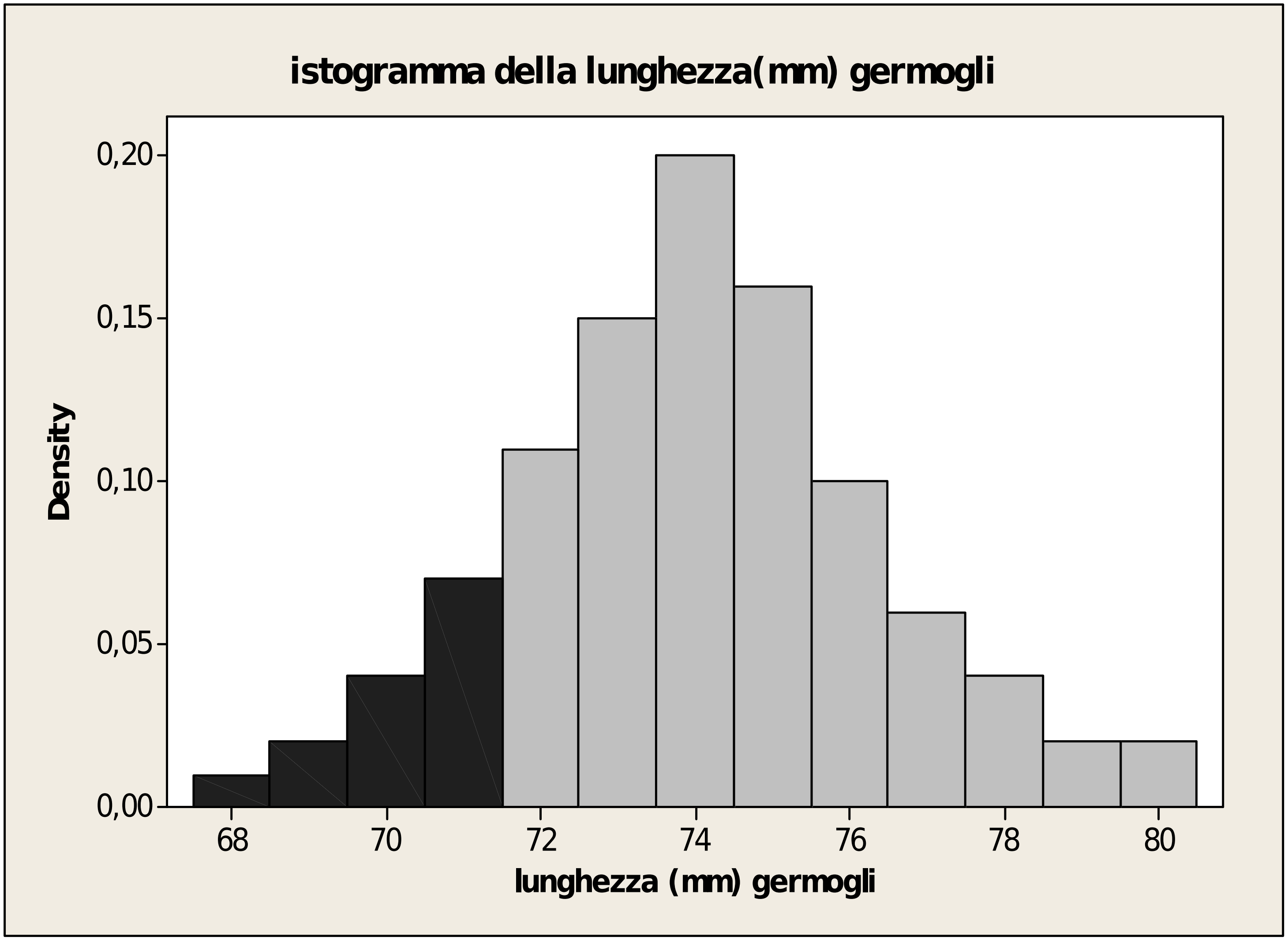


E' unimodale (classe modale 73.5-74.5 mm)

