Corso di Statistica Psicometrica

Canale 1 - Matricole Dispari

CdS triennale in
Scienze e Tecniche
Psicologiche

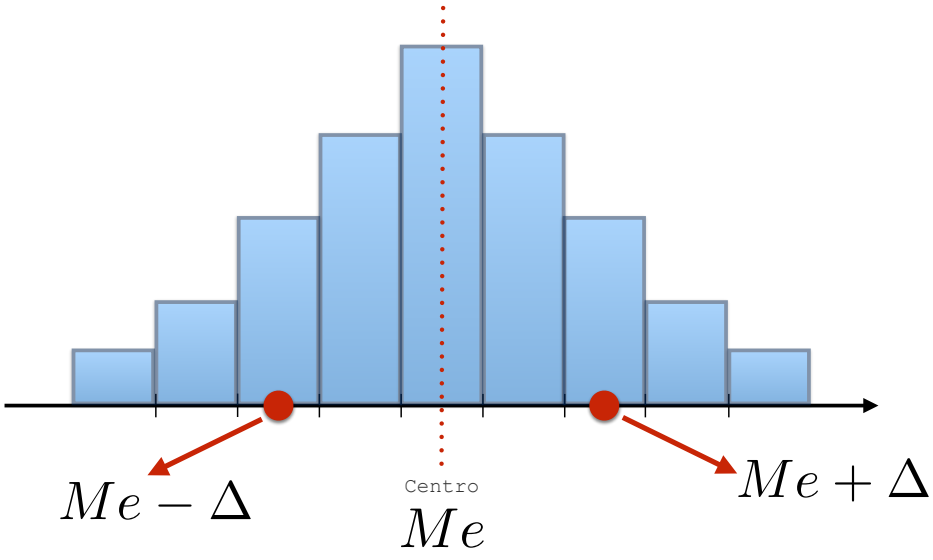
Domenico Vistocco
Dipartimento di Scienze Politiche
Università degli Studi di Napoli Federico II

Attenzione…

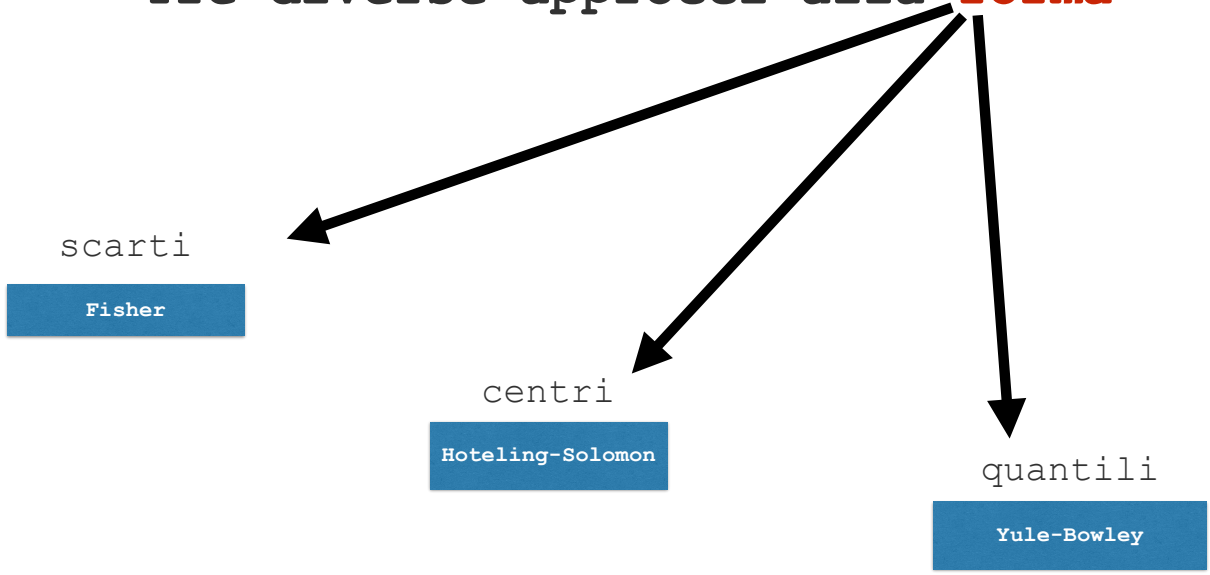
Lei ha il diritto di rimanere in silenzio: qualsiasi cosa dirà / farà potrà essere usata contro di lei…

Nella puntata precedente…

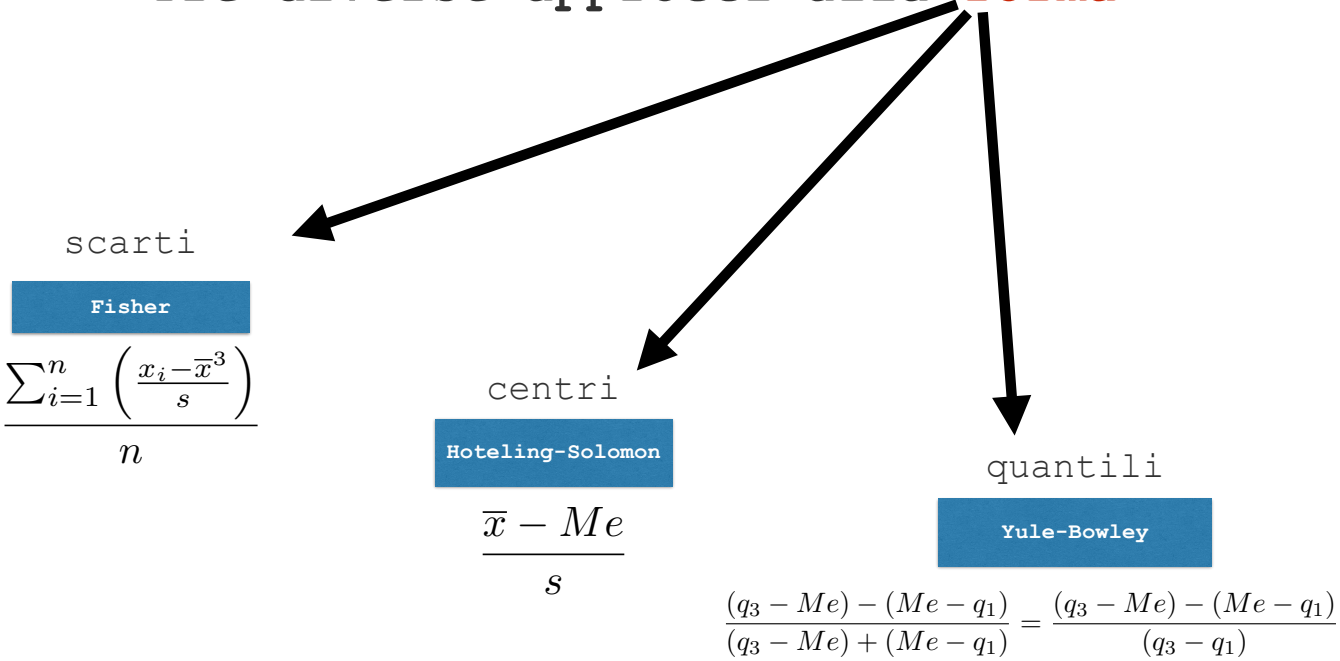
- Tipologia di variabili
- Tipologia di tabelle
- Frequenza assolute, relative e percentuali
- Frequenze cumulate
- Densità di frequenza
- Rappresentazioni grafiche
- Indici di posizione: la classe particolare degli indici di tendenza centrale
- Proprietà di media aritmetica e mediana
- Modelli statistici: i centri come "semplici" modelli
- Misura della variabilità
- Indici di disuguaglianza
- Indici di dispersione
- Indici basati sul confronto tra posizione opposte
- Un indicatore robusto di variabilità: il MAD (median absolute deviation)
- I coefficienti di variazione
- Proprietà desiderabili di un indice di variabilità e comportamento delle tre classi di indicatori
- Tre particolari trasformazioni lineari: centratura, riduzione e standardizzazione
- Proprietà di decomposizione della devianza / varianza
- **Indici di forma**
- **Sintesi a 5 e boxplot**



Tre diverse approcci alla forma

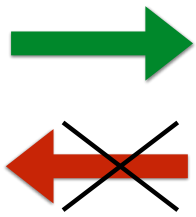


Tre diverse approcci alla forma



Attenzione ai falsi segnali

Distribuzione simmetrica



Indice di forma = 0

Il ruolo degli indici di tendenza centrale: l'indice di Hotelling - Solomon

dati → 2, 8, 10, 10, 10, 15, 15

$$Mo = Me = \bar{x}$$

scarti → -8, -2, 0, 0, 0, +5, +5

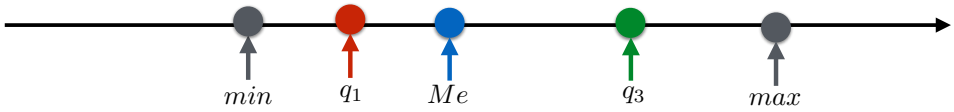
scarti^3 → -512, -8, 0, 0, 0, +125, +125

