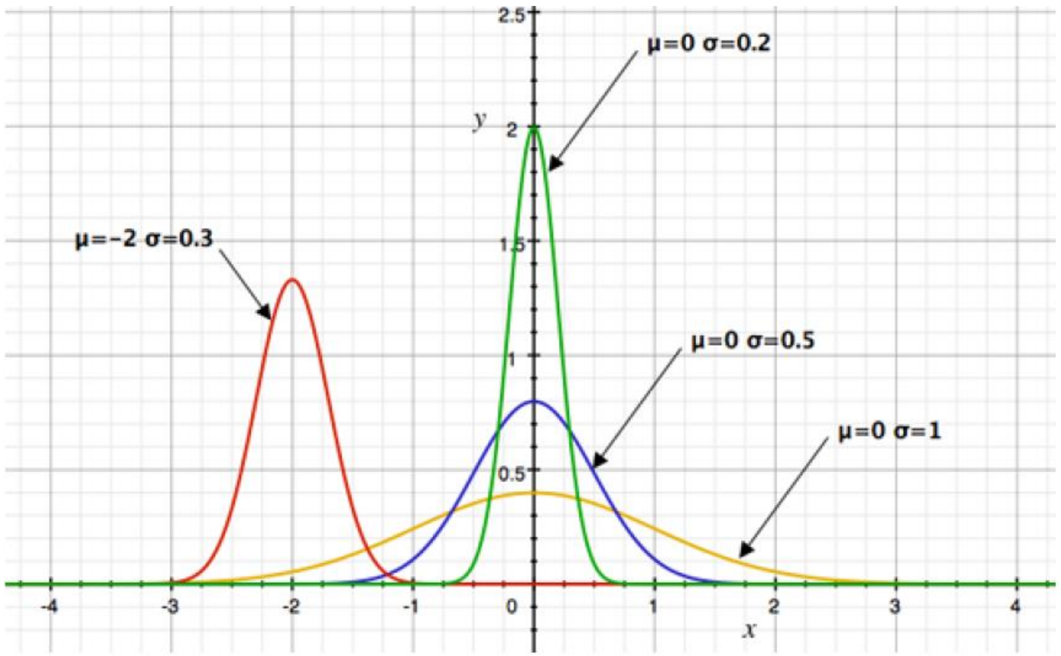
STATISTICA E ANALISI DEI DATI

Distribuzione Normale

Distribuzione Normale

- La funzione di **distribuzione normale**, detta anche di Gauss o gaussiana, riveste estrema importanza nel calcolo delle probabilità e nella statistica
  - Questa funzione di distribuzione costituisce una distribuzione limite alla quale tendono varie altre funzioni di distribuzioni sotto opportune ipotesi
- La distribuzione normale è spesso considerata un buon modello per:
  - Variabili fisiche: peso, altezza, temperatura, voltaggio, livello di inquinamento
  - Analizzare gli errori di misurazione
  - Per descrivere il reddito familiare o voti degli studenti.



Distribuzione Normale

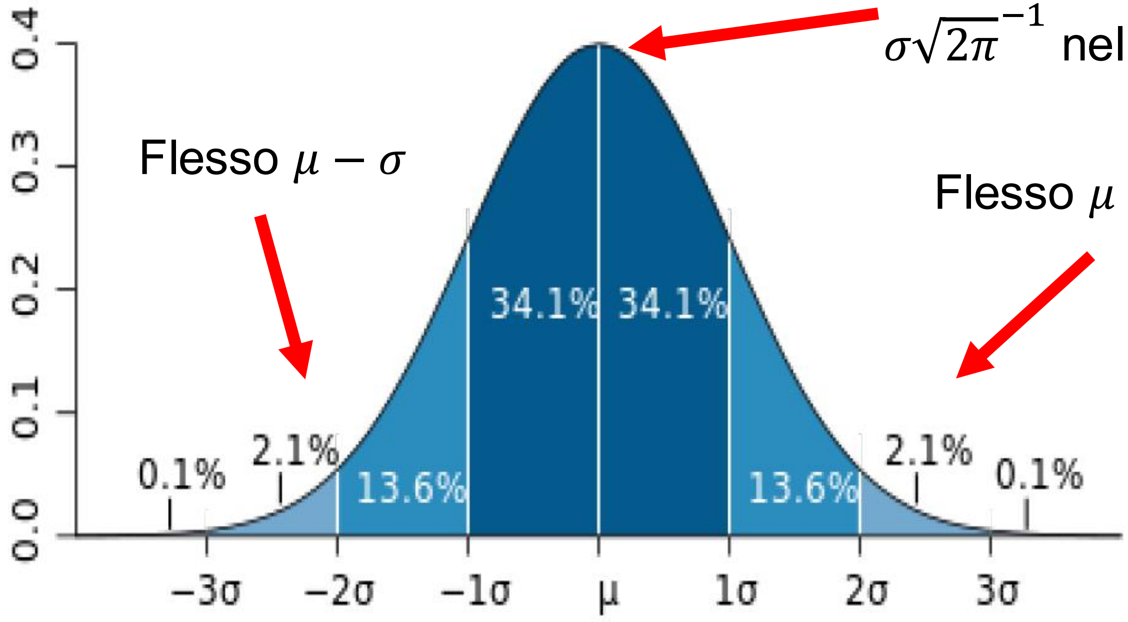
- Una variabile aleatoria X con:

Densità di probabilità: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad con\ x \in R, \mu \in R, \sigma > 0$

La densità presenta il massimo a $\sigma\sqrt{2\pi}^{-1}$ nel punto di ascissa $x = \mu$

Flesso $\mu - \sigma$

Flesso $\mu + \sigma$



Distribuzione Normale

- Una variabile aleatoria X con:

  Densità di probabilità: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad con\ x \in R, \mu \in R, \sigma > 0$

  si dice avere distribuzione normale con parametri μ e σ, dove:

  μ = Valore Atteso $\qquad$ σ = Deviazione Standard

- <u>Notazione</u>:
  - $X \sim N(\mu, \sigma)$ variabile aleatoria con distribuzione normale

- Per una variabile aleatoria normale si ha:

  Valore atteso: $E(X) = \mu$ $\qquad$ **Varianza**: $Var(X) = \sigma^2$

Simmetria Rispetto a μ

- Una variabile aleatoria X con:

Densità di probabilità: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad con\ x \in R, \mu \in R, \sigma > 0$

- Una variabile aleatoria X con:

$$f_X(\mu - x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(\mu\ -\ x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(-x)^2}{2\sigma^2}\right\} = \boxed{\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}}$$

$$f_X(\mu + x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(\mu\ +\ x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(-x)^2}{2\sigma^2}\right\} = \boxed{\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}}$$

- Questa uguaglianza significa che **la densità normale è simmetrica rispetto alla retta verticale** $t = \mu$
- Se prendiamo un punto che dista x a sinistra di μ (cioè $\mu - x$) e un punto che dista x a destra di μ (cioè $\mu + x$), la densità **assume lo stesso valore**

Distribuzione Normale

- Per il calcolo della densità normale si utilizza la funzione:

```
dnorm(x, mean = mu, sd = sigma)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **mean** e **sd** sono il valore medio e la deviazione standard della densità normale

Distribuzione Normale

- Per il calcolo della densità normale si utilizza la funzione:

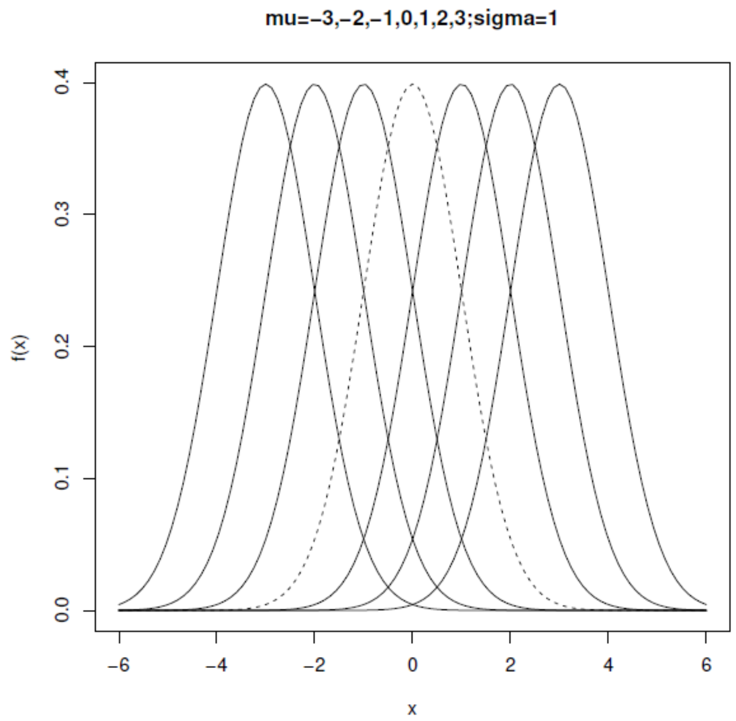
```
dnorm(x, mean = mu, sd = sigma)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **mean** e **sd** sono il valore medio e la deviazione standard della densità normale

- **Esempio**: Valutiamo la densità di $X \sim N(\mu, 1)$ con $\mu = -3, -2, -1{,}0{,}1{,}2{,}3$

```
>curve(dnorm(x,mean=-3,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+main="mu=-3,-2,-1,0,1,2,3;sigma=1")
>curve(dnorm(x,mean=-2,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=-1,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=0,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE,lty=2)
>curve(dnorm(x,mean=1,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=2,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=3,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
```



Distribuzione Normale

- Per il calcolo della <u>densità normale</u> si utilizza la funzione:

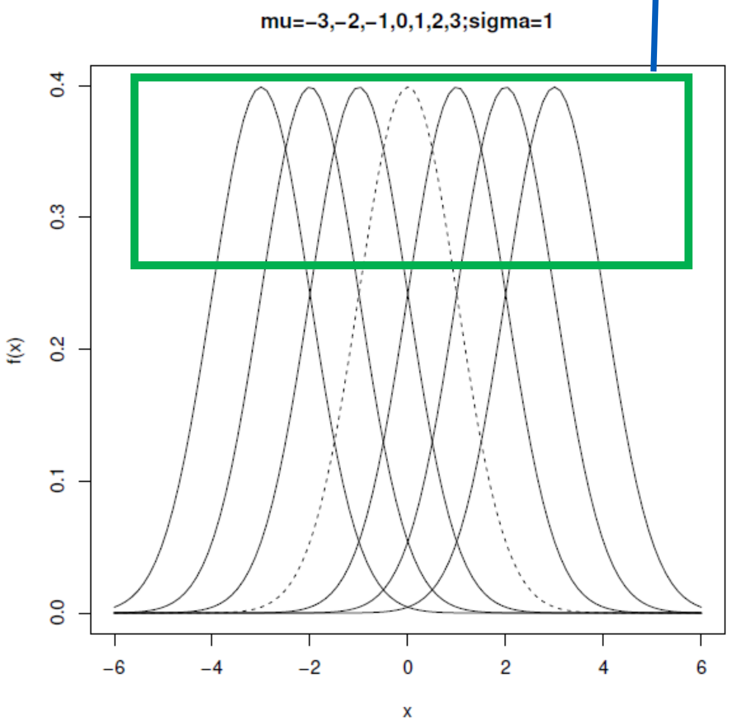
dove

  - **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
  - **mean** e **sd** sono il valore medio e la deviazione standard della densità normale

- **Esempio**: Valutiamo la densità di $X \sim N(\mu, 1)$ con $\mu = -3, -2, -1, 0, 1, 2, 3$

```
>curve(dnorm(x,mean=-3,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+main="mu=-3,-2,-1,0,1,2,3;sigma=1")
>curve(dnorm(x,mean=-2,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=-1,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=0,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE,lty=2)
>curve(dnorm(x,mean=1,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=2,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
>curve(dnorm(x,mean=3,sd=1),from=-6, to=6,xlab="x",ylab="f(x)",
+add=TRUE)
```

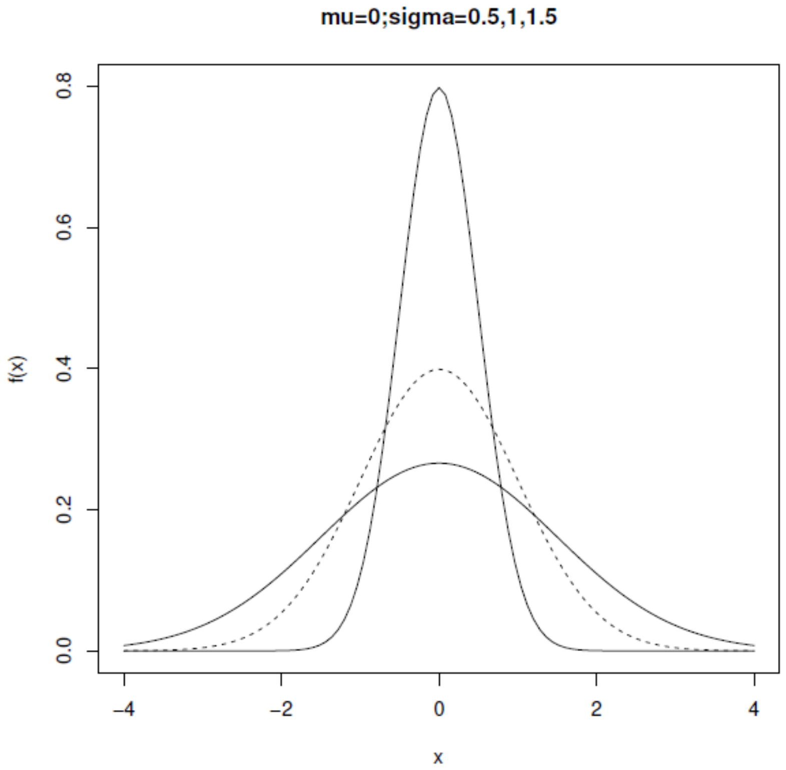
Variando il parametro μ la curva viene traslata lungo l'asse delle ascisse