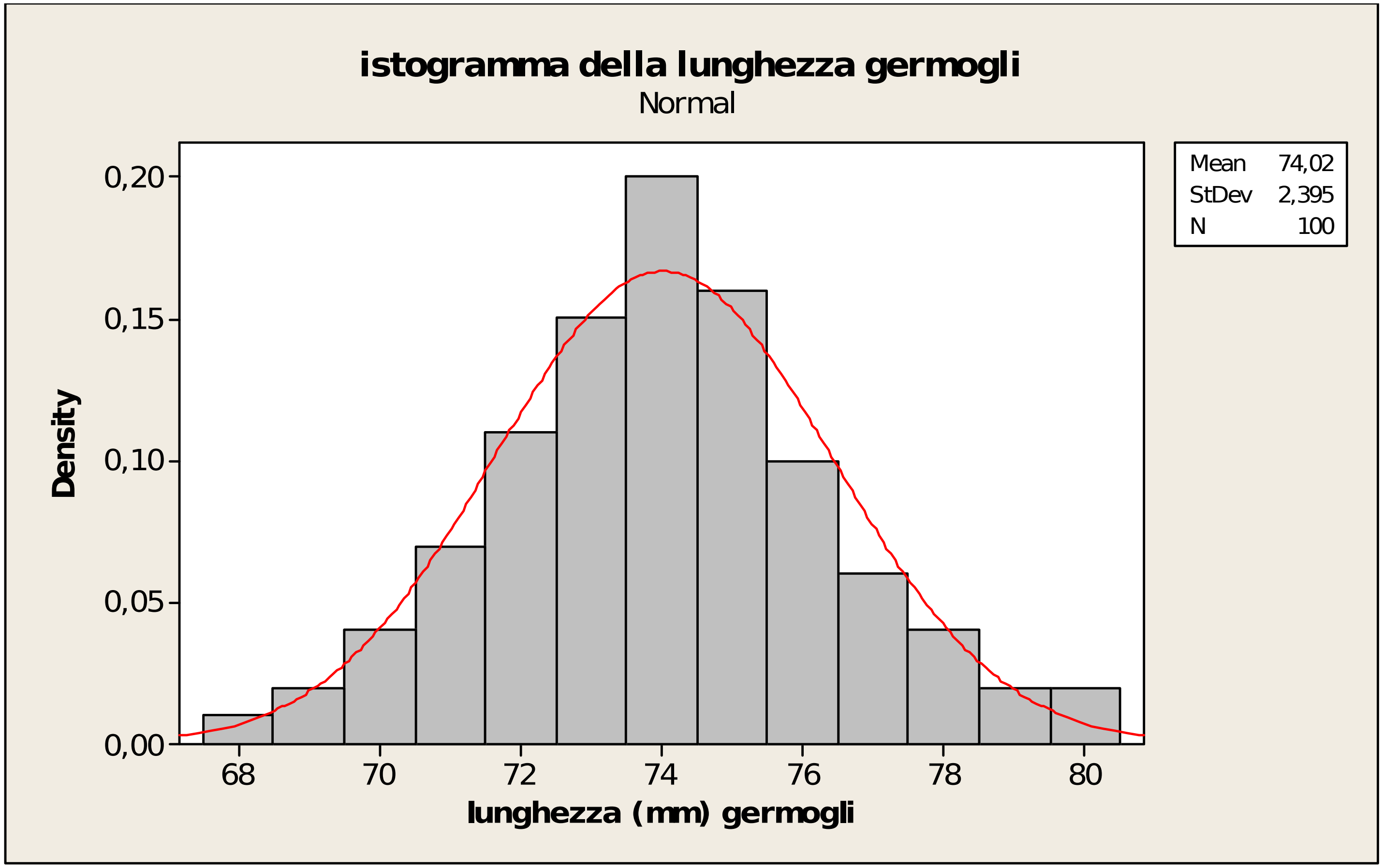
$\bar{X}$ = **74.02mm** e deviazione standard **s = 2.39mm**.