La sintesi a 5: min, q_1, Me, q_3, max

5 indici di posizione

2 (e non solo) indici di variabilità

1 (e non solo) indice di forma

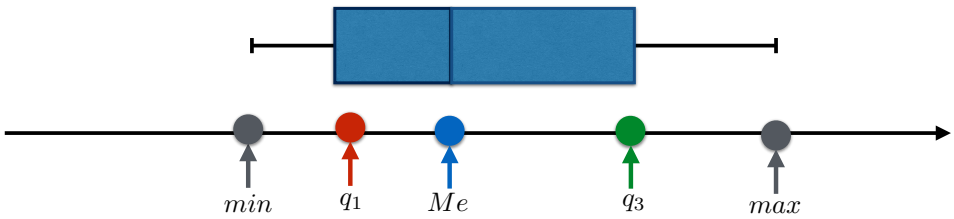


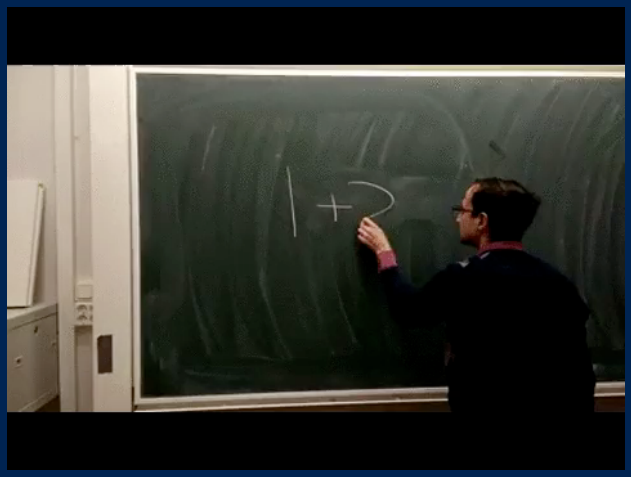
La sintesi a 5: il boxplot

5 indici di posizione

2 (e non solo) indici di variabilità

1 (e non solo) indice di forma

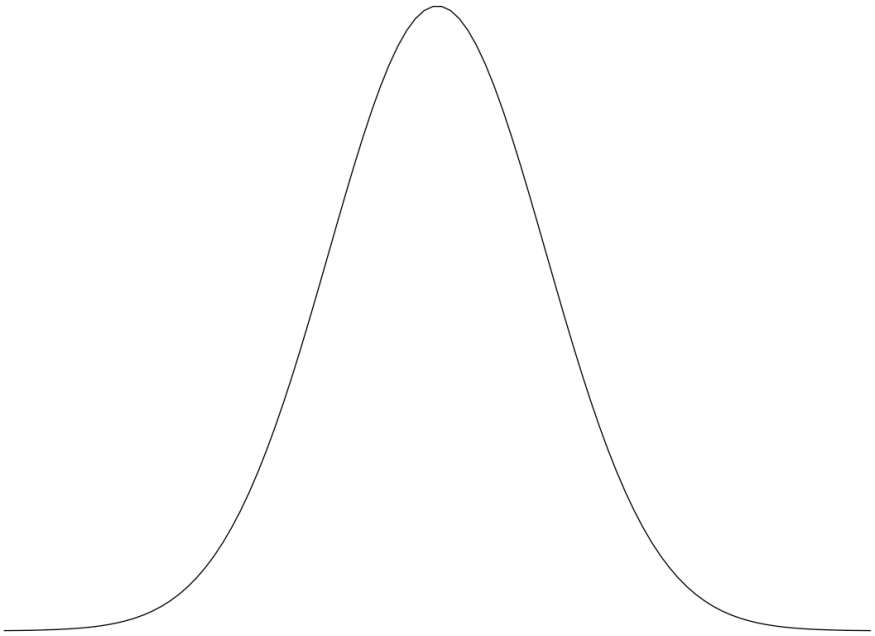




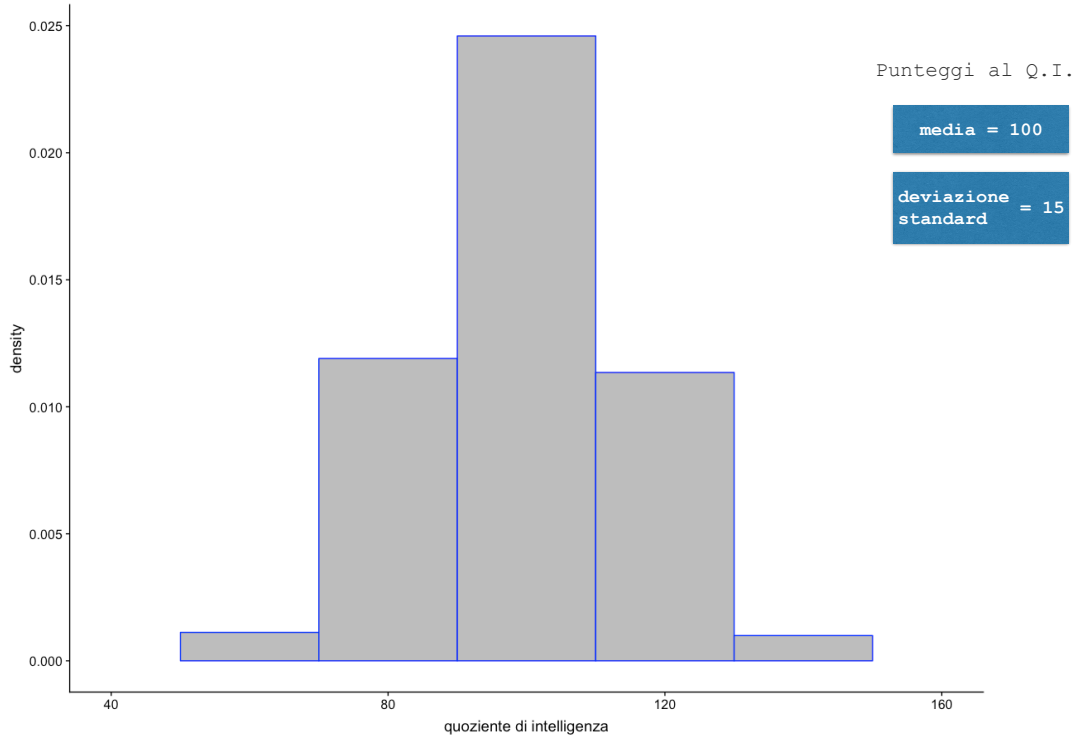
Si riparte?

Normalità...

