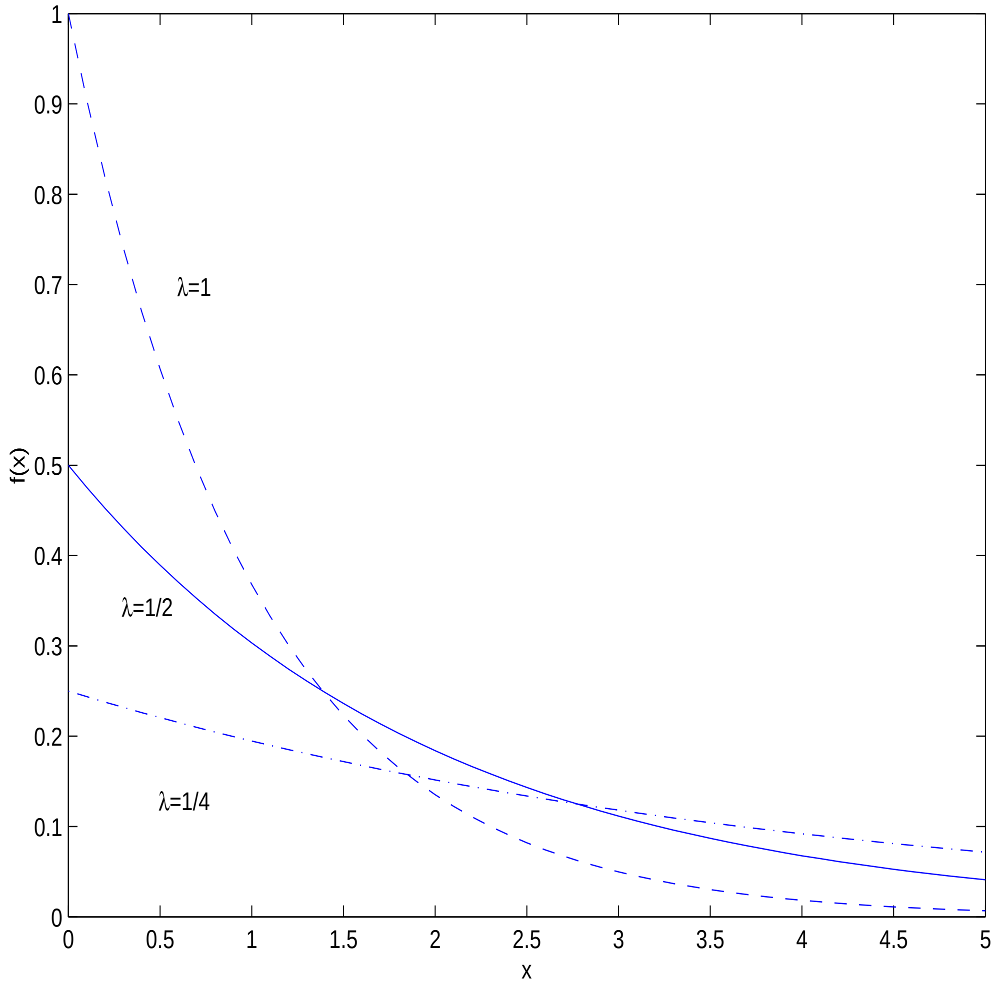
Distribuzione esponenziale

Funzione densità

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Funzione parametrica (λ)

Funzione di densità della distribuzione esponenziale



$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- f è una funzione di densità di distribuzione
- Media: $\mu \equiv E[X] = \frac{1}{\lambda}$ (integrazione per parti)
- Varianza: $Var[X] = \frac{1}{\lambda^2}$.
- La funzione distribuzione cumulativa soddisfa

$$P(X \leq b) = \int_0^b \lambda e^{-\lambda x} dx = 1 - e^{-\lambda b}.$$

Esempio. I tempi di durata di 500 batterie elettriche sono stati raggruppati con le seguenti frequenze

Ore	0-50	50-100	100-150	150-200	200-250	250-300	300-350	350-400
Freq.	208	112	75	40	30	18	11	6

Stimare la media e fare un istogramma. Suggerire una distribuzione che possa essere usata per modellare i dati, e determinare le frequenze per ogni intervallo-tempo sul tale modello di distribuzione (la frequenza per ogni intervallo si determina moltiplicando la probabilità in tale intervallo per il numero totale delle frequenze, cioè $N = 500$). Riportare tali valori sull'istogramma fatto in precedenza e commentare sulla bontà del modello.