L'area del rettangolo sull'intervallo 70.5-71.5 è pari a 0.07, quindi nel campione di 100 germogli il 7% ha lunghezza tra 70.5 e 71.5.



L’area evidenziata rappresenta la frequenza relativa cumulata che fino al valore 71,5 mm è pari a 0.14, quindi il 14% dei germogli del campione ha lunghezza ≤ 71,5 mm.

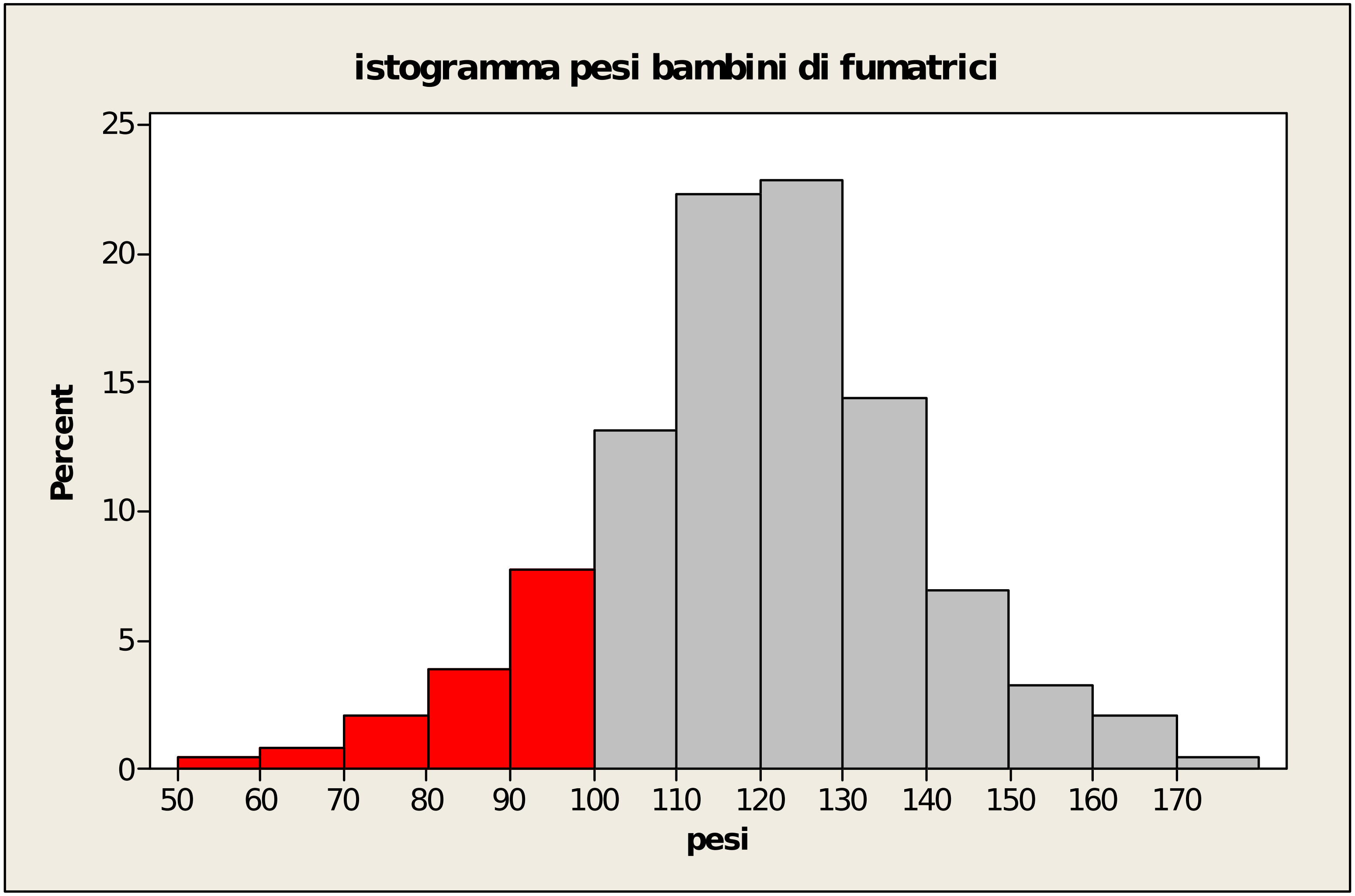


L'area sotto la curva di una distribuzione normale nell'intervallo [70.5 , 71.5] è pari a 0.075 ed è molto vicina alla percentuale (proporzione) osservata di germogli di lunghezza tra 70.5 e 71.5.

ESEMPIO: Pesi alla nascita

- Si considererà un sottoinsieme dei dati di un ampio studio condotto sulle donne in gravidanza tra il 1960 e il 1967 a San Francisco. Allo studio hanno partecipato 15000 famiglie con un livello di studio e di reddito medio-alto.
- Diverse misure del bambino venivano registrate alla nascita.