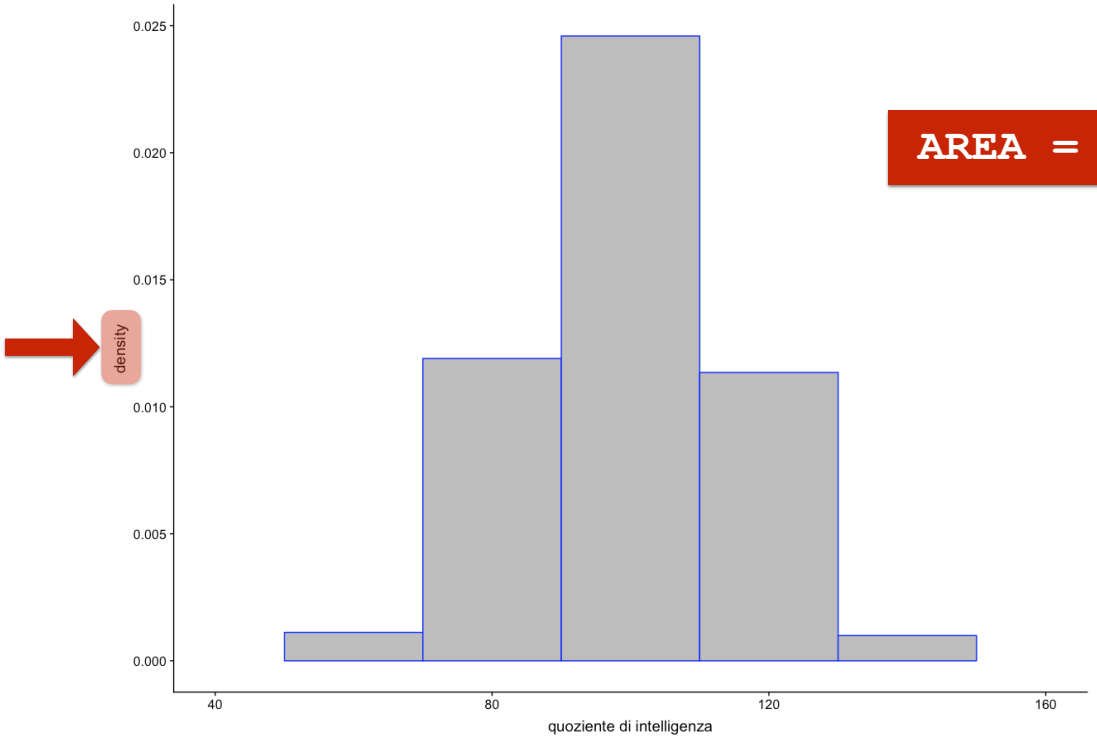
Distribuzione (curva) normale



Distribuzione normale: esempio

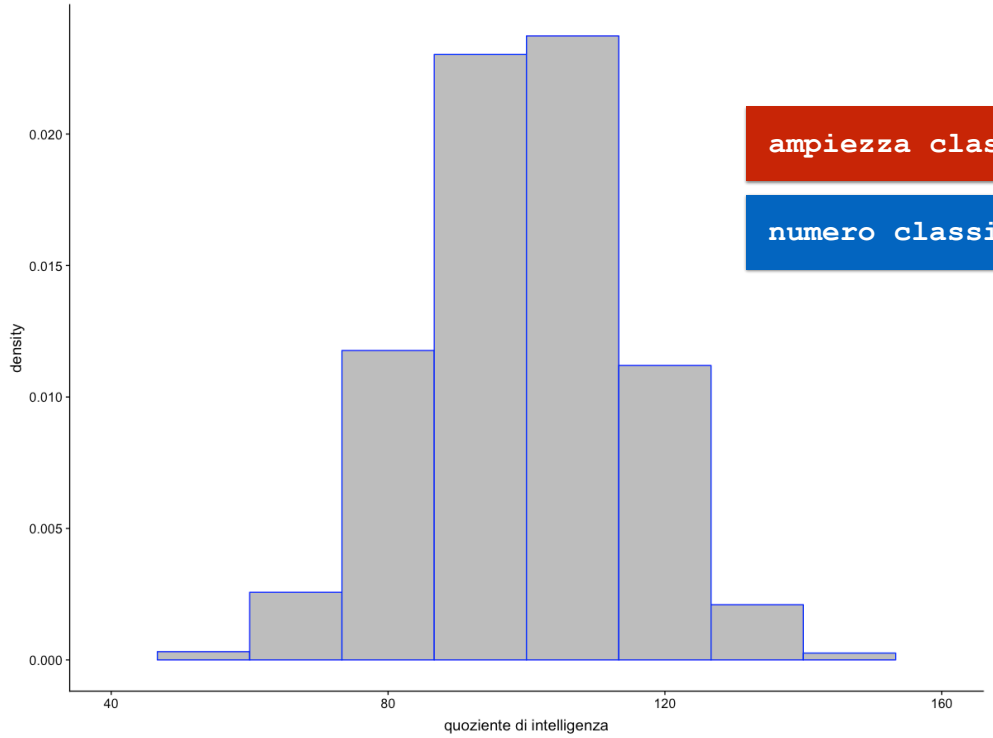


Distribuzione normale: esempio



AREA = ?

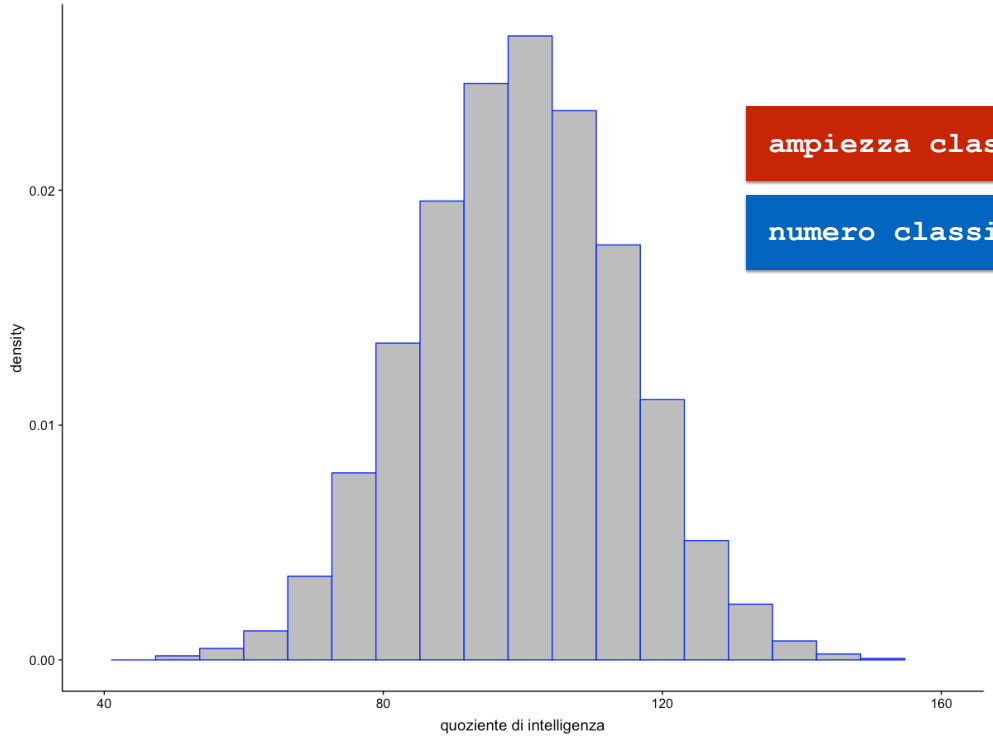
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +OO

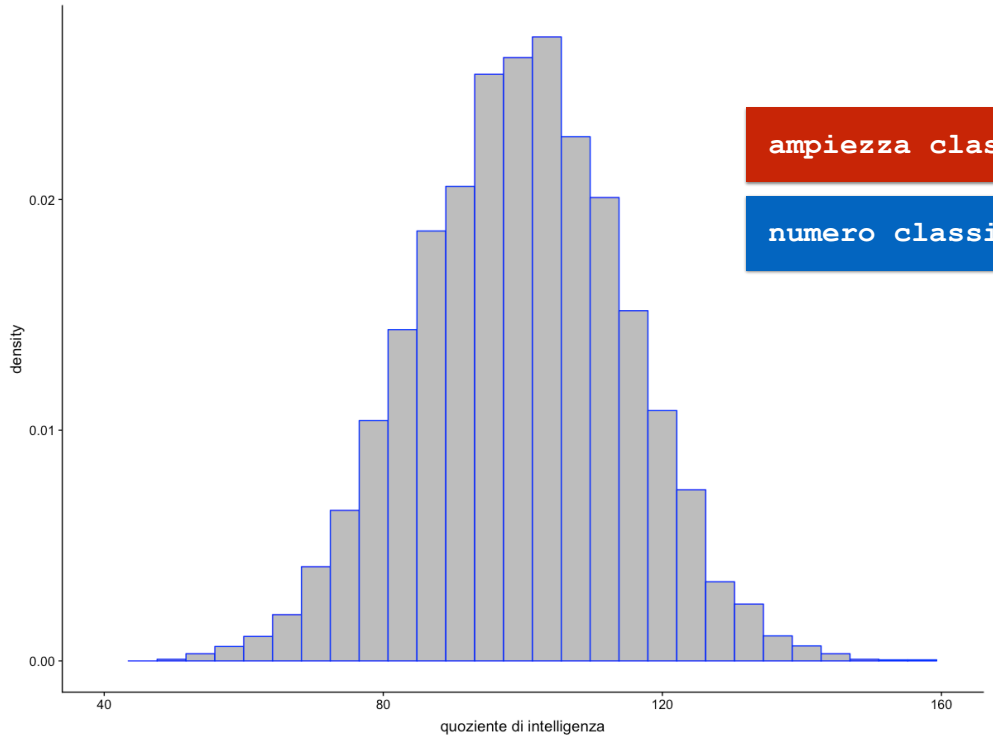
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +OO

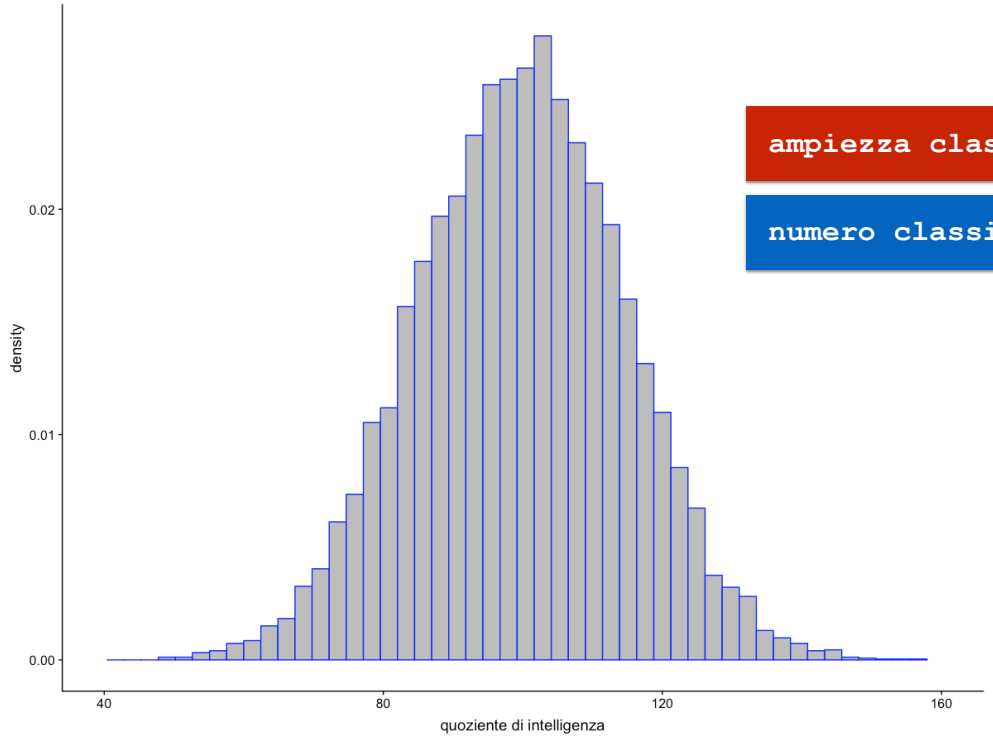
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +OO