Sol. Usando il punto medio x_i di ogni classe ed $N = 500$,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{8} f_i x_i = 95.$$

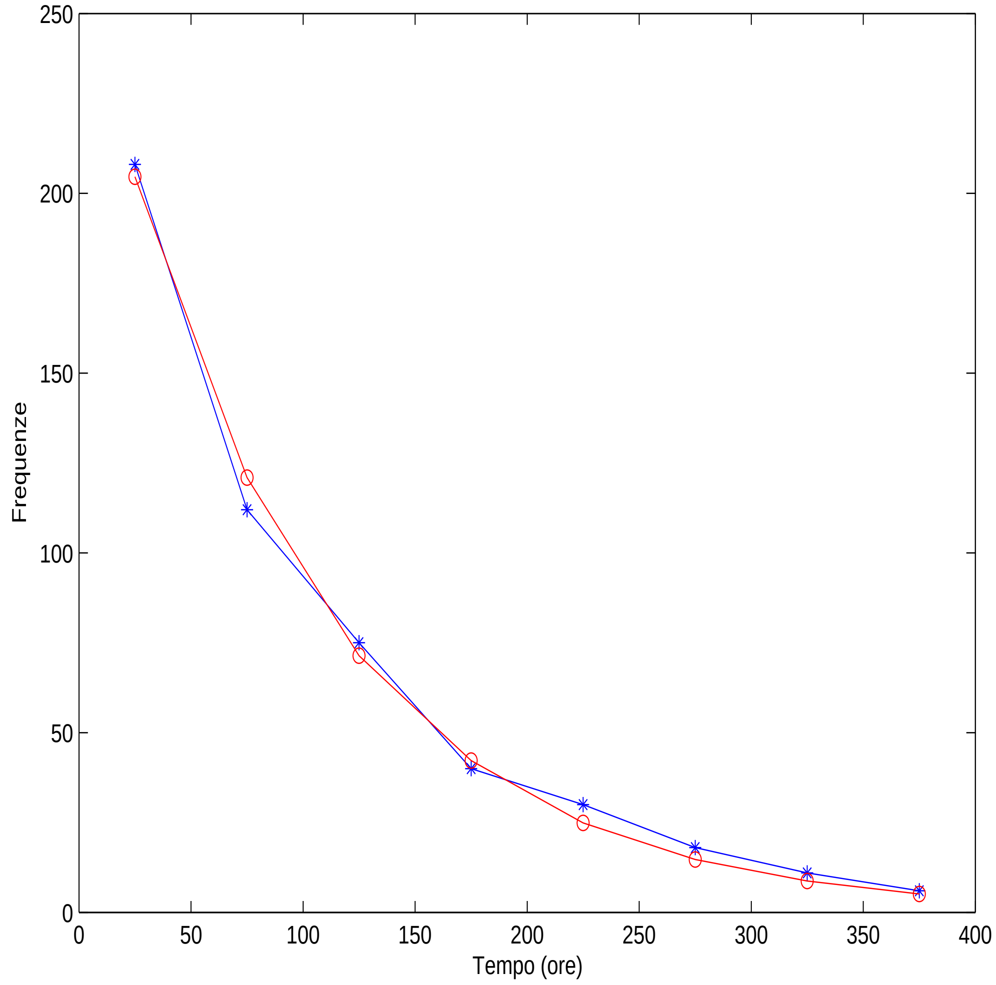
Se il modello di distribuzione è quello esponenziale, dev'essere $E(X) = \frac{1}{\lambda} = \bar{x}$

$$\lambda = \frac{1}{95}$$

Usando $F(b) = 1 - e^{-\lambda b}$:

$$P(x_i \le X \le x_{i+1}) = F(x_{i+1}) - F(x_i) \qquad \widetilde{F}_i = P(x_i \le X \le x_{i+1}) \cdot N$$

Ore	0-50	50-100	100-150	150-200	200-250	250-300	300-350	350-400
$\widetilde{F}$	204.6	120.9	71.4	42.2	24.9	14.7	8.7	5.1


250
200
150
100
50
0
Frequenze
0
50
100
150
200
250
300
350
400
Tempo (ore)

Distribuzione normale (o Gaussiana)

Si deriva dalla distribuzione binomiale per $N \to \infty$

X variabile casuale (media μ e varianza σ^2)