- Inizialmente considereremo 1236 maschi, nati tra il 1960 e il 1961, e che sono sopravissuti almeno 28 giorni. Per tali maschi si considereranno

Variabile	Descrizione
Peso alla nascita	Peso alla nascita in once (0.035 once=1gr)
Abitudine al fumo della madre	Indicatore dell'abitudine al fumo in gravidanza. Fumo si (1), no (0)



Ampiezza classi=10

Campione n=1236

La standardizzazione

Se x è un’ osservabile di media μ e deviazione standard σ, il valore standardizzato di x è

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

z è un’osservabile adimensionale di media 0 e varianza 1.

Variabile aleatoria normale standardizzata

- La standardizzazione trasforma la variabile aleatoria X che ha una distribuzione normale di media μ e dev st σ, in una v. a. Z che ha una distribuzione normale di media μ=0 e dev st =1

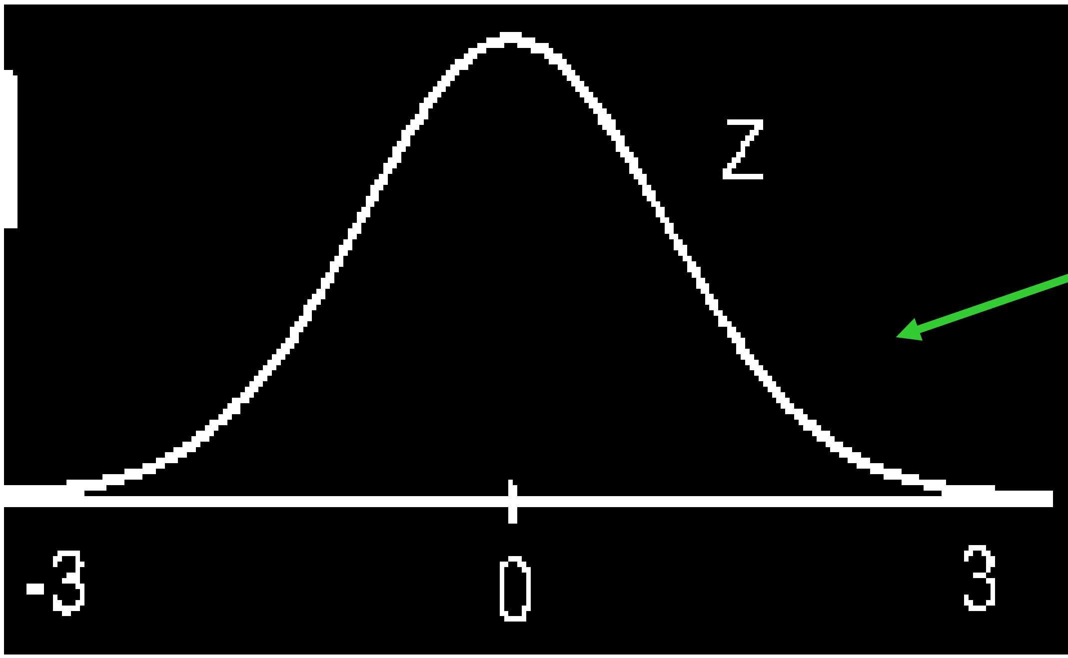
  (variabile aleatoria normale standard)

Es.