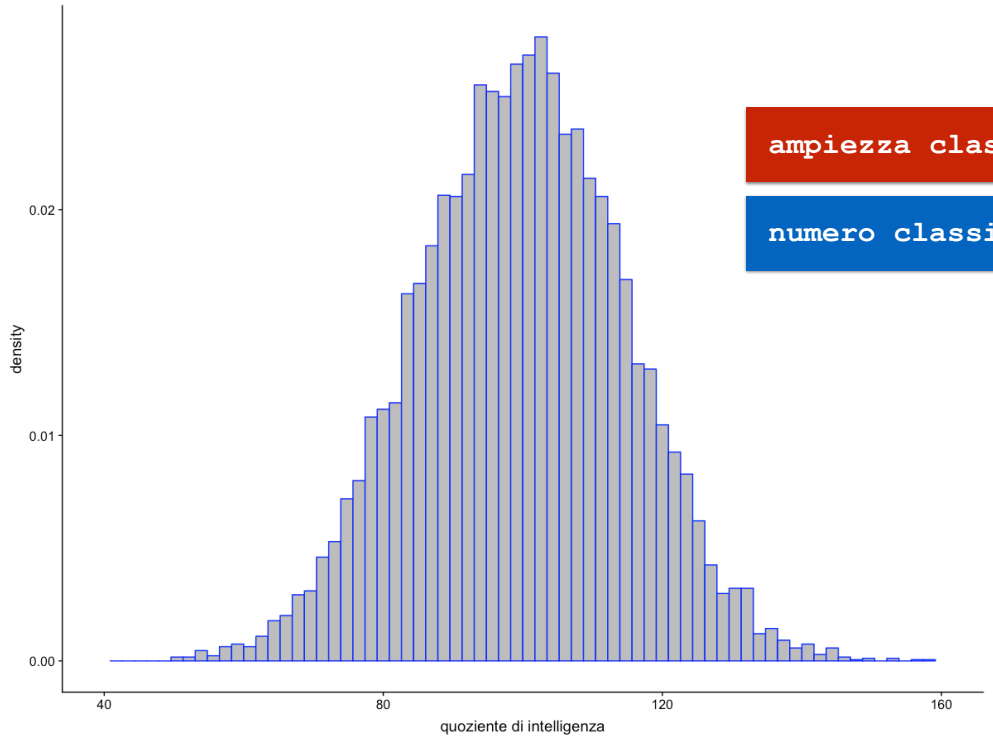
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +∞

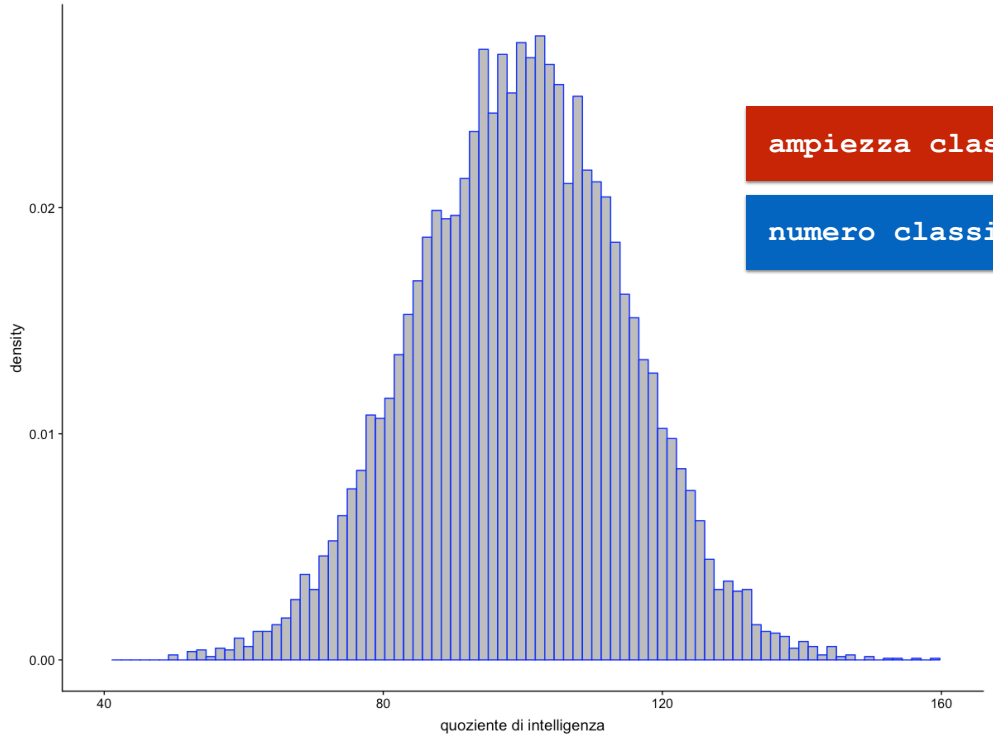
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +∞