Distribuzione Normale

- **Esempio**: Valutiamo la densità di $X \sim N(0, \sigma)$ con $\sigma = 0.5, 1, 1.5$

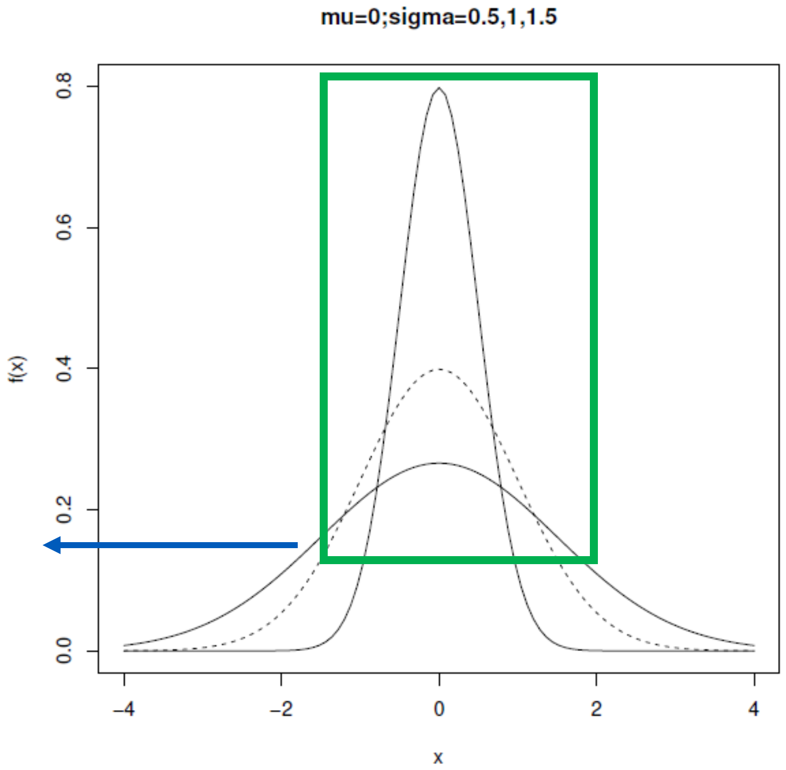
```
>curve(dnorm(x,mean=0,sd=0.5),from=-4, to=4,xlab="x",
+ylab="f(x)",main="mu=0;sigma=0.5,1,1.5")
>curve(dnorm(x,mean=0,sd=1),from=-4, to=4,xlab="x",
+ylab="f(x)",add=TRUE,lty=2)
>curve(dnorm(x,mean=0,sd=1.5),from=-4, to=4,xlab="x",
+ylab="f(x)",add=TRUE)
```



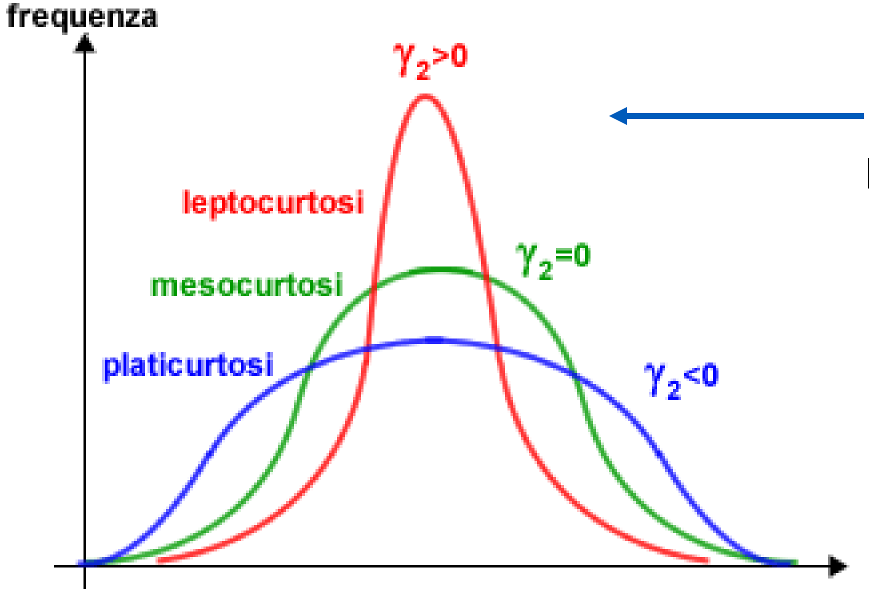
Distribuzione Normale

- **Esempio**: Valutiamo la densità di $X \sim N(0, \sigma)$ con $\sigma = 0.5, 1, 1.5$

```
>curve(dnorm(x,mean=0,sd=0.5),from=-4, to=4,xlab="x",
+ylab="f(x)",main="mu=0;sigma=0.5,1,1.5")
>curve(dnorm(x,mean=0,sd=1),from=-4, to=4,xlab="x",
+ylab="f(x)",add=TRUE,lty=2)
>curve(dnorm(x,mean=0,sd=1.5),from=-4, to=4,xlab="x",
+ylab="f(x)",add=TRUE)
```



Al decrescere di σ la curva si allunga verso l'alto e si restringe nei lati



Distribuzione Normale

- Per calcolare la funzione di distribuzione normale invece utilizziamo la funzione:

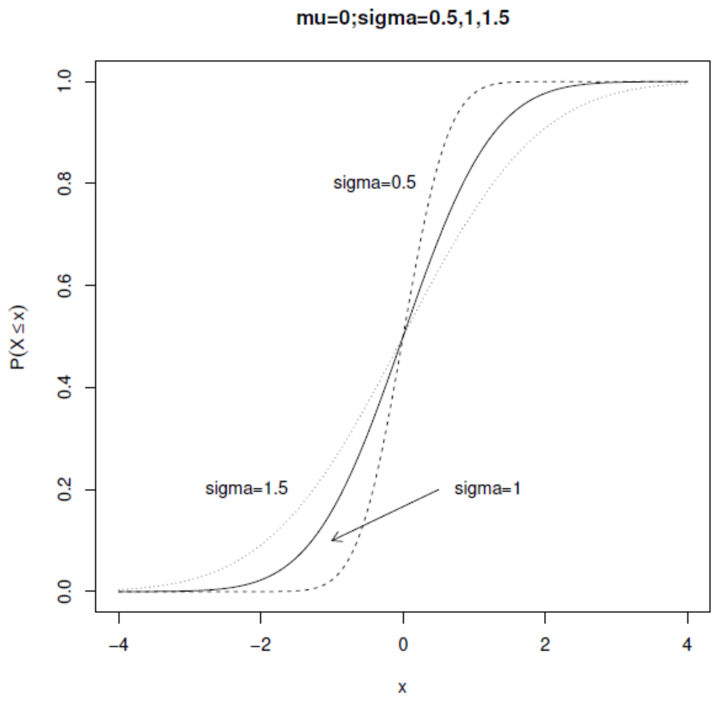
```
pnorm(x, mean = mu, sd = sigma, lower.tail = TRUE)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **mean** e **sd** sono il valore medio e la deviazione standard della densità normale
- **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

- **Esempio**: Valutiamo la densità di $X \sim N(0, \sigma)$ con $\sigma = 0.5, 1, 1.5$