Funzione densità

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

$X \in \mathcal{N}(\mu, \sigma^2)$

$\star$ Dati μ e σ, se X è distribuita seguendo la funzione f sopra, allora

$$E[X] = \mu, \qquad Var[X] = \sigma^2$$

Valgono: $\quad f \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$

Distribuzione normale standardizzata

$$Z = \frac{X - \mu}{\sigma}$$

Funzione di densità *normalizzata*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad z \in (-\infty, \infty)$$

Z ha media zero e varianza uno $\qquad Z \in \mathcal{N}(0, 1)$

Probabilità cumulata

$$P(Z \leq b) = \Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$P(Z \leq b) =$ Area sotto la curva

- 65% della probabilità in $[\mu - \sigma, \mu + \sigma]$,
- 95% in $[\mu - 2\sigma, \mu + 2\sigma]$

Esempio: Supponiamo $X \in \mathcal{N}(5, 4)$. Determinare

(a) $P(4 \leq X \leq 6)$ (b) $P(X \geq 9)$.

Sol. Definiamo Z in $\mathcal{N}(0, 1)$, $Z = (X - 5)/2$.

$4 \leq X \leq 6 \Rightarrow -0.5 \leq Z \leq 0.5$ Si sfrutta

$$P(-0.5 \leq Z \leq 0.5) = P(Z \leq 0.5) - P(Z \leq -0.5)$$

Usando le tabelle: $P(Z \leq 0.5) = .6915$
$P(Z \leq -0.5) = P(Z \geq 0.5) = 1 - P(Z \leq 0.5) = 0.3085.$

$\Rightarrow P(-0.5 \leq Z \leq 0.5) = P(Z \leq 0.5) - P(Z \leq -0.5) = 0.3830.$

La probabilità $P(X \geq 9)$ si trasforma in $P(Z \geq 2)$ e si ha
$P(Z \geq 2) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228.$

Esempio: Le uova di gallina hanno peso medio di 60 gr. con deviazione standard di 15 gr., e si può pensare ad una loro distribuzione normale. Uova di peso inferiore a 45 gr. sono classificate come *piccole*. Il resto viene diviso tra standard e grandi, ed è sperabile che queste si verifichino con uguale frequenza. A quale peso bisognerebbe distinguere tra standard e grandi? (arrotondare al grammo)

Sol. Determiniamo x_0 tale che $P(45 \leq X \leq x_0) = P(X \geq x_0)$. Si ha
$P(X \geq x_0) = 1 - P(X \leq x_0)$ e
$P(45 \leq X \leq x_0) = P(X \leq x_0) - P(X \leq 45)$, quindi deve valere

$$2P(X \leq x_0) - P(X \leq 45) = 1 \quad \Leftrightarrow \quad P(X \leq x_0) = \frac{1 + P(X \leq 45)}{2}.$$

In forma standard, $Z = (X - 60)/15$, per cui $P(X \leq 45)$ si trasforma in
$P(Z \leq -1) = P(Z \geq 1) = 1 - P(Z \leq 1) = 0.1587$ (dalla Tabella).
Quindi, $P(Z \leq z_0) = (1 + 0.1587)/2 = 0.57935$ e sulla Tabella $z_0 = 0.205$.
Nell'unità di misura dei dati, $x_0 = 15z_0 + 60 = 63.075 \approx 63$ gr.

Esempio: In una specie di pesci adulti, maschi e femmine sono distinguibili dalle dimensioni medie: misurate in cm., le femmine hanno $\mu = 37.5$ e $\sigma = 3.8$, mentre i maschi hanno $\mu = 34.5$ e $\sigma = 3.2$. Qual'è la lunghezza minima del 5% delle femmine con dimensioni maggiori? Quale percentuale di maschi avranno le stesse dimensioni del 30% delle femmine di dimensioni maggiori?

Sol. In una distribuzione normale, nella variabile standardizzata, 95% dei valori sono contenuti per $z \leq 1.645$ (si veda la Tabella).

Il valore soglia è quindi $z_0 = 1.645$, al quale corrisponde $x_0 = z_0\sigma + \mu = 43.7510$. La percentuale di maschi è 5.94%.

Addattamento della distr. normale

In un ospedale viene valutata la dipendenza tra il peso (in grammi) dei neonati e la caratteristica che la neo mamma non abbia mai fumato. I dati risultanti sono riportati nella seguente tabella.