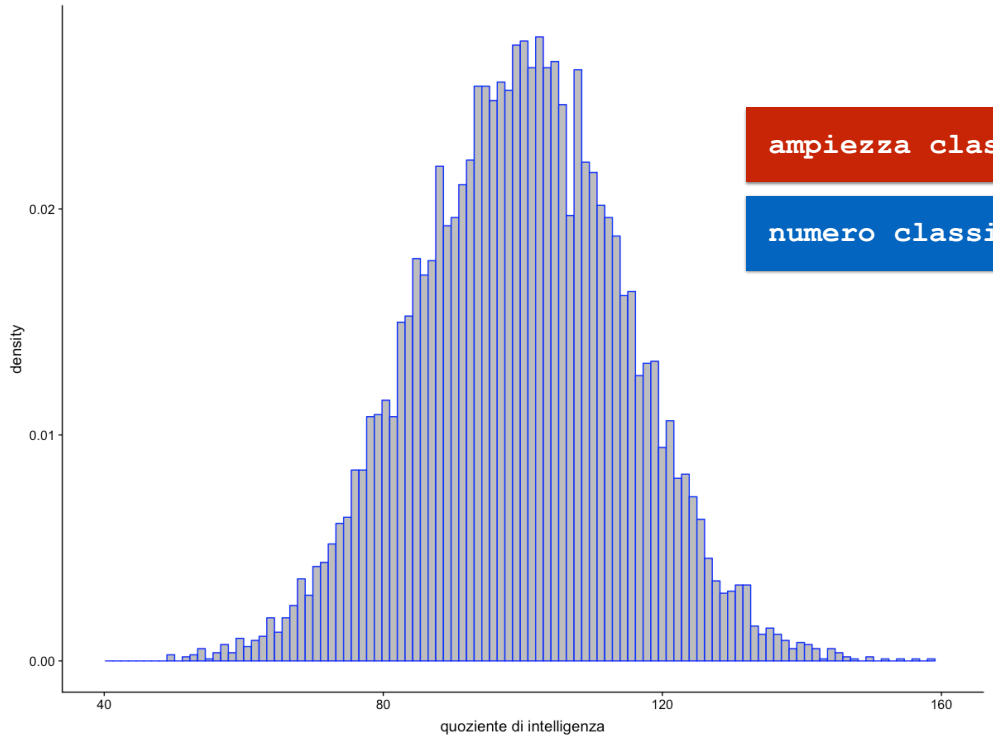
Distribuzione normale: esempio



ampiezza classi -> 0

numero classi -> +∞

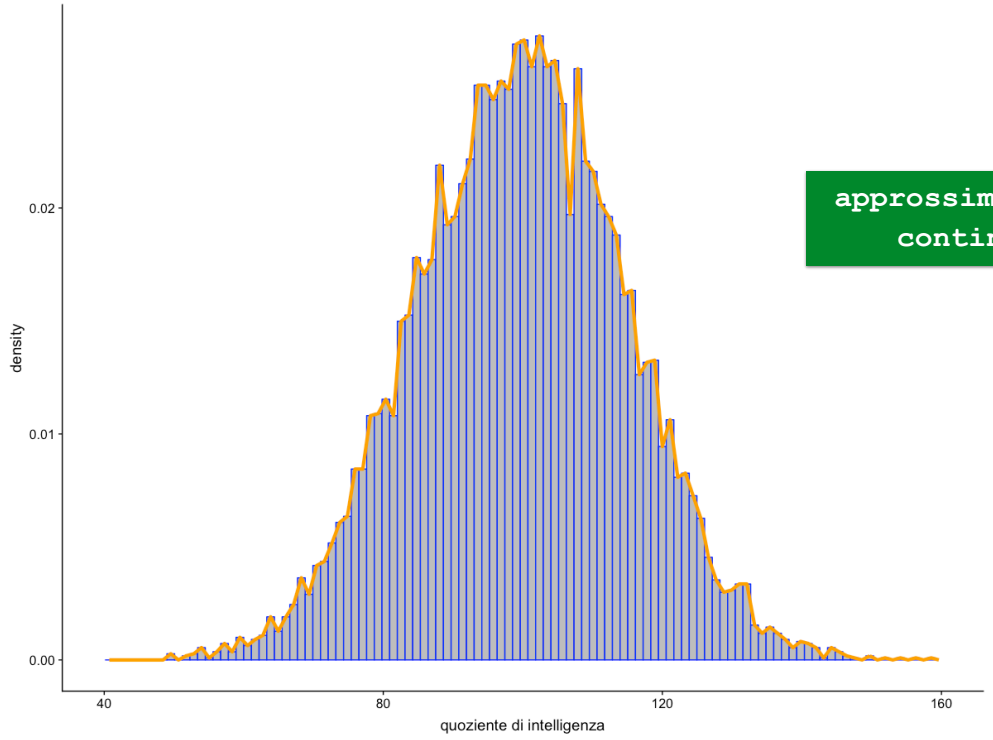
Distribuzione normale: esempio



ampiezza classi -> 0

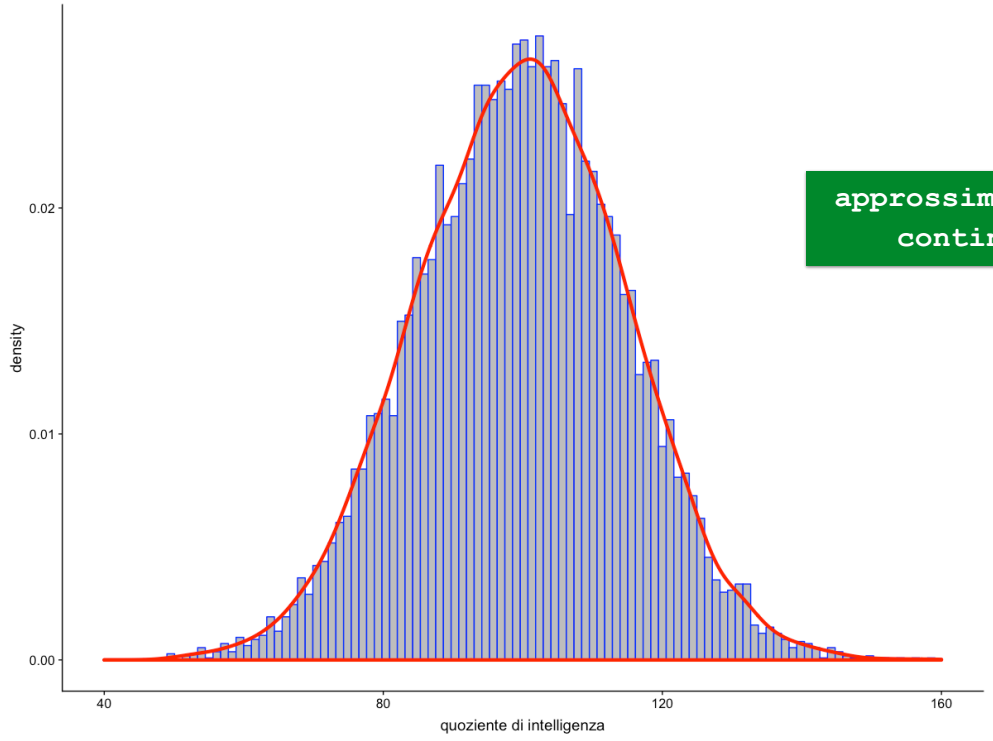
numero classi -> +∞

Distribuzione normale: esempio



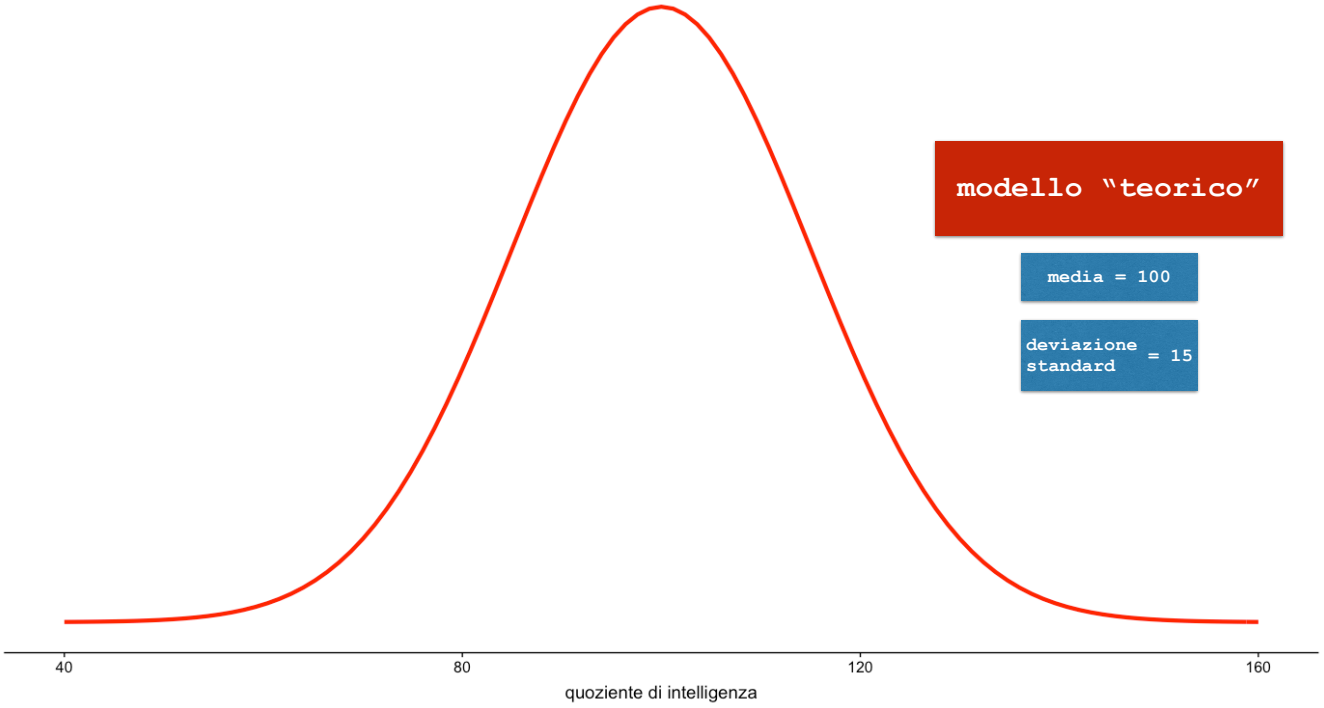
approssimazione continua