X ~ N(μ = 100, σ = 12).

Il valore z corrispondente a X = 128 è **z = (128-100)/12= <u>2.333.</u>**

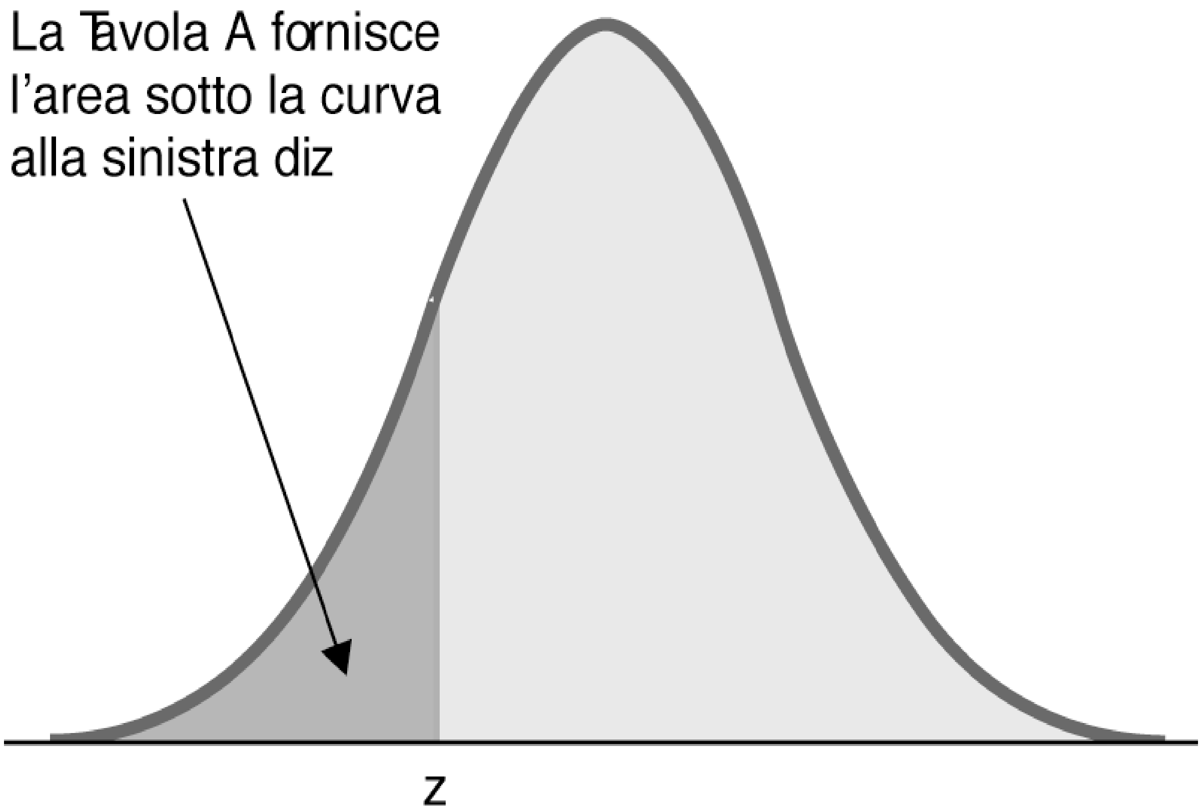
La distribuzione normale standard



$Z \sim N(0,1)$

La tavola della Normale standard

L'area sotto la curva alla sinistra di del valore z corrisponde alla probabilità che Z assuma un valore minore o uguale a z.