Peso	1-	501-	1001-	1501-	2001-	2501-	3001-	3501-	4001-	4501-	5001-	Tot.
Freq.	4	20	25	73	236	1089	2530	1927	551	101	7	6563

Studiare l'adattamento della distribuzione normale ai dati.

Sol. Considerando come osservazioni i punti centrali di ogni classe, si ha che $\bar{x} = 3353.30$, $s^2 = 3.2782 \cdot 10^5$ e $s = 572.55$.

In tabella:

- dati (valore destro dell'intervallo)

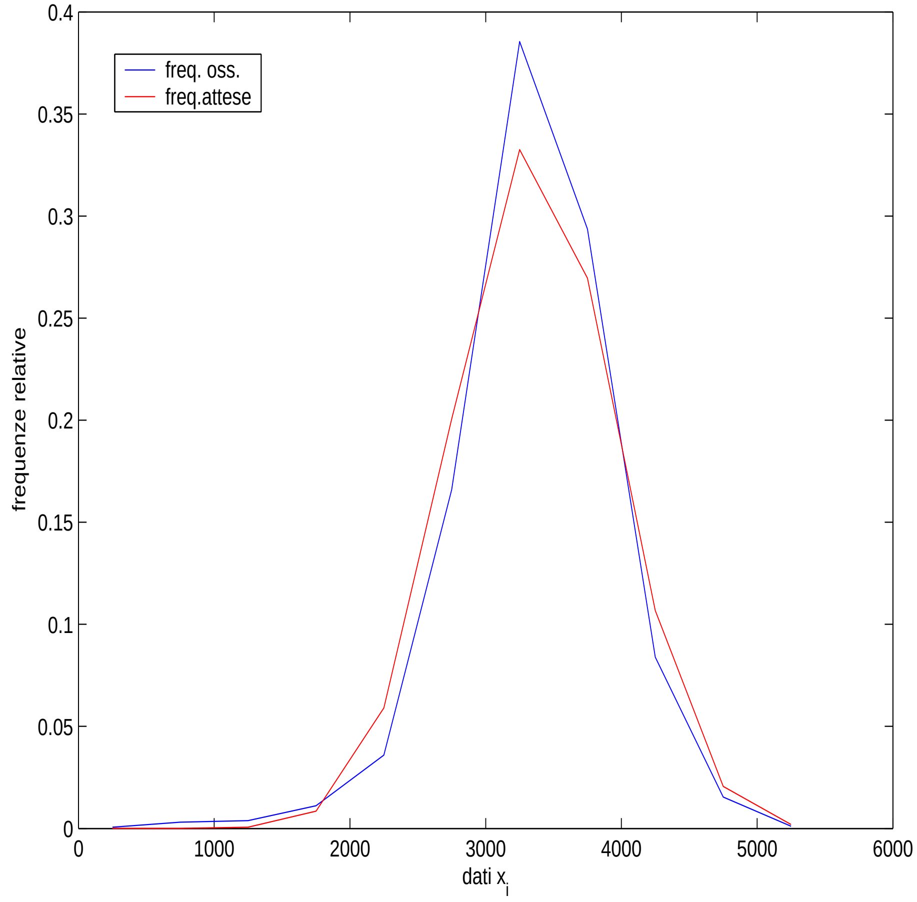
- frequenze relative

- valori standardizzati $z = (x - \bar{x})/s$

- probabilità della distribuzione normale associate ai valori di z

- Frequenze attese (teoriche) relative, $\widetilde{F}(z_i) = P(z_i) - P(z_{i-1})$

x	f_i/N	z	$P(z)$	$\widetilde{F}(z)$
500	0.00060948	-4.9835	3.1225e-007	3.1225e-007
1000	0.0030474	-4.1102	1.9766e-005	1.9453e-005
1500	0.0038092	-3.2369	0.00060414	0.00058437
2000	0.011123	-2.3636	0.0090483	0.0084441
2500	0.035959	-1.4904	0.068065	0.059017
3000	0.16593	-0.61707	0.26859	0.20053
3500	0.38549	0.25621	0.60111	0.33251
4000	0.29362	1.1295	0.87065	0.26955
4500	0.083956	2.0028	0.9774	0.10674
5000	0.015389	2.8761	0.99799	0.020587
5500	0.0010666	3.7493	0.99991	0.0019247


freq. oss.
freq.attese
frequenze relative
dati x_i
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0
1000
2000
3000
4000
5000
6000