```
>curve(pnorm(x,mean=0,sd=0.5),from=-4, to=4,xlab="x",
+ylab=expression(P(X<=x)),main="mu=0;sigma=0.5,1,1.5",lty=2)
>text(-0.4,0.8,"sigma=0.5")
>curve(pnorm(x,mean=0,sd=1),add=TRUE)
>arrows(-1,0.1,0.5,0.2,code=1,length = 0.10)
>text(0.8,0.2,"sigma=1")
>curve(pnorm(x,mean=0,sd=1.5),add=TRUE,,lty=3)
>text(-2.2,0.2,"sigma=1.5")
```



Distribuzione Normale Standard

- Una variabile aleatoria **normale standard**, solitamente denotata con Z, può essere ottenuta da una variabile aleatoria normale non standard $X \sim N(\mu, \sigma)$ standardizzando, ossia **sottraendo il valore medio e dividendo** per la deviazione standard:

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

Usando questa trasformazione, qualsiasi variabile aleatoria normale può essere ottenuta da una variabile aleatoria normale standard Z

- La funzione di distribuzione di una variabile aleatoria $X \sim N(\mu, \sigma)$ è quindi:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\, dy = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

Dove:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{y^2}{2}\right\} dy$$

è la funzione di distribuzione di una variabile aleatoria $Z \sim N(0,1)$

Distribuzione Normale Standard

- Una variabile aleatoria **normale standard**, solitamente denotata con Z, può essere ottenuta da una variabile aleatoria normale non standard $X \sim N(\mu, \sigma)$ standardizzando, ossia **sottraendo il valore medio e dividendo** per la deviazione standard:

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

<u>**Usando questa trasformazione, qualsiasi variabile aleatoria normale può essere ottenuta da una variabile aleatoria normale standard Z**</u>

- La funzione di distribuzione di una variabile aleatoria $X \sim N(\mu, \sigma)$ è quindi:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\, dy = \boxed{\Phi\left(\frac{x-\mu}{\sigma}\right)},$$

L'area sottesa nel grafico
Ricavabile dalla tavola della
Variabile normale standard

Dove:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{y^2}{2}\right\} dy$$

è la funzione di distribuzione di una variabile aleatoria $Z \sim N(0, 1)$

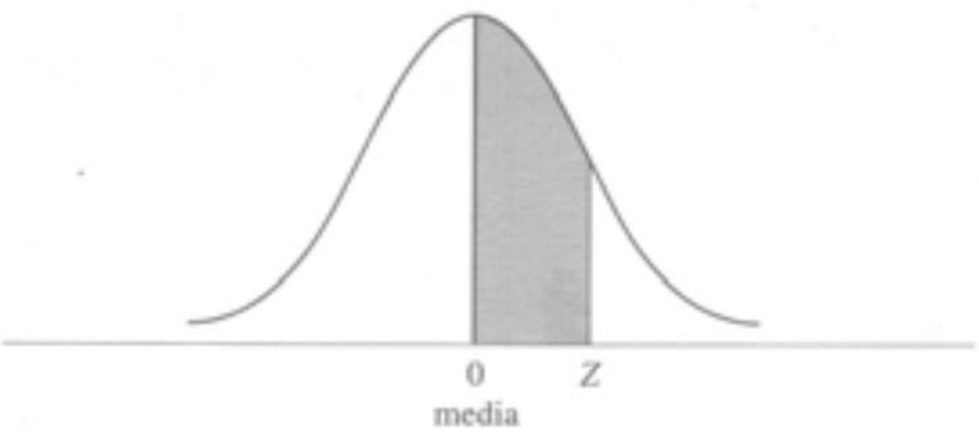
Tavola delle Aree

- La funzione $\Phi(z)$ non dipende da alcun parametro, quindi può essere facilmente tabulata
- E’ cioè possibile costruire una tavola che riporta le aree sottese alla curva in corrispondenza di diversi valori di z
- In questa tavola viene riportata l’area compresa tra 0 e un punto z collocato a destra dell’origine
- La tavola ci fornisce il valore del **percentile** sulla base **dell’area sottesa fino alla percentuale** richiesta

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	0.99995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									

Tavola delle Aree

- La tavola ci fornisce il valore del percentile sulla base dell'area sottesa fino alla percentuale richiesta
- **Esempio**: L'area sottesa dalla $N(0,1)$
  - nell'intervallo [0, 1.65] è pari a: 0,9505
  - nell'intervallo [0, 1.34] è pari a: 0,9099



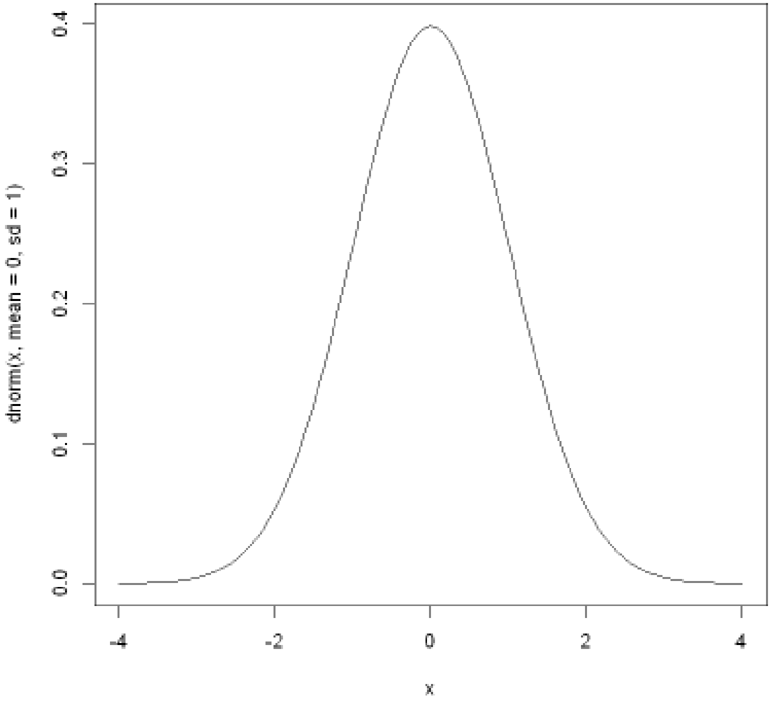
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	0.99995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	0.99997									
5.0	0.9999997									
6.0	0.999999999									

Distribuzione Normale Standard

- Una variabile aleatoria X con valore medio $\mu = 0$ e varianza $\sigma = 1$ si dice variabile aleatoria normale standard $X \sim N(0,1)$:

  Densità di probabilità: $f_X(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad con\; x \in R, \mu \in R, \sigma > 0$

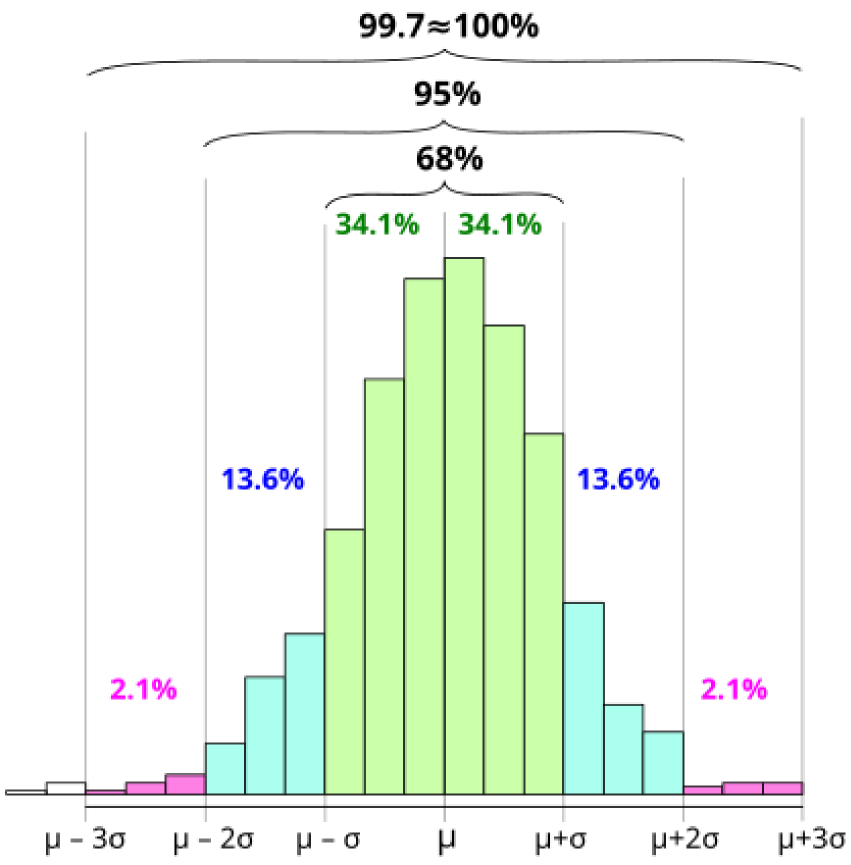
  Funzione di distribuzione: $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x)^2}{2}} dx$



- Proprietà della densità di $X \sim N(0,1)$:
  - Il massimo in $x = 0$ e $\frac{1}{\sqrt{2\pi}} = 0.3989$.
  - La curva è simmetrica rispetto all'asse delle y
  - L'area sottesa alla curva tra -3 e 3 è $P(-3 < X < 3) = 0.9973$
- La variabile aleatoria normale standard Z può essere ottenuta da una variabile aleatoria normale non standard sottraendo il valore medio e dividendo per la deviazione standard: $Z = \frac{X-\mu}{\sigma}$ cioè $X = \mu + \sigma$

Regole del 1σ,2σ,3σ

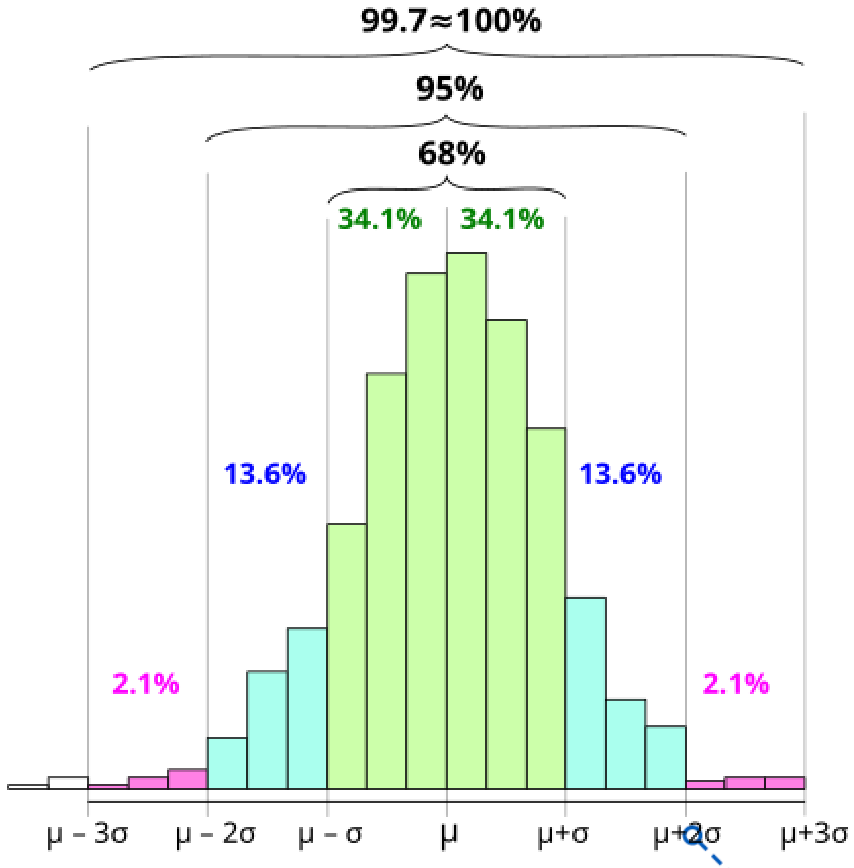
- La regola del 1σ, 2σ, 3σ (o anche nota come 68 - 95 - 99,7) è una **regola empirica** utilizzata per ricordare la percentuale di valori che si trovano all'interno di un intervallo attorno alla media in una distribuzione normale con un'ampiezza di due, quattro e sei deviazioni standard, rispettivamente;
  - Più precisamente, il 68,27%, il 95,45% e il 99,73% dei valori si trovano rispettivamente all'interno di una, due e tre deviazioni standard della media
  - Il resto dei valori sono considerati **outlier**



- Sulla base di questo dobbiamo considerare le seguenti proprietà:
- L'area compresa tra l'asse delle ascisse e le rette
  - $x = \mu - \sigma$ e $x = \mu + \sigma$ è pari al 68.3%: $P(\mu - \sigma < X < \mu + \sigma) = 0.683\%$
  - $x = \mu - 2\sigma$ e $x = \mu + 2\sigma$ è pari al 95.4%: $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.954\%$
  - $x = \mu - 3\sigma$ e $x = \mu + 3\sigma$ è pari al 99.7%: $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997\%$

Alcuni Esempi

- **Esempio 1**:
  - In una fabbrica si producono bulloni con un diametro medio di 10 mm e una deviazione standard di 0.2 mm
    - Se i bulloni rientrano nell'intervallo $10 \pm 2\sigma$ (cioè [9.6,10.4]), si **accetta** il prodotto
    - Bulloni fuori da 3σ (cioè <9.4 o >10.6) sono considerati **difettosi**
- **Esempio 2**:
  - Una compagnia assicurativa analizza i reclami per danni medi di 10.000€ con una deviazione standard di 2.000€
  - La maggior parte dei reclami (95%) è compresa tra $\mu \pm 2\sigma$ (cioè [6.000€,14.000€]).
  - Reclami superiori a $\mu + 3\sigma$ (>16.000€) potrebbero essere investigati come potenziali **frodi**

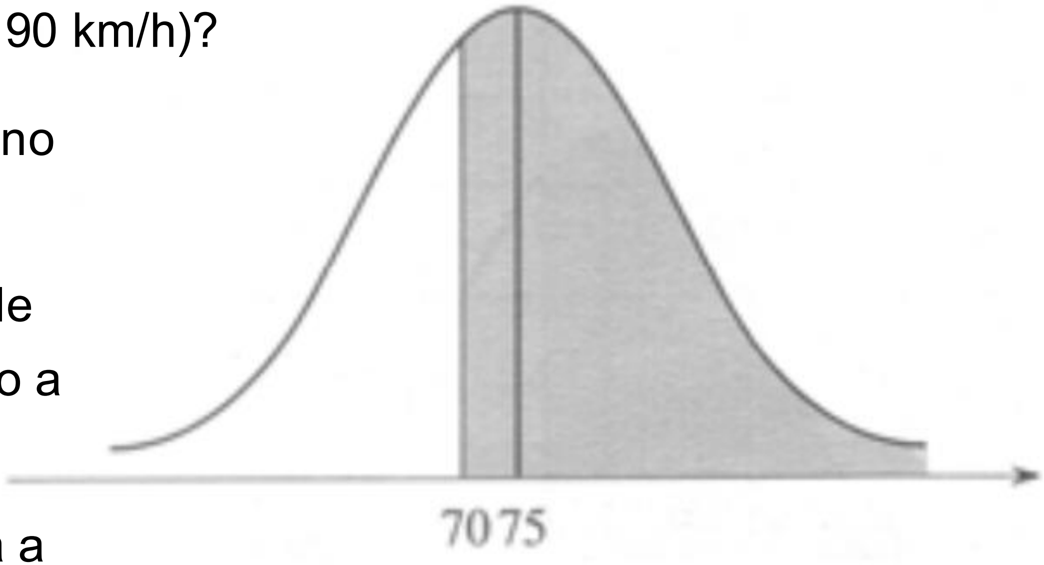


Esempio

- La velocità delle auto rilevata dall'autovelox sulla tangenziale est si distribuisce normalmente con $\mu = 75$ km/h e $\sigma = 8\ km/h$
  - Che percentuale di auto superano il limite di velocità di 70 km/h?
  - A quanti automobilisti su 100 viene ritirata la patente (oltre 90 km/h)?

Esempio

- La velocità delle auto rilevata dall'autovelox sulla tangenziale est si distribuisce normalmente con $\mu = 75$ km/h e $\sigma = 8\ km/h$
  - Che percentuale di auto superano il limite di velocità di 70 km/h?
  - A quanti automobilisti su 100 viene ritirata la patente (oltre 90 km/h)?



- Disegniamo le aree che rispondono alle domande che ci vengono poste:
  - Supponiamo di avere a disposizione la tavola della Normale che riporta l'area compresa nell'intervallo $[0, z]$, e pensiamo a come conviene procedere
  - Per rispondere alla prima domanda, considerato che l'area a destra della media è ben nota (=0,5), dobbiamo solo calcolare l'area compresa tra 70 e 75

Esempio

- Il primo passo è calcolare il valore z standardizzato corrispondente a 70 km/h:

$$z = \frac{(70 - 75)}{8} = -0.625$$

- Osserviamo che l'area compresa nell'intervallo $[-0.625, 0]$, per la simmetria della Normale, è uguale a quella dell'intervallo $[0, 0.625]$
- Cerchiamo 0.625 sulla tavola:
  - vediamo che non troviamo il valore esatto, dobbiamo quindi interpolare i due valori più prossimi:

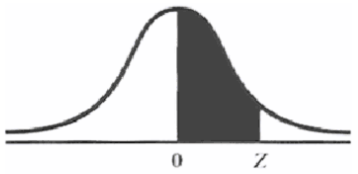
$$\frac{(0.2324 + 0.2357)}{2} = 0.2341$$

- A questo punto dobbiamo solo ricordarci di aggiungere l'area a destra del valore medio:

$$0.2341 + 0.5 = 0.7341$$

- **Risposta 1**: il 73,41% degli autoveicoli in transito superano il limite di velocità

Tavola della distribuzione Normale Standardizzata

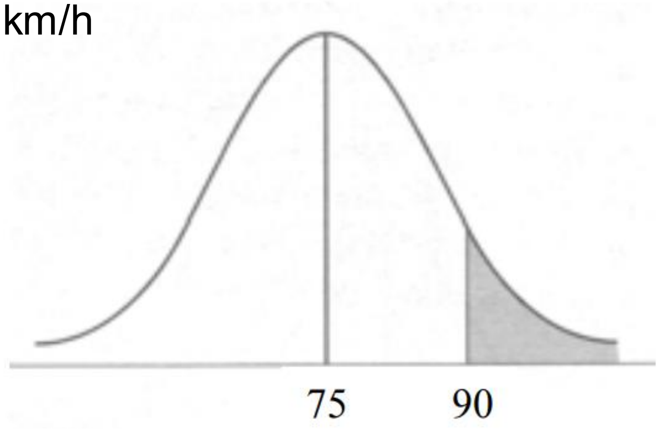


Area sottesa alla curva di densità normale standardizzata calcolata tra **0** e **Z**

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	$\boxed{0,2324}$	$\boxed{0,2357}$	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Esempio

- Per rispondere alla seconda domanda dobbiamo calcolare l'area a destra di 90 km/h



- Calcoliamo il valore z standardizzato corrispondente a 90 km/h:

$$z = \frac{(90 - 75)}{8} = 1.88$$

- Osserviamo che l'area a destra di z=1.88 è uguale a 0,5 meno l'area dell'intervallo [0, 1.88], fornita dalla nostra tavola
- Cerchiamo 1.88 sulla tavola ed è: 0.4699
- L'area che ci interessa è quindi uguale a: 0.5 – 0.4699 = 0.0301
- **Risposta 2**: viene ritirata la patente al 3% degli automobilisti in transito

Tavola della distribuzione Normale Standardizzata

	Area sottesa alla curva di densità normale standardizzata calcolata tra **0** e **Z**									
Z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	$\boxed{0,4699}$	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Quantili

- Per calcolare i quantili (percentili) della distribuzione normale