Distribuzione normale: esempio



approssimazione continua

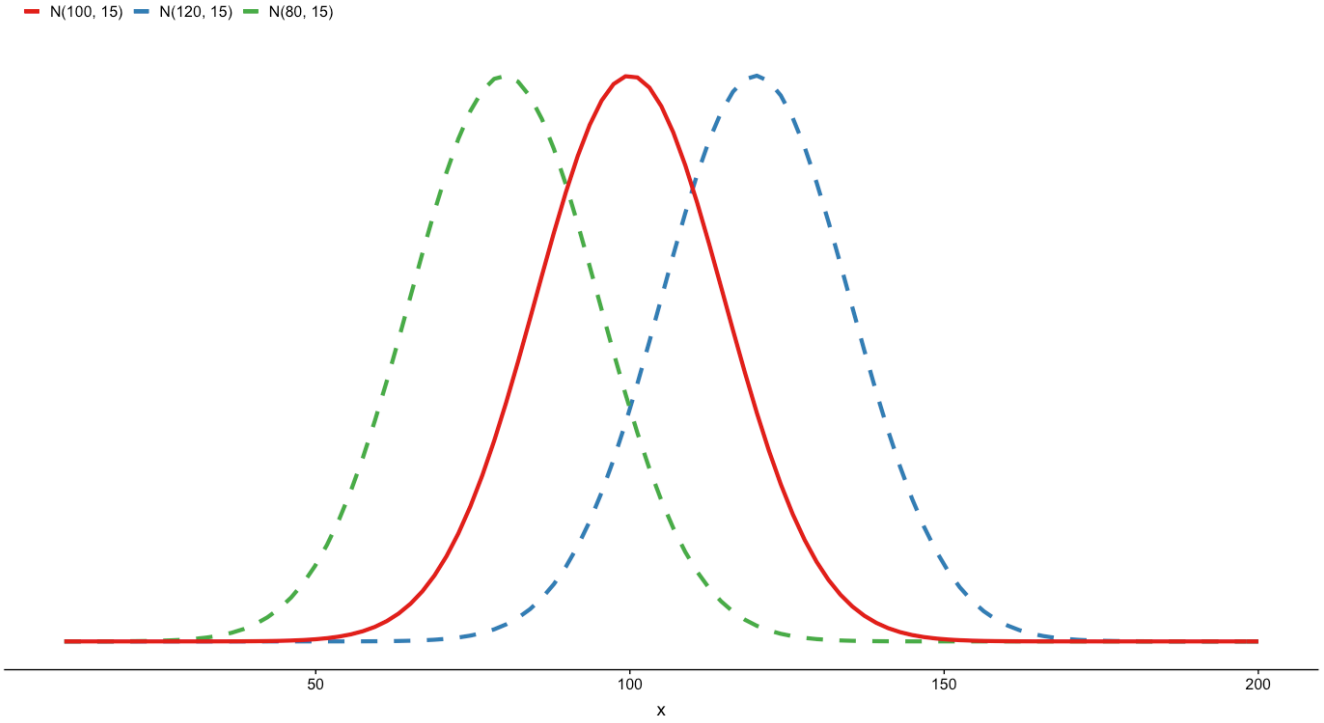
Distribuzione normale: esempio



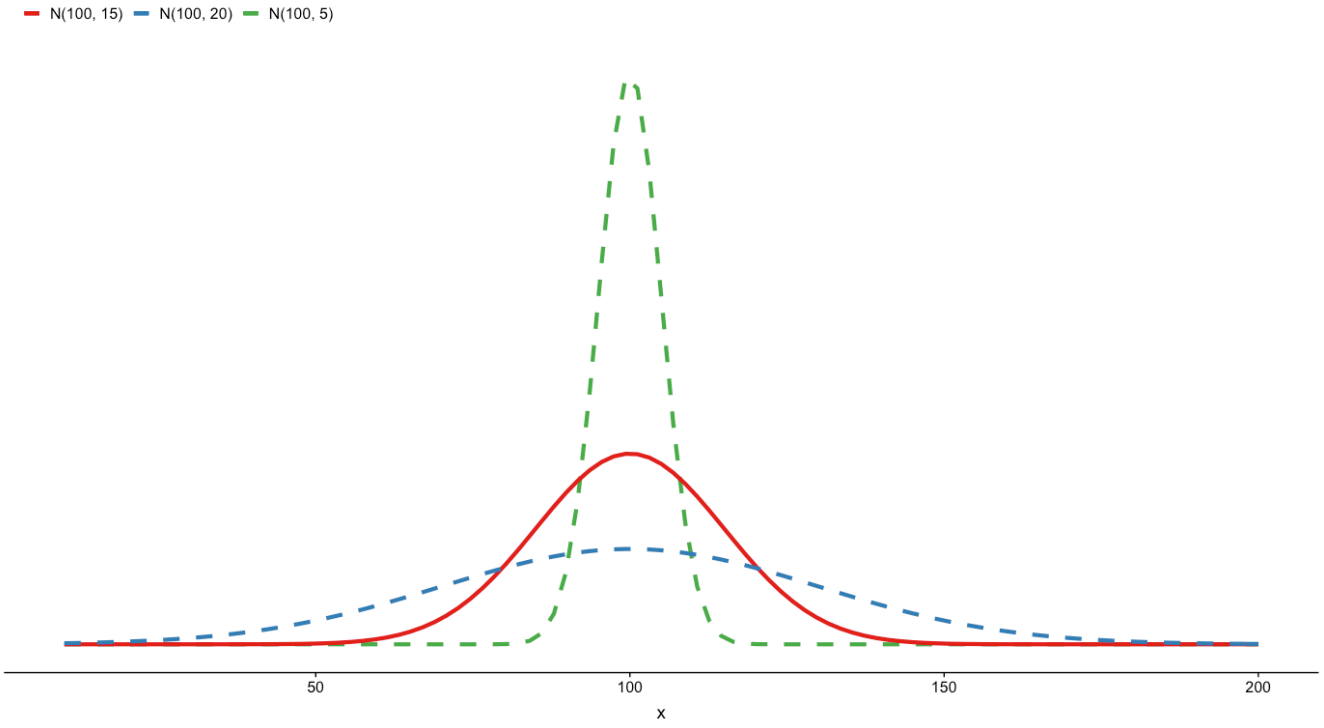
modello "teorico"

- media = 100
- deviazione standard = 15

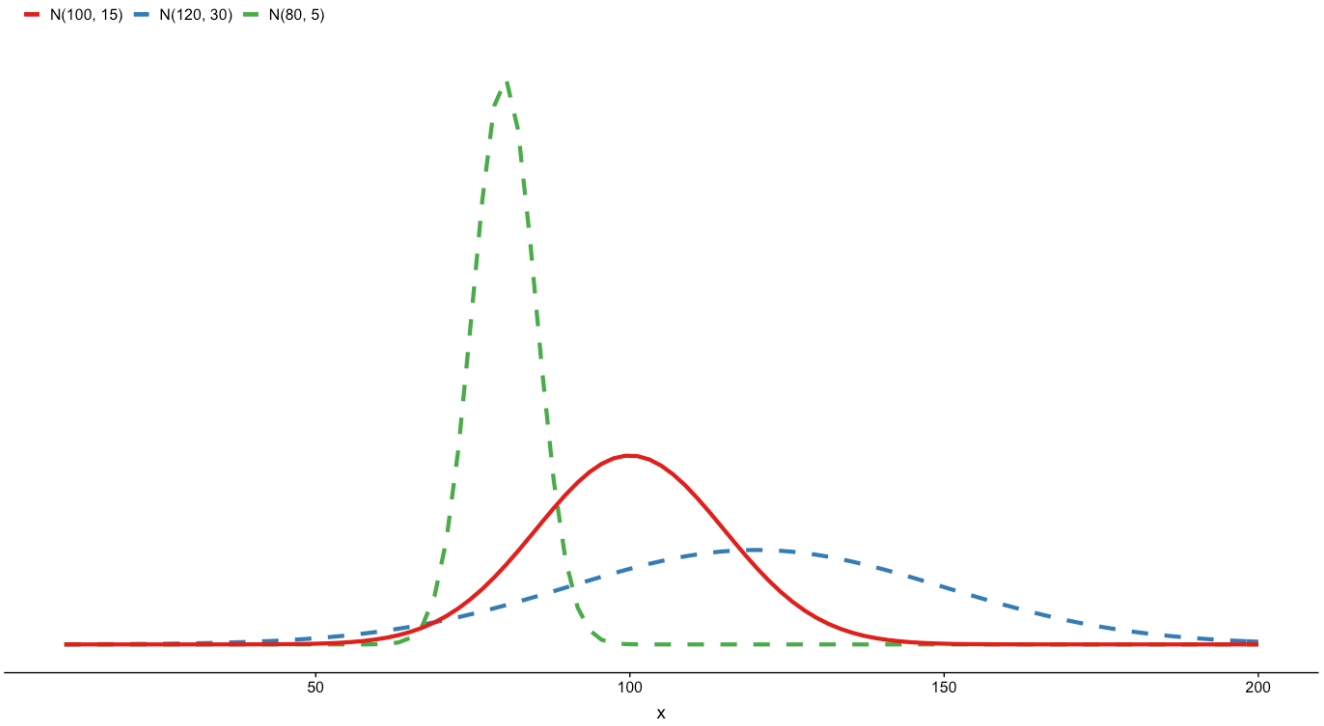
Distribuzione normale: cambia la media



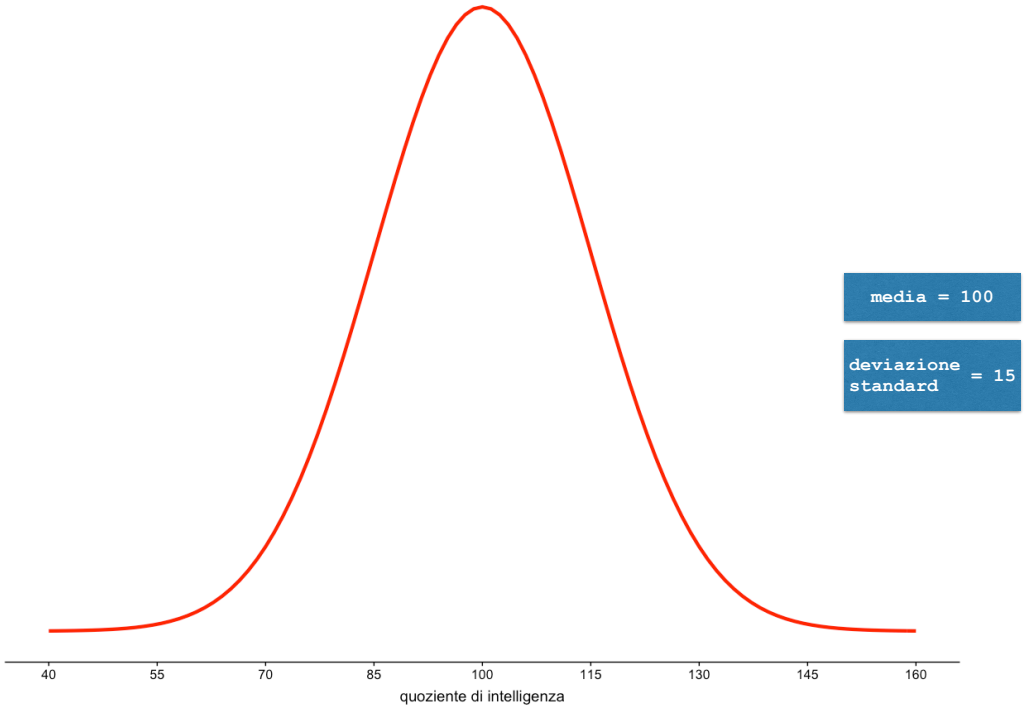
Distribuzione normale: cambia la varianza