Valutazione della normalità: Q-Q plot

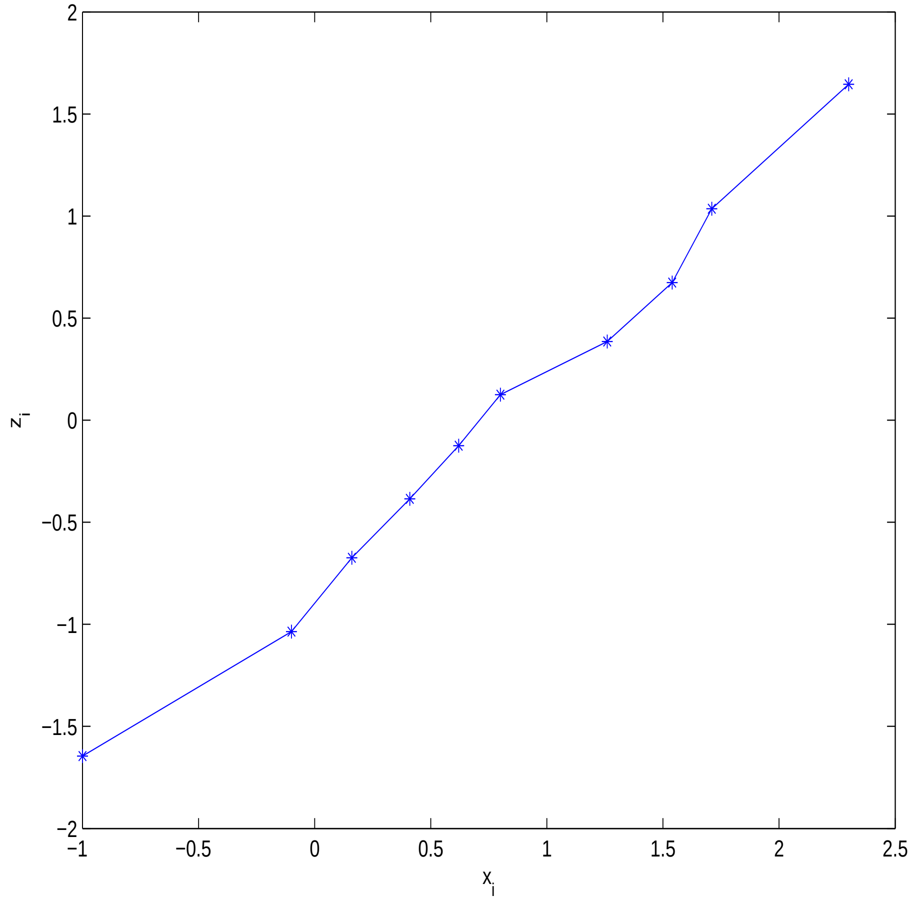
$x_i \; i = 1, \ldots, N$ ordinati crescenti

z_i tali che $P(z \leq z_i) = \frac{i - \frac{1}{2}}{N}$

Q-Q plot: grafico di (x_i, z_i) (quantili)

Osserv. x_i	Prob $(i-\frac{1}{2})/n$	quantili (std. normal) z_i
-1.00	0.05	-1.645
-0.10	0.15	-1.036
0.16	0.25	-0.674
0.41	0.35	-0.385
0.62	0.45	-0.125
0.80	0.55	0.125
1.26	0.65	0.385
1.54	0.75	0.674
1.71	0.85	1.036
2.30	0.95	1.645