```
qnorm(z, mean = mu, sd = sigma)
```

dove

  - **z** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile z · 100–esimo, cioè il vettore delle probabilità;
  - **mean** e **sd** sono il valore medio e la deviazione standard della densità normale

- **Esempio**: Calcoliamo i quartili Q_0, Q_1, Q_2, Q_3, Q_4 della variabile aleatoria normale standard:

```
>z<-c(0,0.25,0.5,0.75,1)
>qnorm(z, mean = 0, sd = 1)
[1]    -Inf -0.6744898  0.0000000  0.6744898      Inf
```

Generazione Randomica di Numeri

- È possibile simulare la variabile aleatoria normale generando una sequenza di numeri pseudocasuali mediante la funzione

```
>rnorm(N, mean = mu, sd = sigma)
```

dove

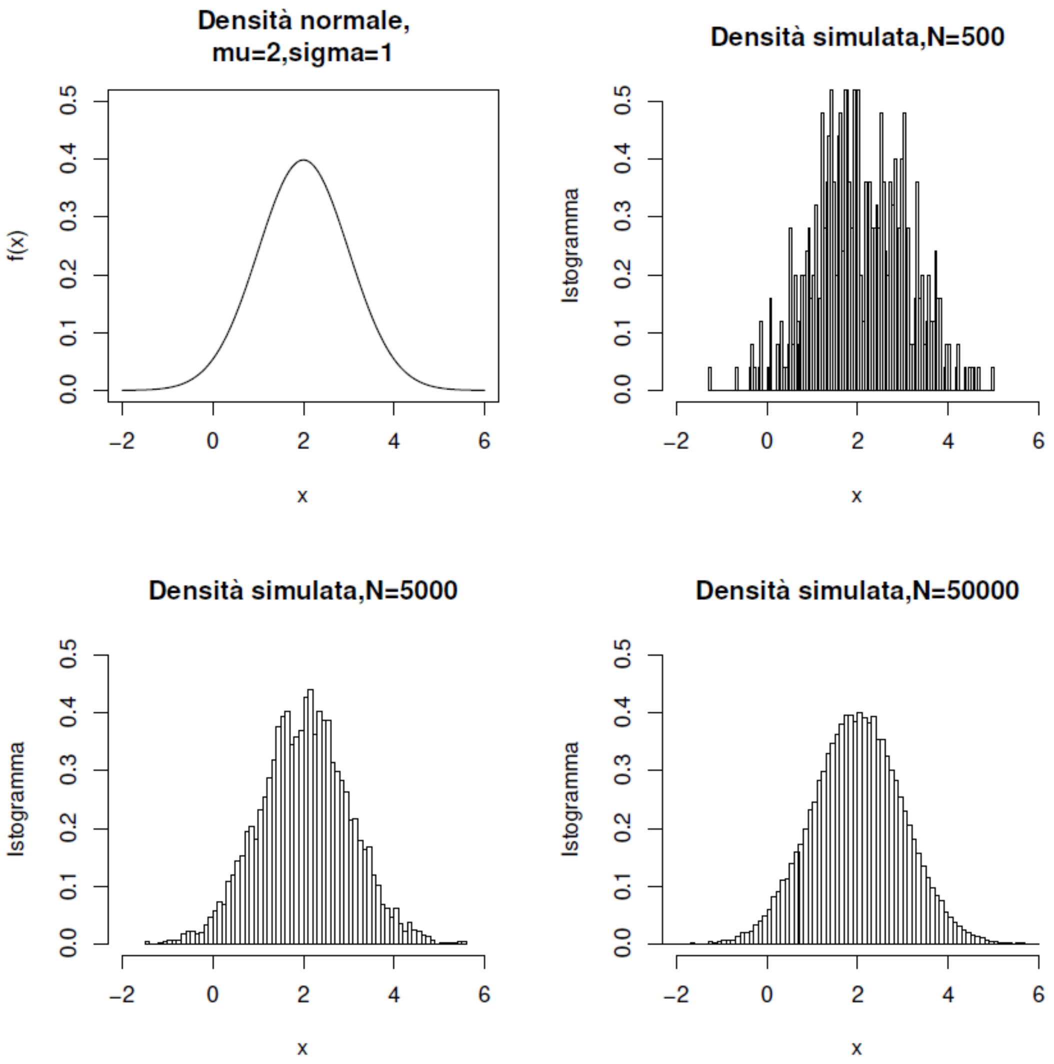
- **N è** lunghezza della sequenza da generare;
- **mean** e **sd** sono il valore medio e la deviazione standard della densità normale

Generazione Randomica di Numeri

- **Esempio**:

```
>par(mfrow=c(2,2))
>curve(dnorm(x,mean=2,sd=1),from=-2, to=6,xlab="x",ylab="f(x)",
+ylim=c(0,0.5),main="Densita’ normale,mu=2,sigma=1")
>
>sim1<-rnorm(500,mean=2,sd=1)
>hist(sim1,freq=F,xlim=c(-2,6),ylim=c(0,0.5),breaks=100,xlab="x"
+ylab="Istogramma",main="Densita’ simulata,N=500")
>
>sim2<-rnorm(5000,mean=2,sd=1)
>hist(sim2,freq=F,xlim=c(-2,6),ylim=c(0,0.5),breaks=100,xlab="x"
+ylab="Istogramma",main="Densita’ simulata,N=5000")
>
>sim3<-rnorm(50000,mean=2,sd=1)
>hist(sim3,freq=F,xlim=c(-2,6),ylim=c(0,0.5),breaks=100,xlab="x"
+ylab="Istogramma",main="Densita’ simulata,N=50000")
```

- Notiamo che al crescere di N la densità simulata si avvicina alla densità teorica normale



Media Campionaria

- Siano $X_1, X_2, \dots, X_n$ sono variabili aleatorie normali indipendenti con $X_i \sim N(\mu_i, \sigma_i)$ per $i = 1,2, \dots, n$
- Siano $a_1, a_2, \dots, a_n$ numeri reali, allora la variabile aleatoria:

$$Y = X_1 a_1 + X_2 a_2 + \cdots + X_n a_n$$

  - ha distribuzione normale con valore medio e varianza:

$$E(Y) = \mu_1 a_1 + \mu_2 a_2 + \cdots + \mu_n a_n$$
$$Var(Y) = \sigma_1^2 a_1^2 + \sigma_2^2 a_2^2 + \cdots + \sigma_n^2 a_n^2$$

- Se $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$ si ha che:

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \longleftarrow \text{Media Campionaria}$$

che ha una distribuzione normale con valore medio e varianza: $E(\overline{X_n}) = \mu$ $\quad Var(\overline{X_n}) = \frac{\sigma^2}{n}$

- Questo risultato mostra che se la popolazione è descritta da una variabile aleatoria $X \sim N(\mu, \sigma)$ e da essa si estrae un campione $X_1, X_2, \dots, X_n$, allora la media campionaria $\overline{X_n}$ è anche normale

STATISTICA E ANALISI DEI DATI

Approssimazione di Distribuzioni note alla Normale

Approssimazione della binomiale con la normale

- Il **teorema di De Moivre-Laplace** è un caso particolare del **teorema centrale del limite**, ed è una delle prime dimostrazioni formali che lega la distribuzione binomiale alla distribuzione normale.
- Esso afferma che, per un numero di prove n **sufficientemente grande**, la distribuzione binomiale $B(n, p)$ può essere approssimata da una distribuzione normale.
- **Teorema di De Moivre-Laplace**
  - Siano $X_1, X_2, \dots$ una successione di variabili aleatorie indipendenti di Bernoulli con $(0 < p < 1)$
  - Sia $Y_n = X_1 + X_2 + \cdots + X_n$ per ogni $x \in \mathrm{R}$ si ha:

$$\lim_{n\to\infty}\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-y^2}{2}} dx$$

Ossia

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \to Z$$

converge in distribuzione alla variabile aleatoria Z normale standard

Approssimazione della binomiale con la normale

- Il **teorema di De Moivre-Laplace** è un caso particolare del **teorema centrale del limite**, ed è una delle prime dimostrazioni formali che lega la distribuzione binomiale alla distribuzione normale.
- Esso afferma che, per un numero di prove n **sufficientemente grande**, la distribuzione binomiale $B(n, p)$ può essere approssimata da una distribuzione normale.
- **Teorema di De Moivre-Laplace**
  - Siano $X_1, X_2, \ldots$ una successione di variabili aleatorie indipendenti di Bernoulli con $(0 < p < 1)$
  - Sia $Y_n = X_1 + X_2 + \cdots + X_n$ per ogni $x \in \mathrm{R}$ si ha:

$$\lim_{n\to\infty}\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-y^2}{2}} dx$$

Ossia

Valore Atteso binomiale

$$\frac{Y_n - \boxed{np}}{\boxed{\sqrt{np(1-p)}}} \to Z$$

Varianza binomiale

converge in distribuzione alla variabile aleatoria Z normale standard

Approssimazione della binomiale con la normale

- Ricordiamo che se $X_1, X_2, \dots$ sono variabili aleatorie indipendenti di Bernoulli di parametro p, si ha che:

$$Y_n = X_1 + X_2 + \cdots + X_n$$

con Y_n variabile aleatoria <u>binomiale</u> di valore medio np e varianza $np(1-p)$

- Inoltre, si ha che sottraendo a Y_n la media np e dividendo per la varianza $np(1-p)$ si ottiene una **variabile aleatoria standard** la cui funzione di distribuzione è per n grande approssimativamente normale standard
- L'approssimazione della binomiale alla normale è quindi:

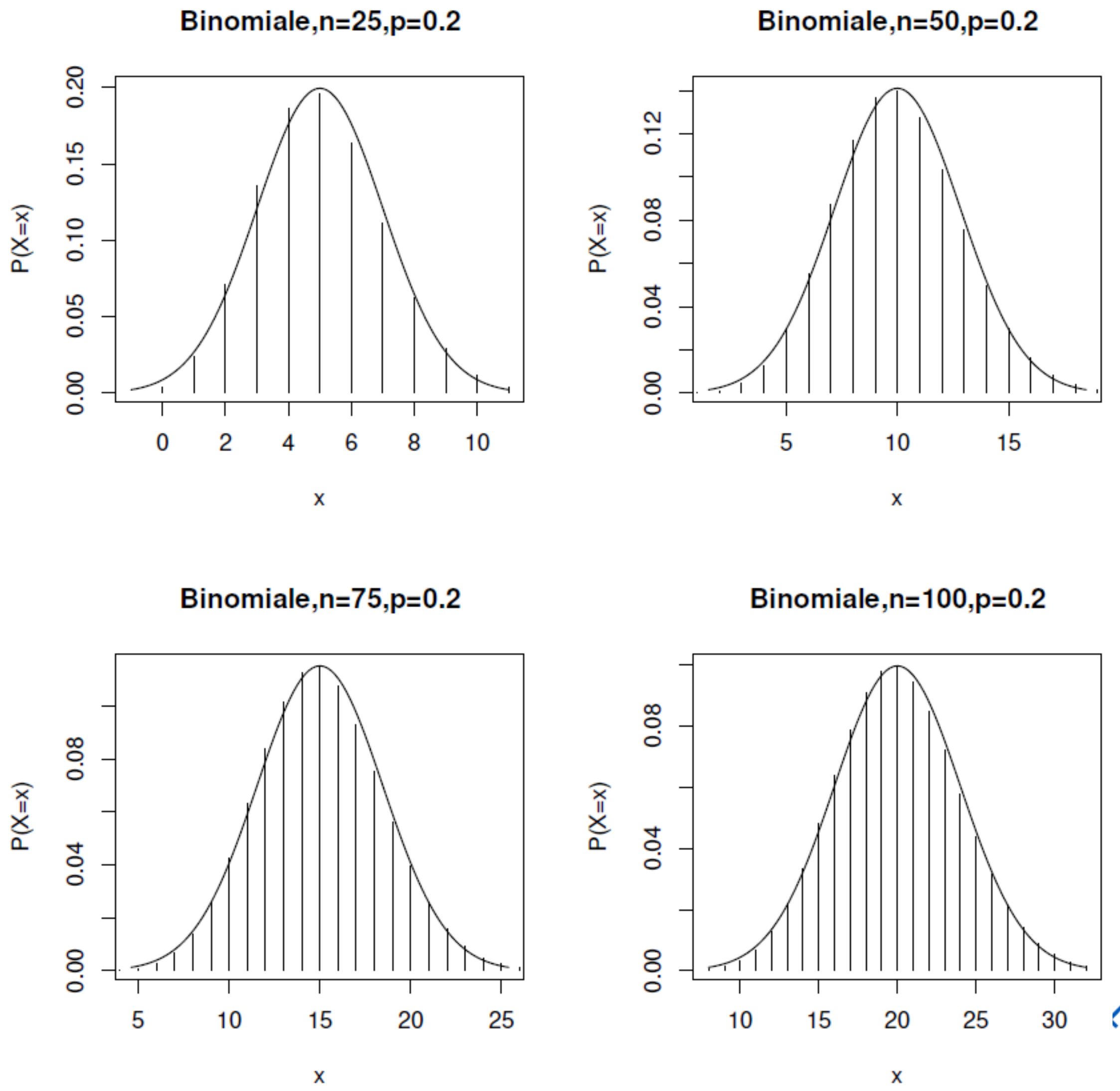
$$Da\ \frac{Y_n - np}{\sqrt{np(1-p)}} \rightarrow Z \quad si\ ha\ che\ \ Y_n \cong np + Z\sqrt{np(1-p)}$$

al variare di n con p fissato

Approssimazione della binomiale con la normale

- Confrontiamo la densità normale di valore medio np e varianza $np(1-p)$ e la funzione di probabilità binomiale per $n = 20,50,75,100$ e p fissato a $p = 0.2$

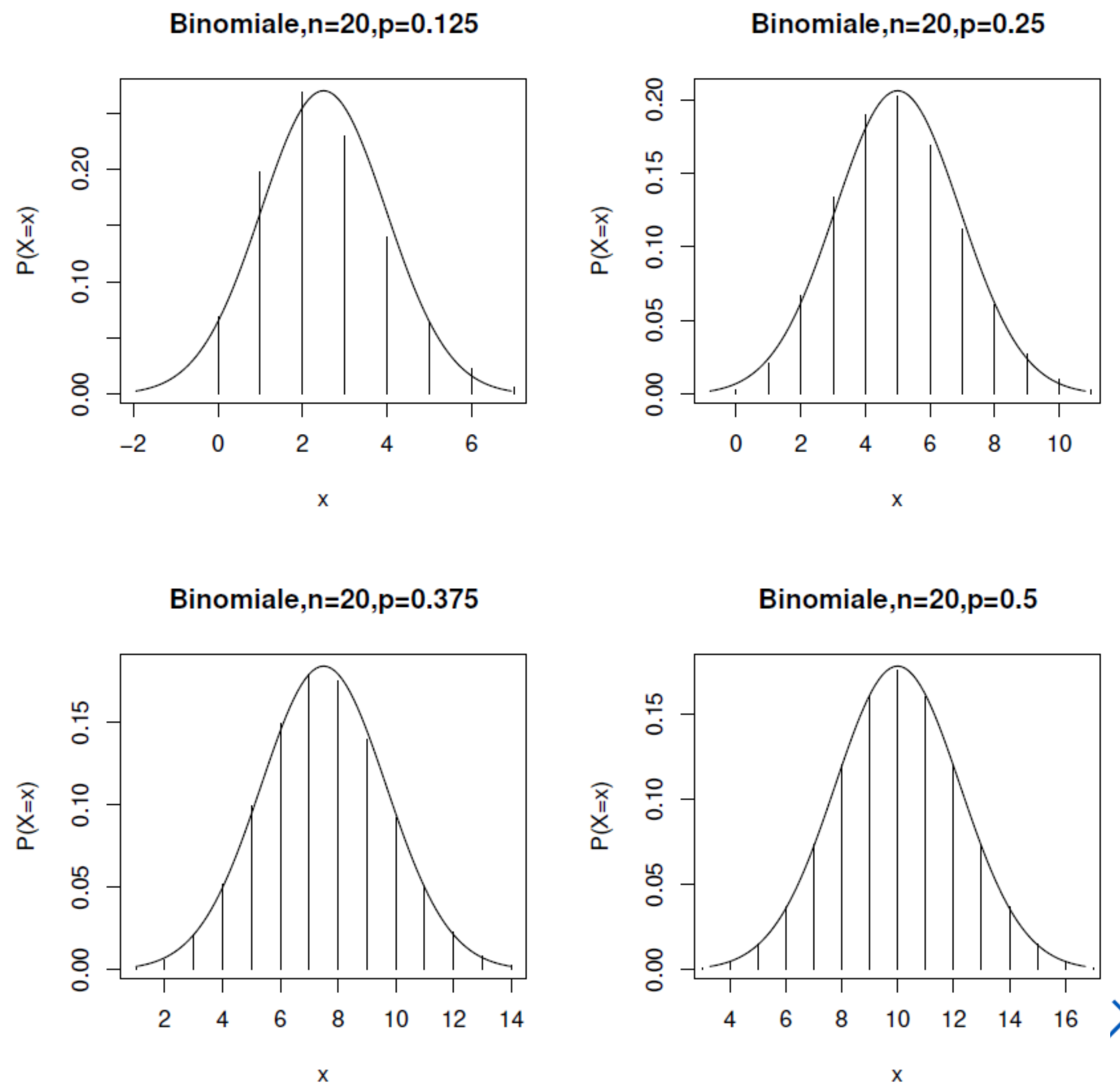
```
par(mfrow = c(2, 2))
p <- 0.2
q <- 1 - p