Distribuzione normale: cambiano media e varianza



Distribuzione normale: esempio



Dettagli "pratici"…

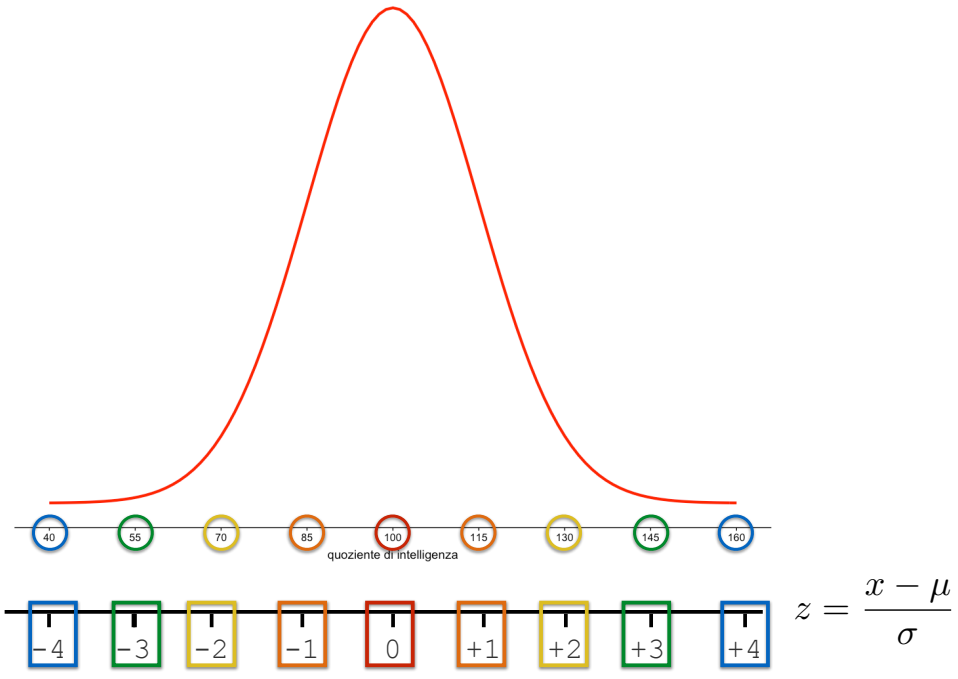
Distribuzione normale: proprietà (1)