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706

Esempio

Probabilità che Z assuma un valore più piccolo di 0,67. E’ l’area sottesa dalla curva fino al valore 0,67

- Cerco la riga corrispondente al valore 0,6 e la colonna 0,07. Il valore trovato è 0,7486

$$P(Z < 0,67) = 0,7486$$

Osservazioni

- $P(Z>1,5) = 1 - P(Z \leq 1,5)$

  Dalla tabella $P(Z < 1,5)= 0,9332$, quindi

  $$P(Z < 1,5)= 1 - 0,9332 =0,0668$$

- Per la simmetria della curva

  $$P(Z < -2)= P(Z > 2)=1-P(Z \leq 2)$$

  Dalla tabella $P(Z < 2)=0,9772$, quindi

  $$P(Z < -2)= 1-0,9772=0,0228$$

Esempi

- $P(1 < Z < 2)$= area sottesa dalla curva nell'intervallo $[1 , 2]$

  $= P(Z < 2)- P(Z < 1) = 0,9772- 0,8413 =0, 1359$

- $P(-1 < Z < 2) = P(Z< 2) – P(Z < -1)$

  ma $P(Z < -1)= P (Z > 1) = 1 – P (Z < 1)$

  $= 1 – 0,8413=0,1587$

  quindi $P(-1 < Z < 2)= 0,9772-0,1587=0,8185$

Esempio

Sia X v.a. normale di media μ= 540 e deviazione standard σ=111. Calcolare $P(X < 420)$

Il valore x = 420 corrisponde a $z=(x-504)/111$

$= (420-504)/111=-0.76$

$P(Z< -0.76)= P(Z>0.76)=1-P(Z<0.76)= 1-0.7764=0.2236$

- Qual è la probabilità che X>646.?

Esercizio

I punteggi di QI ottenuti dagli studenti di prima media hanno distribuzione normale di media 100 e deviazione standard 14,2.

a) Qual è la probabilità che uno studente di prima media scelto a caso abbia un punteggio maggiore di 130?

b) E che abbia un punteggio compreso tra 80 e 115?

esercizio

- An exam is normally distributed with a mean of 200 points and a standard deviation of 25 points.

- (a) What percentage of the students score above 200 points?
- (b) What percentage of the students score below 175 points?
- (c) What percentage of the students score more than 250 points?

Soluzioni: a) 50% b) 16% c) 2%

Percentili delle variabili aleatorie normali

Il 75-percentile di una variabile aleatoria normale è il valore x tale che

$$P(X<x)= 0,75$$

Es. Z variabile aleatoria normale standard; per calcolare z tale che $P(Z < z)=0,75$ Dalla tabella z il valore più vicino è 0,67.
0,67 è il 75-percentile.

Esempio

Calcolare il 25-percentile di Z (v.a. normale standard)

- z : $P(Z < z) = 0{,}25$. Osservazione: $z < 0$. Per usare le tavole utilizzo la simmetria della distribuzione

  $$P(Z < z) = P(Z > -z) = 1 - P(Z \leq -z) = 0{,}25$$

  quindi

  $P(Z \leq -z) = 1 - 0{,}25 = 0{,}75$

  Dalla tabella $-z = 0{,}67$ quindi $-0{,}67$ è il 25-percentile.

Esempio

Calcolare il 95-percentile di una v.a. normale X di media 40 e deviazione standard 5.

- Dobbiamo determinare x: $P(X < x) = 0{,}95$.
- Usiamo la v.a. standardizzata $Z = (X-40)/5$
- $P(X < x) = P((X-40)/5 < (x-40)/5) = P(Z < (x-40)/5)$
- $P(Z < z) = 0{,}95$ si ha per $z = 1{,}645$ quindi

$$z = (x-40)/5 = 1{,}645$$

$$x = 5z + 40 = 48{,}225$$ (è il 95-percentile di X)

Esercizio

I punteggi di un test del QI hanno distribuzione normale con media 100 e deviazione standard 14,2. In quale insieme si trova il miglior 1% dei punteggi?

(Soluzione: è l'insieme dei punteggi ≥ 134, quindi il miglior 1% è costituito da chi totalizza almeno 134 punti)