# Primo grafico: n = 25
x <- 0:25
n <- 25
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 25, p = 0.2")
lines(x, dbinom(x, n, 0.2), type = "h")

# Secondo grafico: n = 50
x <- 0:50
n <- 50
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 50, p = 0.2")
lines(x, dbinom(x, n, 0.2), type = "h")

# Terzo grafico: n = 75
x <- 0:75
n <- 75
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 75, p = 0.2")
lines(x, dbinom(x, n, 0.2), type = "h")

# Quarto grafico: n = 100
x <- 0:100
n <- 100
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 100, p = 0.2")
lines(x, dbinom(x, n, 0.2), type = "h")
```



Approssimazione della binomiale con la normale

- Confrontiamo la densità normale di valore medio np e varianza $np(1-p)$ e la funzione di probabilità binomiale per n fissato e $p = 0.125, 0.25, 0.375, 0.5$

```r
par(mfrow = c(2, 2))
x <- 0:20
n <- 20

# Primo grafico: p = 0.125
p <- 0.125
q <- 1 - p
curve(dnorm(x, n * p, sqrt(n * p * q)),  from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 20, p = 0.125")
lines(x, dbinom(x, n, 0.125), type = "h")

# Secondo grafico: p = 0.25
p <- 0.25
q <- 1 - p
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 20, p = 0.25")
lines(x, dbinom(x, n, 0.25), type = "h")

# Terzo grafico: p = 0.375
p <- 0.375
q <- 1 - p
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 20, p = 0.375")
lines(x, dbinom(x, n, 0.375), type = "h")

# Quarto grafico: p = 0.5
p <- 0.5
q <- 1 - p
curve(dnorm(x, n * p, sqrt(n * p * q)), from = n * p - 3 * sqrt(n * p * q),
      to = n * p + 3 * sqrt(n * p * q), xlab = "x", ylab = "P(X = x)",
      main = "Binomiale, n = 20, p = 0.5")
lines(x, dbinom(x, n, 0.5), type = "h")
```



Teorema centrale del limite

- Uno dei più importanti risultati della teoria della probabilità è il **teorema centrale di convergenza** o **teorema centrale del limite**
- Il teorema fornisce una semplice ed utile approssimazione della distribuzione della somma di variabili aleatorie indipendenti
- Il Teorema dice che:
  - **La somma di molte variabili aleatorie indipendenti tende a distribuirsi come una normale, INDIPENDENTEMENTE da come sono distribuite le variabili originali.**
    - È come se la natura avesse una tendenza alla normalità quando si combinano molti fenomeni casuali
- Le variabili $X_1, X_2, \ldots$ devono essere:
  - **Indipendenti**
  - **Identicamente distribuite** (stessa distribuzione)
  - Con **media finita** μ
  - Con **varianza finita** σ^2

Teorema centrale del limite

- **Teorema**:
  - Sia $X_1, X_2, \ldots$ una successione di variabili aleatorie definite nello stesso spazio di probabilità, indipendenti e identicamente distribuite con valore medio μ e varianza σ^2
  - Posto per ogni intero n positivo $Y_n = X_1 + X_2 + \cdots + X_n$ per ogni $x \in \mathrm{R}$ si ha:

$$\frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}} = \frac{Y_n - n\mu}{\sigma\sqrt{n}} \to Z$$

ossia la successione delle variabili aleatorie standardizzate:

$$\lim_{n\to\infty}\left(\frac{Y_n - \ n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-y^2}{2}} dx$$

converge in distribuzione alla variabile aleatoria normale standard

Approssimazione della distribuzione di somme di Var. Aleatorie

- Dal Teorema di De Moivre-Laplace:
  - ricordiamo che sottraendo a:

$$Y_n = X_1 + X_2 + \cdots + X_n$$

    la media $n\mu$ e dividendo per la deviazione standard $n\sqrt{\sigma}$ si ottiene una variabile aleatoria **normale standard** la cui funzione di distribuzione è per n grande approssimativamente una variabile normale standard

- Quindi Y_n è approssimativamente una variabile normale con valore medio $n\mu$ e varianza $n\sigma^2$, ossia:

$$Y_n \cong n\mu + Z\, n\sqrt{\sigma}$$

Media campionaria

- <u>Media campionaria</u>: se la popolazione da cui è estratto il campione è descritta da una variabile aleatoria discreta o continua X, con valore medio μ e varianza σ^2, entrambi finiti, allora la media campionaria $\overline{X_n}$ per campioni **numerosi**, è approssimativamente normale con valore medio μ e varianza $\frac{\sigma^2}{n}$
  - Se denotiamo con:

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \longleftarrow \text{Media Campionaria}$$

Dal teorema del limite centrale ricordiamo che: $\frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}} = \frac{Y_n - n\mu}{\sigma\sqrt{n}} \rightarrow Z$

e si ha che: $\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow Z$

cioè converge in distribuzione alla variabile aleatoria normale standard.

- Quindi, per n grande la distribuzione della media campionaria $\overline{X_n}$ è **approssimativamente normale** con valore medio μ e varianza $\frac{\sigma^2}{n}$, ossia: $\overline{X_n} \approx \mu + \frac{\sigma}{\sqrt{n}} \rightarrow Z$

Approssimazione Poisson con la Normale

- Siano $X_1, X_2, \ldots$ una successione di variabili aleatorie indipendenti di Poisson ognuna di parametro λ allora $Y_n = X_1 + X_2 + \cdots + X_n$ è ancora una variabile aleatoria di Poisson di parametro $n\lambda$

Valore atteso: $E(Y_n) = n\lambda$ **Varianza**: $Var(Y_n) = n\lambda$

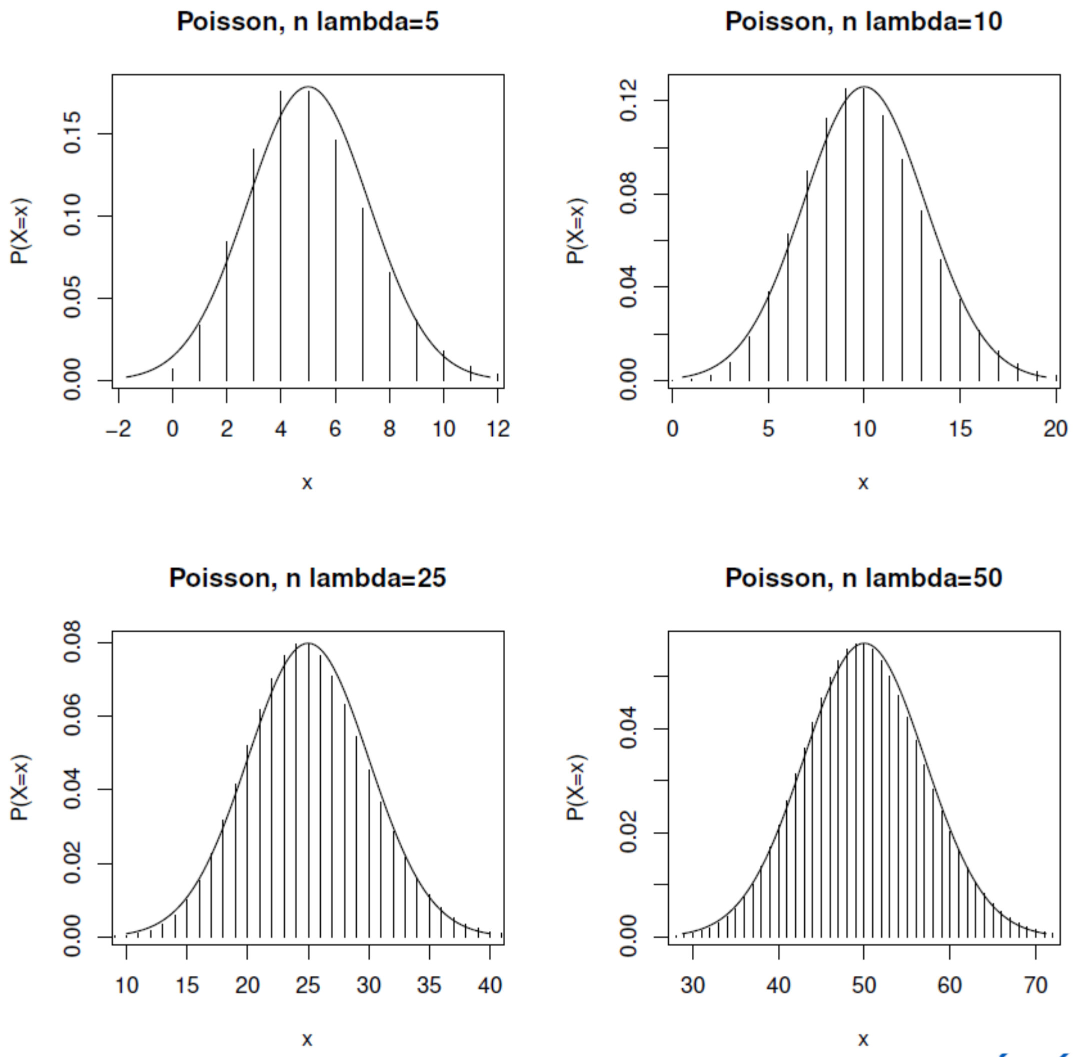
- Il **teorema centrale di convergenza** afferma che per n grande la distribuzione di $Y_n = X_1 + X_2 + \cdots + X_n$ è approssimativamente normale con valore medio $n\lambda$ e varianza $n\lambda$, ossia:

$$Y_n \cong n\lambda + Z\sqrt{n\lambda}$$ Variabile aleatoria Normale

Approssimazione Poisson con la Normale

- L'approssimazione della distribuzione di Poisson della variabile al variare del parametro $n\lambda$:

```
par(mfrow = c(2, 2))

# Primo grafico: lambda = 5
x <- 0:100
curve(dnorm(x, 5, sqrt(5)), from = 5 - 3 * sqrt(5),
      to = 5 + 3 * sqrt(5), xlab = "x", ylab = "P(X = x)",
      main = "Poisson, lambda = 5")
lines(x, dpois(x, 5), type = "h")

# Secondo grafico: lambda = 10
curve(dnorm(x, 10, sqrt(10)), from = 10 - 3 * sqrt(10),
      to = 10 + 3 * sqrt(10), xlab = "x", ylab = "P(X = x)",
      main = "Poisson, lambda = 10")
lines(x, dpois(x, 10), type = "h")

# Terzo grafico: lambda = 25
curve(dnorm(x, 25, sqrt(25)), from = 25 - 3 * sqrt(25),
      to = 25 + 3 * sqrt(25), xlab = "x", ylab = "P(X = x)",
      main = "Poisson, lambda = 25")
lines(x, dpois(x, 25), type = "h")

# Quarto grafico: lambda = 50
curve(dnorm(x, 50, sqrt(50)), from = 50 - 3 * sqrt(50),
      to = 50 + 3 * sqrt(50), xlab = "x", ylab = "P(X = x)",
      main = "Poisson, lambda = 50")
lines(x, dpois(x, 50), type = "h")
```

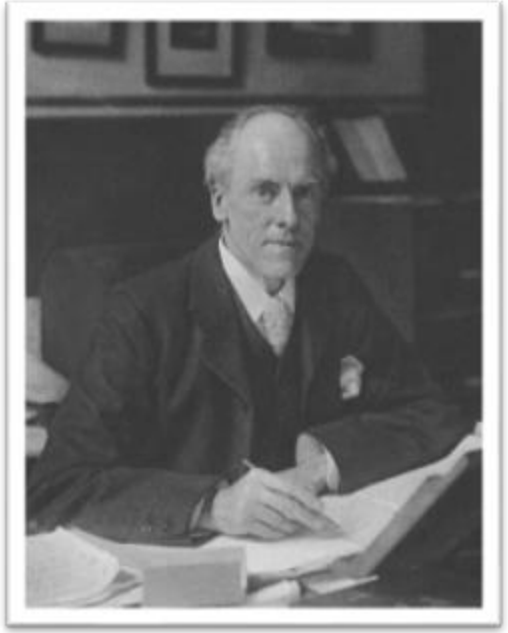


Si nota che l'approssimazione migliora al crescere di $n\lambda$

STATISTICA E ANALISI DEI DATI

Distribuzione chi-quadrato

Distribuzione chi-quadrato



- La distribuzione chi-quadrato fu introdotta intorno al 1900 da un famoso matematico inglese Karl Pearson, considerato uno dei fondatori della statistica matematica
- Riveste un ruolo importante nell'inferenza statistica, in particolare nella **stima intervallare** della varianza di una popolazione normale, ed anche in molti test di **verifica di ipotesi statistiche**
- Per definire la densità chi-quadrato occorre introdurre la **funzione gamma**:

$$\Gamma(v) = \int_0^{+\infty} x^{v-1} e^{-x}\, dx \qquad v > 0$$

  - Se $v > 1$, per la funzione $\Gamma(v)$ sussiste la seguente proprietà di fattorizzazione:

$$\Gamma(v) = (v - 1)\,\Gamma(v - 1) \qquad v > 1$$

- La funzione Γ è una generalizzazione dei **fattoriali**; infatti, se v è un intero positivo, usando iterativamente $\Gamma(v)$ si ha:

$$\Gamma(1) = (0)! = 1 \qquad \Gamma(2) = (2)! = 1 \qquad \Gamma(3) = (2)! = 2 \qquad \Gamma(v) = (v - 1)! \;\; v > 1$$

- **Nota**: La funzione Γ restituisce il valore fattoriale (!) per numeri non interi

Distribuzione chi-quadrato

- Per calcolare la funzione gamma in R si usa:

```
> gamma(1/2)
[1] 1.772454
> gamma(3/2)
[1] 0.886227
```

- Che mostra:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

per la proprietà di fattorizzazione:

$$\Gamma(v) = (\nu - 1)\Gamma(\nu - 1) \qquad v > 1$$

```
x <- seq(1, 10, length.out = 100)
print(x)
```

```
 [1]  1.000000  1.090909  1.181818  1.272727  1.363636  1.454545  1.545455
 [8]  1.636364  1.727273  1.818182  1.909091  2.000000  2.090909  2.181818
[15]  2.272727  2.363636  2.454545  2.545455  2.636364  2.727273  2.818182
[22]  2.909091  3.000000  3.090909  3.181818  3.272727  3.363636  3.454545
[29]  3.545455  3.636364  3.727273  3.818182  3.909091  4.000000  4.090909
[36]  4.181818  4.272727  4.363636  4.454545  4.545455  4.636364  4.727273
[43]  4.818182  4.909091  5.000000  5.090909  5.181818  5.272727  5.363636
[50]  5.454545  5.545455  5.636364  5.727273  5.818182  5.909091  6.000000
[57]  6.090909  6.181818  6.272727  6.363636  6.454545  6.545455  6.636364
[64]  6.727273  6.818182  6.909091  7.000000  7.090909  7.181818  7.272727
[71]  7.363636  7.454545  7.545455  7.636364  7.727273  7.818182  7.909091
[78]  8.000000  8.090909  8.181818  8.272727  8.363636  8.454545  8.545455
[85]  8.636364  8.727273  8.818182  8.909091  9.000000  9.090909  9.181818
[92]  9.272727  9.363636  9.454545  9.545455  9.636364  9.727273  9.818182
[99]  9.909091 10.000000
```

Distribuzione chi-quadrato

- Per calcolare la funzione gamma in R si usa:

```
> gamma(1/2)
[1] 1.772454
> gamma(3/2)
[1] 0.886227
```

- Che mostra:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

per la proprietà di fattorizzazione:

$$\Gamma(v) = (\nu - 1)\Gamma(\nu - 1) \qquad v > 1$$

```r
x <- seq(1, 10, length.out = 100)
print(x)

# Calcolare i valori di Gamma(x)
gamma_values <- gamma(x)
```

```
  [1] 1.000000e+00 9.550795e-01 9.232014e-01 9.020056e-01 8.898568e-01
  [6] 8.856248e-01 8.885444e-01 8.981248e-01 9.140893e-01 9.363360e-01
 [11] 9.649125e-01 1.000000e+00 1.041905e+00 1.091056e+00 1.148007e+00
 [16] 1.213441e+00 1.288181e+00 1.373205e+00 1.469659e+00 1.578881e+00
 [21] 1.702429e+00 1.842106e+00 2.000000e+00 2.178529e+00 2.380486e+00
 [26] 2.609107e+00 2.868133e+00 3.161900e+00 3.495431e+00 3.874555e+00
 [31] 4.306040e+00 4.797755e+00 5.358853e+00 6.000000e+00 6.733634e+00
 [36] 7.574275e+00 8.538896e+00 9.647358e+00 1.092293e+01 1.239289e+01
 [41] 1.408929e+01 1.604979e+01 1.831870e+01 2.094824e+01 2.400000e+01
 [46] 2.754668e+01 3.167424e+01 3.648437e+01 4.209756e+01 4.865668e+01
 [51] 5.633133e+01 6.532308e+01 7.587172e+01 8.826282e+01 1.028368e+02
 [56] 1.200000e+02 1.402377e+02 1.641301e+02 1.923721e+02 2.257960e+02
 [61] 2.654000e+02 3.123828e+02 3.681846e+02 4.345380e+02 5.135292e+02
 [66] 6.076722e+02 7.200000e+02 8.541748e+02 1.014623e+03 1.206698e+03
 [71] 1.436884e+03 1.713037e+03 2.044687e+03 2.443407e+03 2.923256e+03
 [76] 3.501335e+03 4.198462e+03 5.040000e+03 6.056876e+03 7.286836e+03
 [81] 8.775986e+03 1.058069e+04 1.276991e+04 1.542810e+04 1.865874e+04
 [86] 2.258880e+04 2.737407e+04 3.320602e+04 4.032000e+04 4.900563e+04
 [91] 5.961957e+04 7.260134e+04 8.849304e+04 1.079638e+05 1.318401e+05
 [96] 1.611437e+05 1.971386e+05 2.413896e+05 2.958354e+05 3.628800e+05
```

Distribuzione chi-quadrato

- Per calcolare la funzione gamma in R si usa:

```
> gamma(1/2)
[1] 1.772454
> gamma(3/2)
[1] 0.886227
```

- Che mostra:

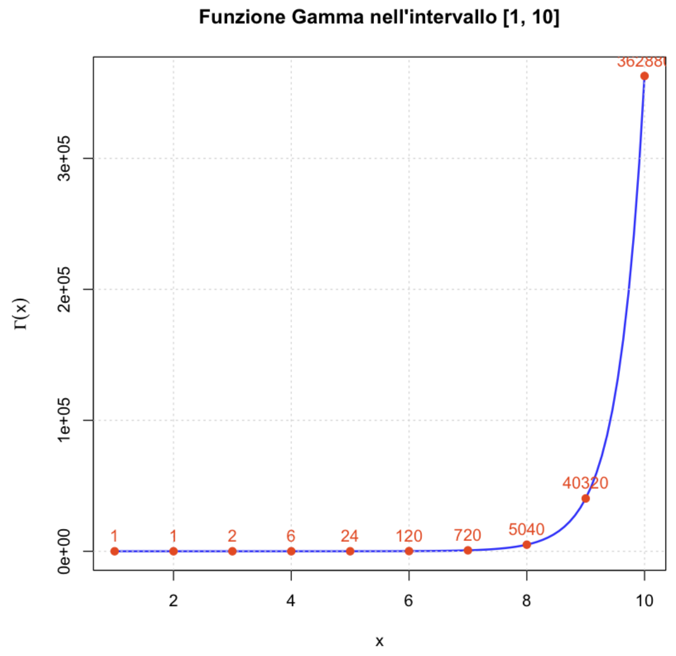
$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

per la proprietà di fattorizzazione:

$$\Gamma(v) = (\nu - 1)\Gamma(\nu - 1) \qquad v > 1$$

```r
x <- seq(1, 10, length.out = 100)
print(x)

# Calcolare i valori di Gamma(x)
gamma_values <- gamma(x)

# Disegnare la funzione Gamma
plot(x, gamma_values, type = "l", col = "blue", lwd = 2,
     xlab = "x", ylab = expression(Gamma(x)),
     main = "Funzione Gamma nell'intervallo [1, 10]")
grid()

points(1:10, gamma(1:10), pch = 19, col = "red")
text(1:10, gamma(1:10), labels = round(gamma(1:10), 2), pos = 3, col = "red")
```



Distribuzione chi-quadrato

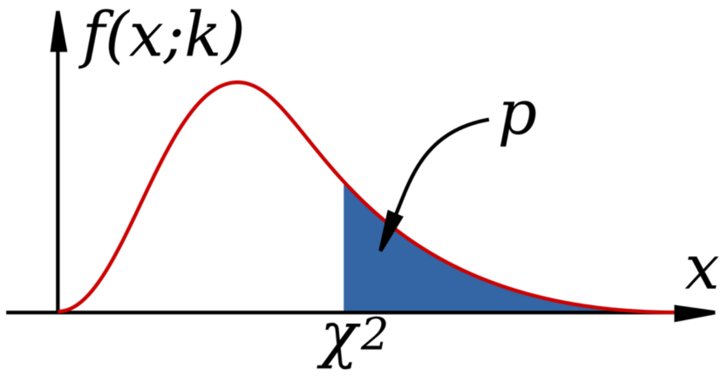
- Siano $X_1, X_2, \dots, X_n$ sono variabili aleatorie normali indipendenti con $\mu = 0$ e $\sigma = 1$ la somma dei loro quadrati:

$$Y_n = {X_1}^2 + {X_2}^2 + \dots + {X_n}^2$$

si distribuisce come un chi quadrato con $\boldsymbol{n}$ **gradi di libertà** (numero degli addendi considerati)

- Una variabile aleatoria X con:

Densità di probabilità: $$f_X(x) = \begin{cases} \dfrac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\dfrac{1}{2}\right)^{\frac{n}{2}} (x)^{\frac{n}{2}-1} (e)^{-\frac{x}{2}} & con\ x > 0 \\ 0 & con\ x \leq 0 \end{cases}$$



si dice avere distribuzione chi-quadrato con $\boldsymbol{n}$ **gradi di libertà**

Notazione:

- $X \sim \chi^2(n)$ variabile aleatoria con distribuzione chi-quadrato

Distribuzione chi-quadrato

- Per il calcolo della densità chi-quadrato si utilizza la funzione:

```
dchisq(x, df)
```

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

- Per calcolare la funzione di distribuzione chi-quadrato invece utilizziamo la funzione:

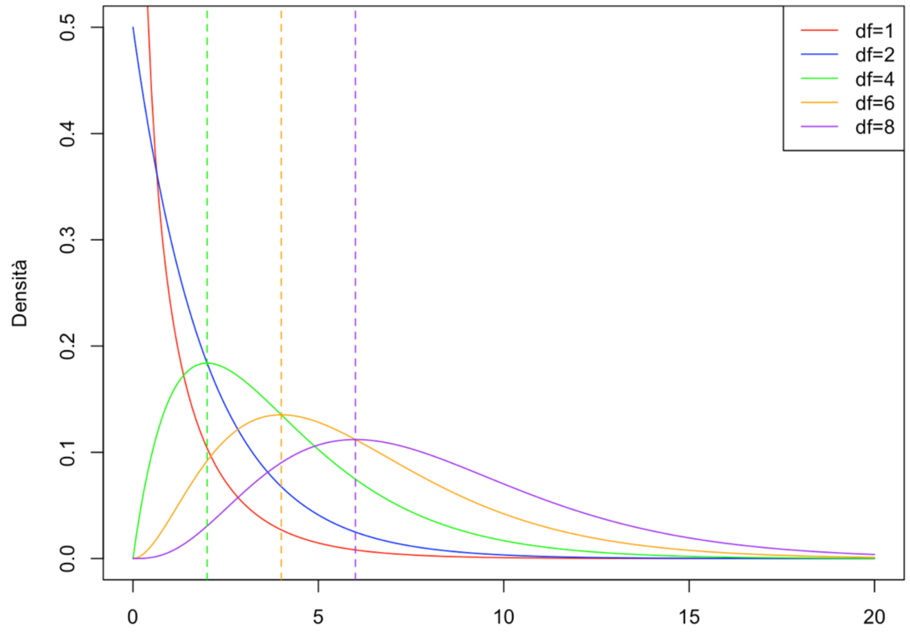
```
pchisq(x, df, lower.tail = TRUE)
```

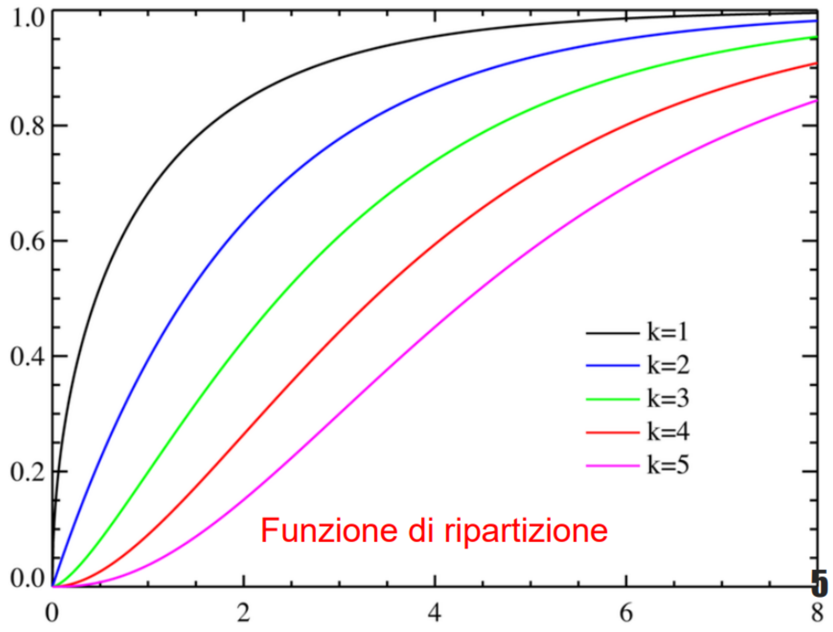
dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
- **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)
- **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

Distribuzione chi-quadrato

- Si nota che la funzione densità di $X \sim \chi^2(n)$ con n gradi di libertà è strettamente decrescente per $n = 1,2,\ldots$ mentre per $n > 2$ presenta un unico punto di massimo in $x = n - 2$





- Quando $n = 2$ la funzione di densità corrisponde ad una densità esponenziale di valore medio $\frac{1}{\lambda} = 2$
- Per una variabile aleatoria normale si ha:

Valore atteso: $E(X) = n$ **Varianza**: $Var(X) = 2n$

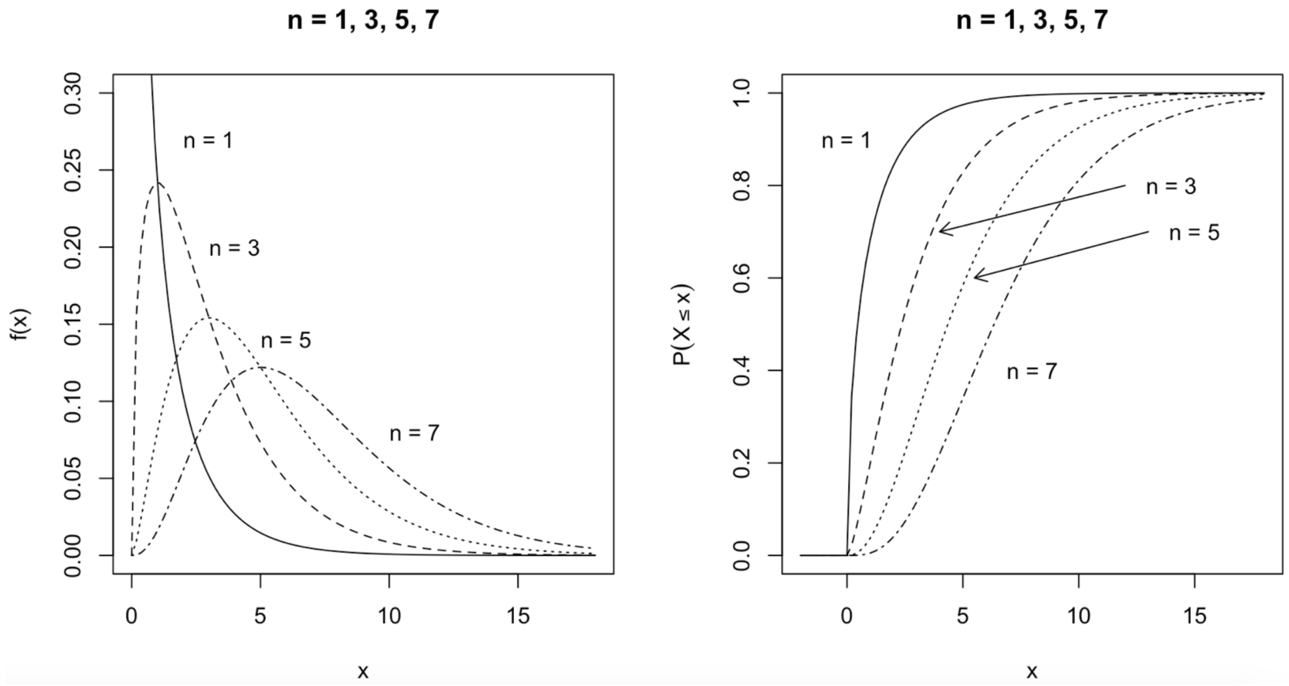
Distribuzione chi-quadrato

- **Esempio**:

```
par(mfrow = c(1, 2))

# Primo grafico: funzioni di densità
curve(dchisq(x, df = 1), from = 0, to = 18, ylim = c(0, 0.3),
      xlab = "x", ylab = "f(x)", main = "n = 1, 3, 5, 7")
curve(dchisq(x, df = 3), add = TRUE, lty = 2)
curve(dchisq(x, df = 5), add = TRUE, lty = 3)
curve(dchisq(x, df = 7), add = TRUE, lty = 4)

# Etichette per le densità
text(3, 0.27, "n = 1")
text(4, 0.20, "n = 3")
text(6, 0.14, "n = 5")
text(11, 0.08, "n = 7")

# Secondo grafico: funzioni di ripartizione
curve(pchisq(x, df = 1), from = -2, to = 18, ylim = c(0, 1),
      xlab = "x", ylab = expression(P(X <= x)), main = "n = 1, 3, 5, 7")
curve(pchisq(x, df = 3), add = TRUE, lty = 2)
curve(pchisq(x, df = 5), add = TRUE, lty = 3)
curve(pchisq(x, df = 7), add = TRUE, lty = 4)

# Etichette per le ripartizioni
text(0, 0.9, "n = 1")
arrows(4, 0.7, 12, 0.8, code = 1, length = 0.10)
text(14, 0.8, "n = 3")
arrows(5.5, 0.6, 13, 0.7, code = 1, length = 0.10)
text(15, 0.7, "n = 5")
text(8, 0.4, "n = 7")
```

Densità

Distribuzione



La densità è decrescente per $n = 1$ e presenta un unico punto di massimo in $n - 2$ quando $n = 3{,}5{,}7$

Quantili

- Per calcolare i quantili (percentili) della distribuzione chi-quadrato

```
qchisq(z, df)
```

dove

  - **z** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile $z \cdot 100$–esimo, cioè il vettore delle probabilità;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

- È possibile simulare la variabile aleatoria chi-quadrato generando una sequenza di numeri pseudocasuali mediante la funzione

```
rchisq(N, df)
```

dove

  - **N** è lunghezza della sequenza da generare;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

STATISTICA E ANALISI DEI DATI

Distribuzione Student

La storia di William Gosset (alias “Student”)

- All’inizio del Novecento, **William Sealy Gosset** non era un professore universitario né uno statistico accademico.
    - Era un **chimico inglese**, assunto dalla celebre fabbrica di birra **Guinness** a Dublino. Il suo lavoro? Migliorare la qualità della birra usando approcci scientifici.
- **Il problema: pochi dati, grandi decisioni**
    - Nella produzione della birra, molte decisioni importanti, come **selezionare il miglior orzo**, dovevano essere prese su **campioni molto piccoli**, perché analizzare grandi quantità era costoso.
    - All’epoca, però, la statistica si concentrava soprattutto su campioni grandi.
    - Le tabelle dei valori critici erano costruite assumendo che si potesse applicare la distribuzione normale. Con pochi dati, questa approssimazione era **ingannevole**
- Gosset si trovò quindi davanti a una domanda fondamentale:
    - *Come si stima la media di una popolazione quando si hanno pochissimi campioni, e la varianza è incerta?*
- Questo lo portò a sviluppare una nuova distribuzione, nata proprio per gestire queste situazioni

Distribuzione di Student

- Una variabile aleatoria X:

Densità di probabilità: $f_X(x) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$ $\quad con\ x \in \mathrm{R}$

si dice avere distribuzione di Student con $n-1$ gradi di libertà

<u>Notazione</u>:

- $X \sim T(n)$ variabile aleatoria con distribuzione di Student

- Al limite per $n \to +\infty$ la densità di Student converge alla **densità normale standard**, si ha che:

$$\lim_{n\to+\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\} \qquad con\ x \in \mathrm{R}$$

Distribuzione di Student

- Per il calcolo della densità di Student si utilizza la funzione:

```
dt(x, df)
```

dove

  - **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

- Per calcolare la funzione di distribuzione chi-quadrato invece utilizziamo la funzione:

```
pt(x, df, lower.tail = TRUE)
```

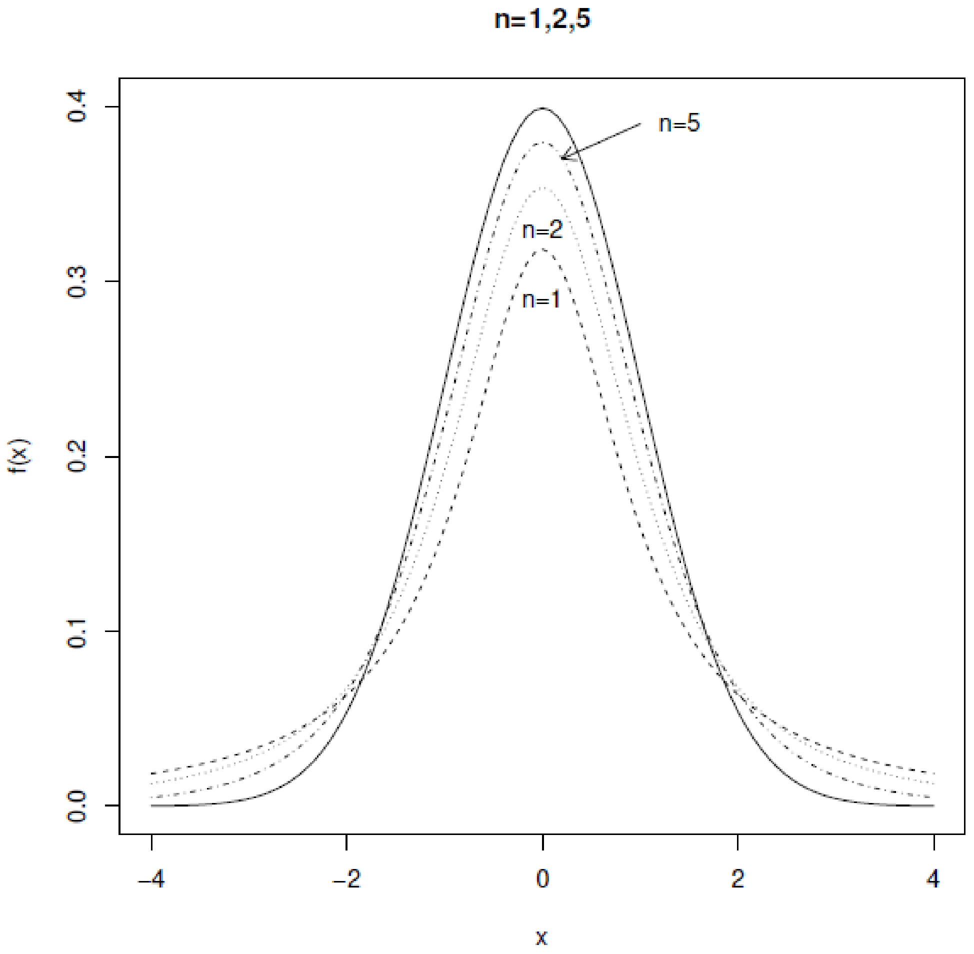
dove

  - **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)
  - **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

Distribuzione di Student

- **Esempio**:

```
curve(dnorm(x, mean = 0, sd = 1), from = -4, to = 4,
      ylim = c(0, 0.4),
      xlab = "x", ylab = "f(x)", main = "n = 1, 2, 5")
curve(dt(x, df = 1), add = TRUE, lty = 2)
text(0, 0.29, "n = 1")

curve(dt(x, df = 2), add = TRUE, lty = 3)
text(0, 0.33, "n = 2")

curve(dt(x, df = 5), add = TRUE, lty = 4)
arrows(0.2, 0.37, 1.0, 0.39, code = 1, length = 0.10)
text(1.4, 0.39, "n = 5")
```

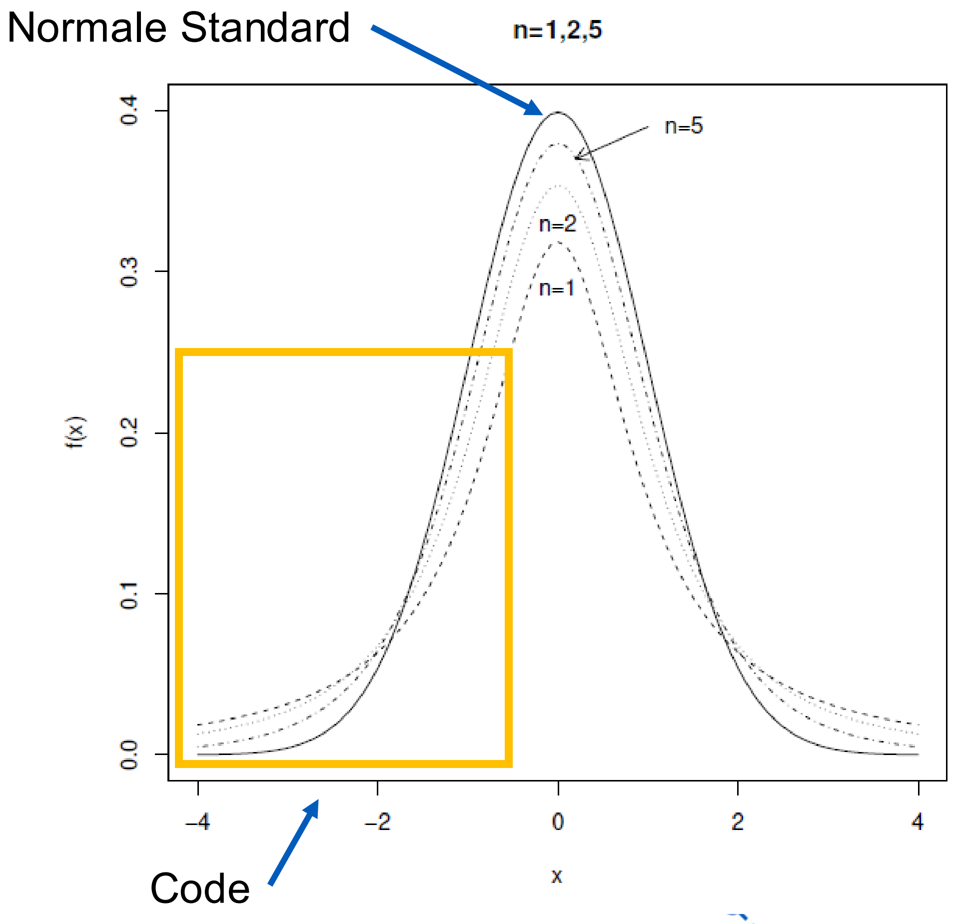


Distribuzione di Student

- Si nota che la distribuzione di Student:
  - è <u>simmetrica intorno all'asse</u> $x = 0$
  - dipende dal solo parametro n rappresentante il numero di gradi di libertà
    - Per $n = 1$ si identifica con la densità di Cauchy

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

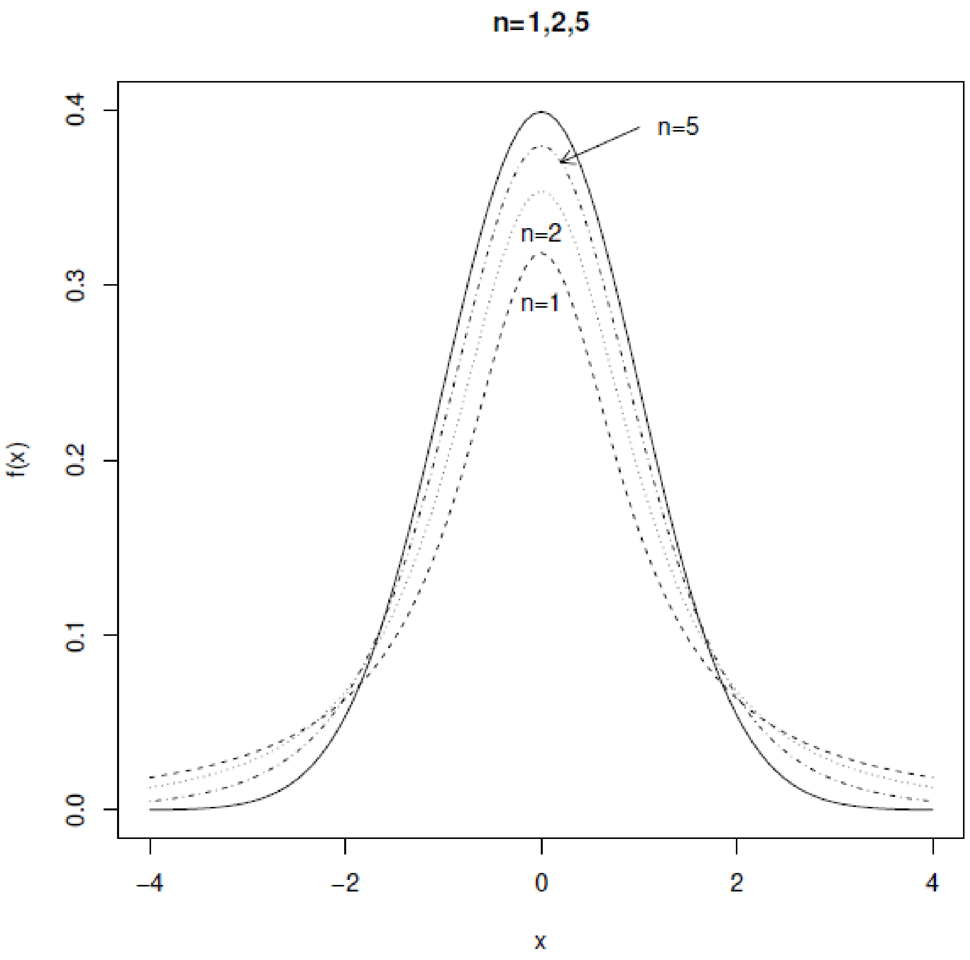
  - È molto simile alla densità normale standard
    - Il picco della densità di Student è più basso
    - le sue code sono più allungate rispetto alla densità normale standard



Distribuzione di Student

- Il valore medio della variabile aleatoria X con distribuzione di Student ha:
  - **Valore medio**:
    - Non esiste se $n = 1$
    - È zero se $n = 2,3, \dots$
  - **Varianza:**
    - Non esiste se $n = 1$
    - Se $n = 2,3, \dots$ si ha che:

$$Var(X) = \frac{n}{n-2}$$



Quantili

- Per calcolare i quantili (percentili) della distribuzione di Student

```
qt(p, df)
```

dove

  - **p** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile $z \cdot 100$–esimo, cioè il vettore delle probabilità;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

- È possibile simulare la variabile aleatoria di Student generando una sequenza di numeri pseudocasuali mediante la funzione

```
rt(N, df)
```

dove

  - **N è** lunghezza della sequenza da generare;
  - **df** numero di gradi di libertà (non negativo, può anche essere un valore non intero)

Student, chi–quadrato e normale standard

- Il teorema seguente mostra la stretta connessione esistente tra una variabile a distribuzione di Student e variabili a distribuzione chi–quadrato e normale standard
- **Teorema**:
  - Siano $Y \sim \chi^2(n)$ e $Z \sim N(0,1)$ variabili aleatorie indipendenti
  - Allora:

$$X = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

ha distribuzione di Student con n gradi di libertà

Student, chi–quadrato e normale standard

- Siano $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e normali con valore medio μ e varianza σ^2 e sia $S_n{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ la variabile aleatoria che denota la varianza campionaria di $X_1, X_2, \dots, X_n$:

  - $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ è distribuita con legge normale standard

  - $T_n = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}}$ è distribuita con legge di Student con $n-1$ gradi di libertà

  - $Q_n = \frac{(n-1)S_n{}^2}{\sigma^2}$ è distribuita con legge Normale standard

- Inoltre, se $X_1, X_2, \dots, X_n$ sono variabili aleatorie indipendenti e <u>identicamente distribuite</u> con valore medio μ e varianza σ^2 allora per il teorema centrale di convergenza:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow Z$$

ossia converge in distribuzione alla variabile aleatoria normale standard

Tabelle per le distribuzioni continue

Distribuzione	Notazione	Funzione densità di probabilità
Uniforme	$X \sim \mathcal{U}(a,b)$	$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{altrimenti} \end{cases} \quad (a, b \in \mathbb{R})$
Esponenziale	$X \sim \mathcal{E}(1,\lambda)$	$f_X(x) = \begin{cases} \lambda\, e^{-\lambda\, x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases} \quad (\lambda > 0)$
Normale	$X \sim \mathcal{N}(\mu,\sigma)$	$f_X(x) = \dfrac{1}{\sigma\sqrt{2\,\pi}}\,\exp\left\{-\dfrac{(x-\mu)^2}{2\,\sigma^2}\right\}, \quad x \in \mathbb{R}$ $(\mu \in \mathbb{R}, \sigma > 0)$
Chi-quadrato	$X \sim \chi^2(n)$	$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)}\, e^{-x/2}\, x^{n/2-1}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ $(n = 1, 2, \ldots)$
Student	$X \sim \mathcal{T}(n)$	$f_X(x) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\,\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \dfrac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}$ $(n = 1, 2, \ldots)$

Distribuzione	Valore medio $E(X)$	Varianza $\mathrm{Var}(X)$	Coefficiente di variazione $CV(X)$
Uniforme	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{b-a}{\sqrt{3}\,(a+b)}$
Esponenziale	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	1
Normale	μ	σ^2	$\dfrac{\sigma}{\mu}$
Chi-quadrato	n	$2\,n$	$\sqrt{\dfrac{2}{n}}$
Student	0 se $n > 1$	$\begin{cases} +\infty & \text{se } n = 2 \\ \dfrac{n}{n-2} & \text{se } n = 3, 4, \ldots \end{cases}$	

Tabelle per le distribuzioni continue (R)

Nome	Densità Distribuzione Quantili	Simulazione
Uniforme	dunif(x,min=a,max=b) punif(x,min=a,max=b) qunif(z,min=a,max=b)	runif(N,min=a,max=b)
Esponenziale	dexp(x,rate=lambda) pexp(x,rate=lambda) qexp(z,rate=lambda)	rexp(N,rate=lambda)
Normale	dnorm(x,mean=mu,sd=sigma) pnorm(x,mean=mu,sd=sigma) qnorm(z,mean=mu,sd=sigma)	rnorm(N,mean=mu,sd=sigma)
Chi-quadrato	dchisq(x, df) pchisq(x, df) qchisq(z, df)	rchisq(N, df)
Student	dt(x, df) pt(x, df) qt(z, df)	rt(N, df)