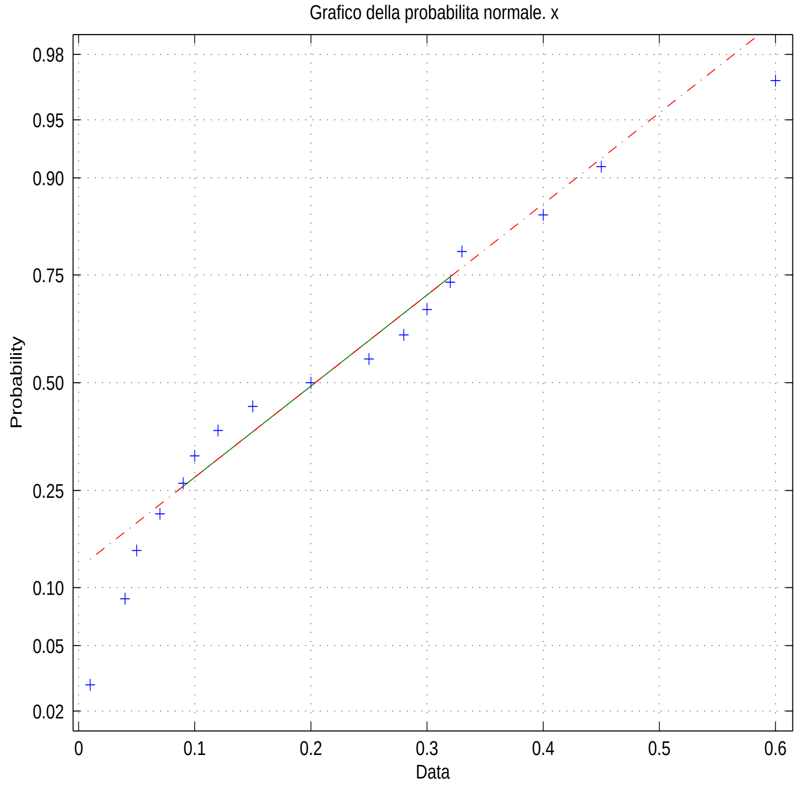
es. $P(z \leq z_7) = (7 - 0.5)/10 = 0.65$ e si ottiene $z_7 = 0.385$.

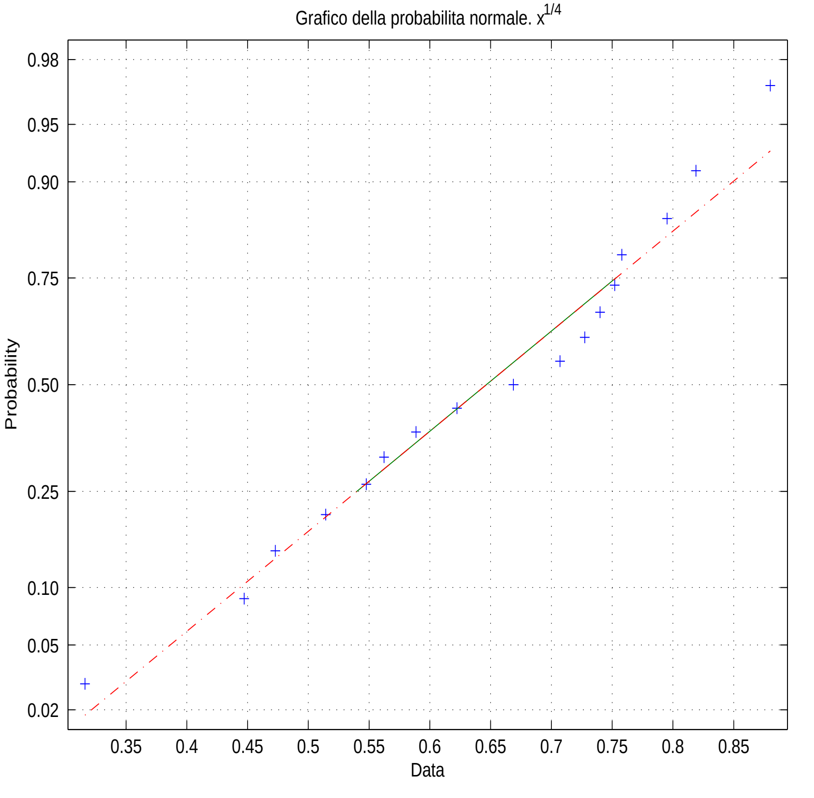


2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−1
−0.5
0
0.5
1
1.5
2
2.5
x_i
z_i

Trasformazioni verso la normalità

$$x^2, x^3, x^4, \sqrt{x}, \sqrt[4]{x}, \log x, \ln x, e^x, \ldots$$

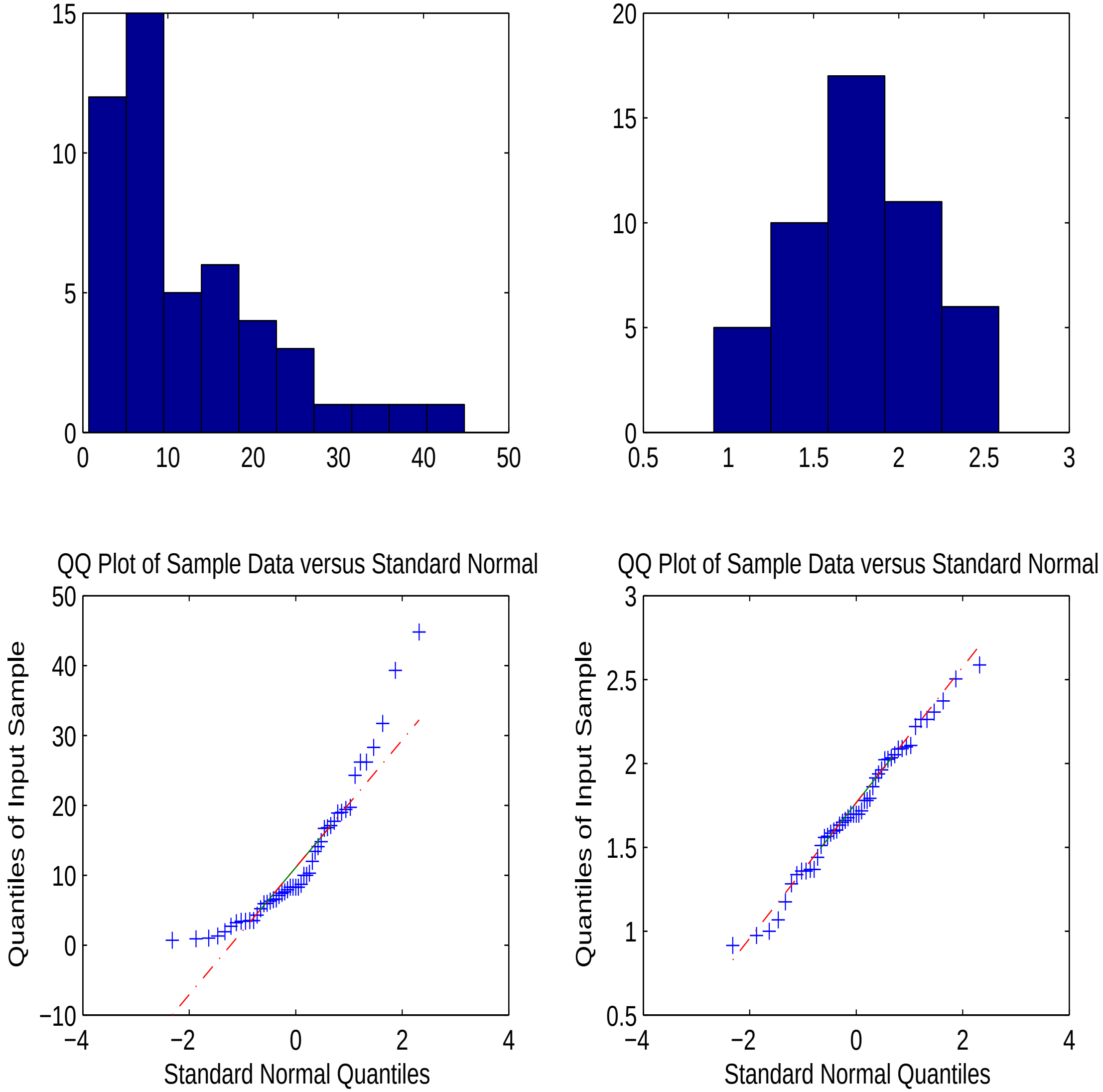




Trasformazioni verso la normalità

Volume di legno da segheria (stimato dal campionamento del numero di alberi in zone casuali della foresta)

$$\ell = \begin{bmatrix} 39.3 & 3.5 & 6.0 & 2.7 & 7.4 & 3.5 & 19.4 \\ 19.7 & 1.0 & 8.7 & 14.8 & 8.3 & 17.1 & 26.2 \\ 6.6 & 8.3 & 19.0 & 10.3 & 7.6 & 18.9 & 6.3 \\ 10.0 & 16.8 & 24.3 & 5.2 & 44.8 & 14.1 & 3.4 \\ 28.3 & 3.4 & 0.9 & 1.3 & 0.7 & 17.7 & 8.3 \\ 8.3 & 1.9 & 16.7 & 26.2 & 10.0 & 6.5 & 7.1 \\ 7.9 & 3.2 & 5.9 & 13.4 & 12.0 & 4.3 & 31.7 \end{bmatrix}$$



Volume $\qquad\qquad\qquad$ $\sqrt[4]{\text{Volume}}$