Qualunque trasformazione lineare di una variabile normale è ancora normale

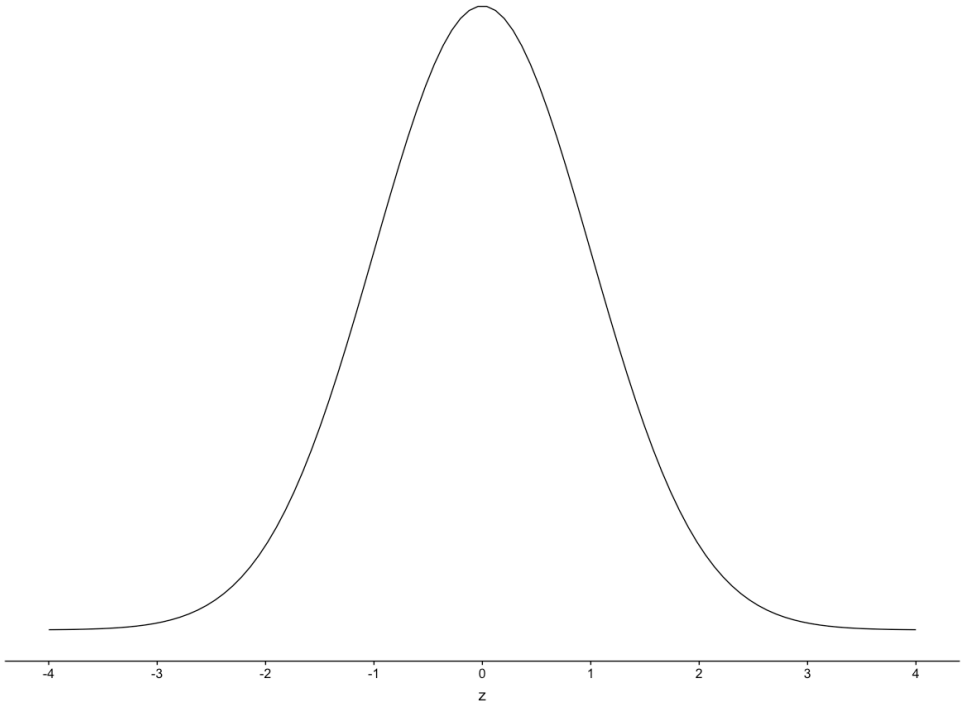
$$X \sim N(\mu_X, \sigma_X)$$

$$z = \frac{x - \mu_X}{\sigma_X} \sim N(\mu_Z = 0, \sigma_Z^2 = 1)$$

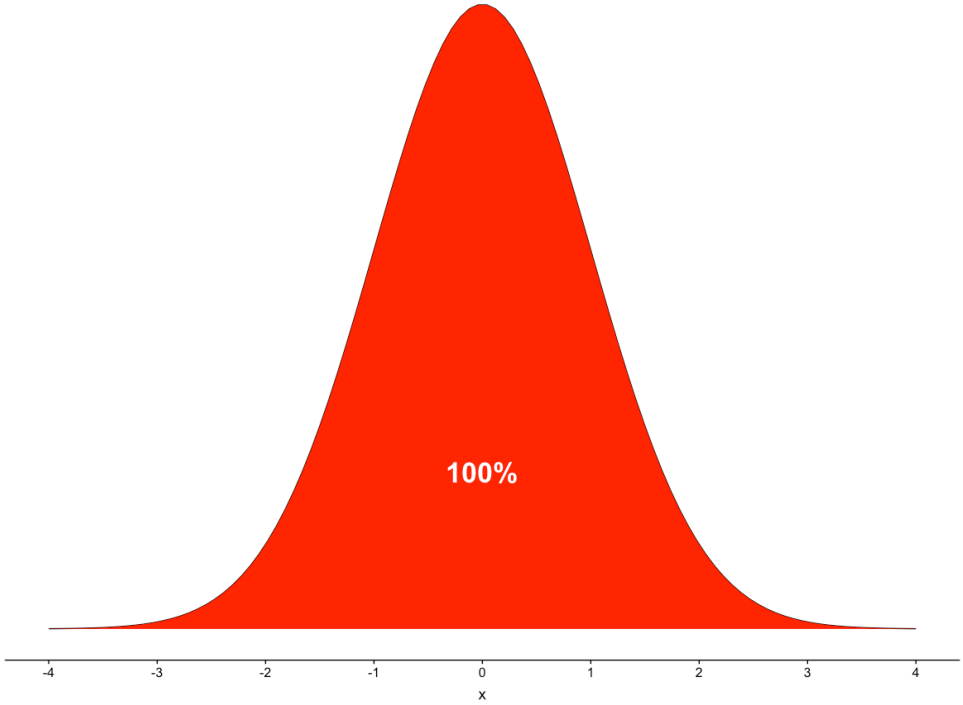
Distribuzione normale: esempio



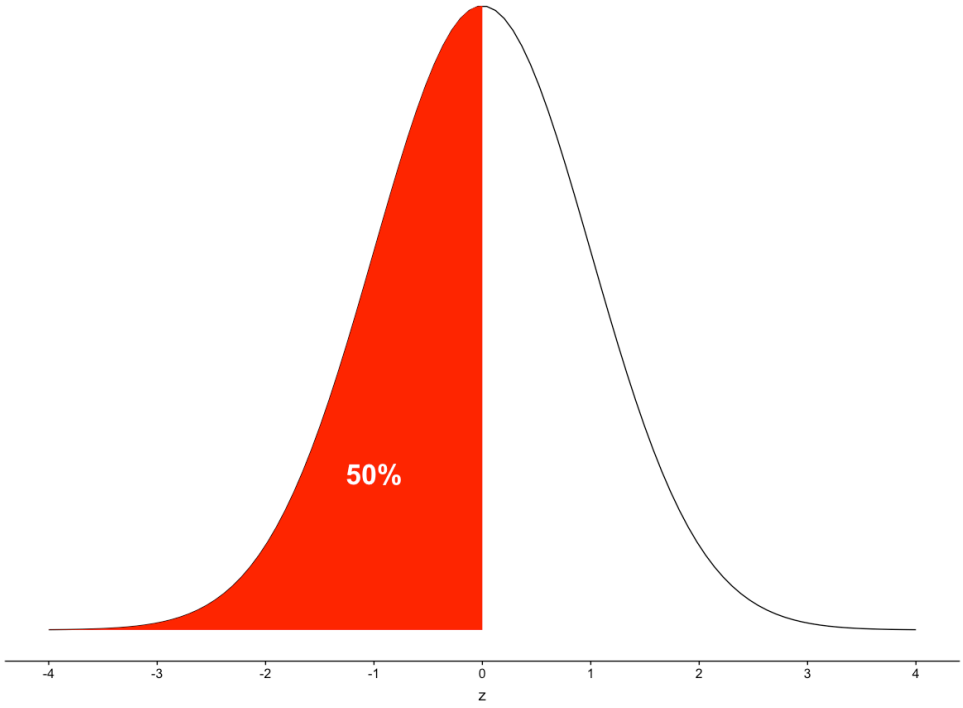
Distribuzione normale: caratteristiche



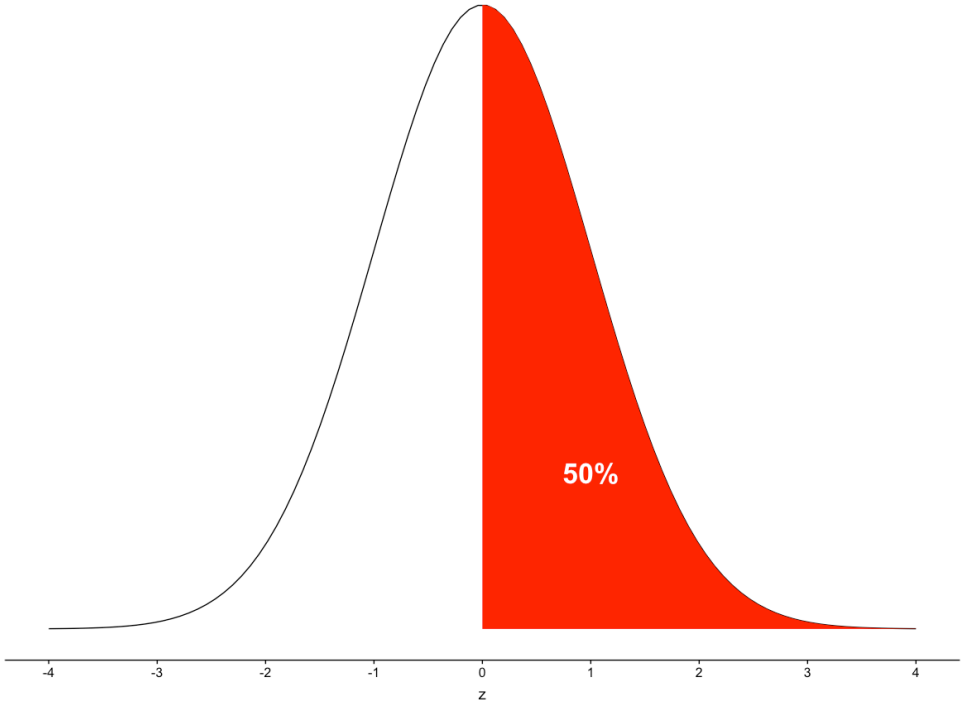
Distribuzione normale: caratteristiche



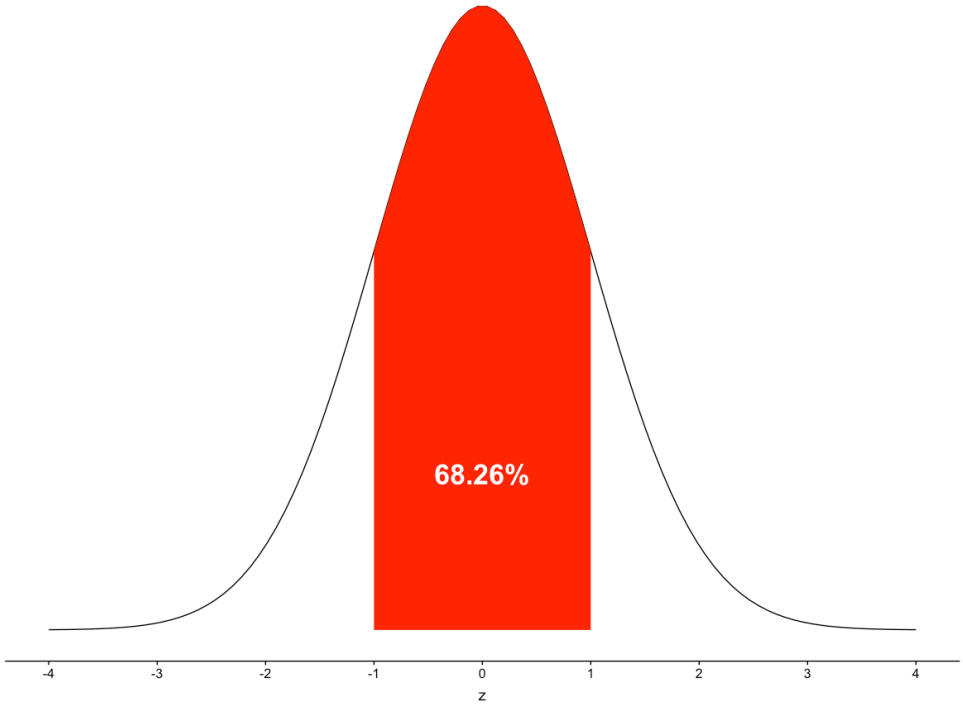
Distribuzione normale: caratteristiche



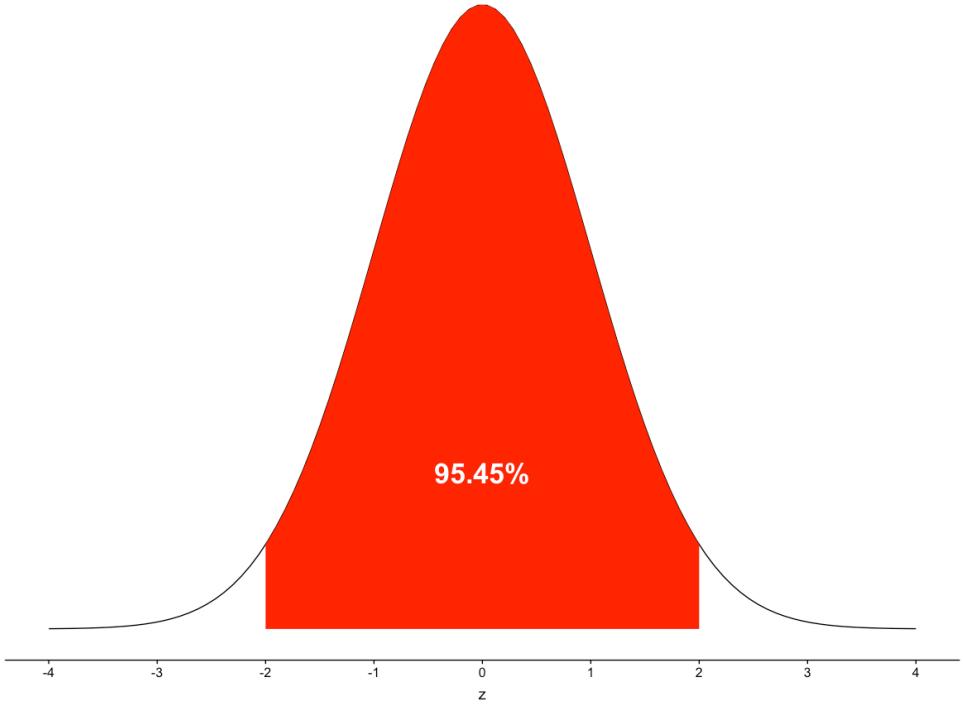
Distribuzione normale: caratteristiche



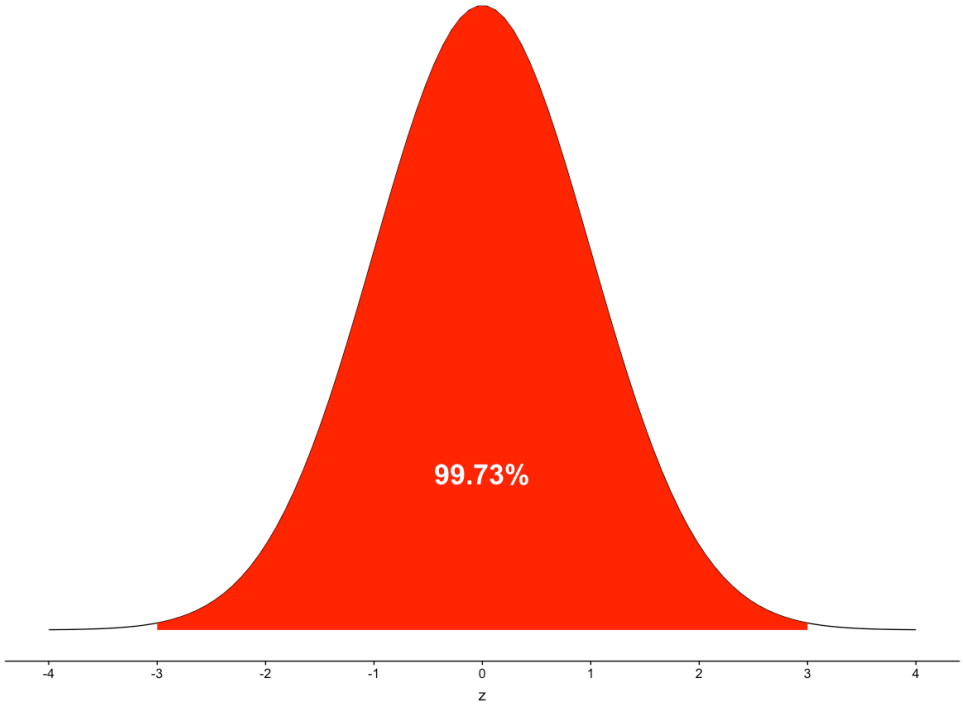
Distribuzione normale: caratteristiche



Distribuzione normale: caratteristiche



Distribuzione normale: caratteristiche



Dettagli tecnici…

Distribuzione normale: tecnicismi

Legge normale → $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Costanti:

2

no comment

$\pi = 3.14$

nella geometria piana viene definito come il rapporto tra la lunghezza della circonferenza e quella del suo diametro, o anche come l'area di un cerchio di raggio 1

$e = 2.72$

costante matematica il cui valore è approssimativamente 2,7182818284…
È la base della funzione esponenziale e del logaritmo naturale

$$e = \lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^n$$

$$e = \sum_{n=0}^{+\infty}\frac{1}{n!}$$

Distribuzione normale: tecnicismi

Legge normale → $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Parametri:

μ

media

(mediana)

(moda)

σ^2

varianza

Distribuzione normale: tecnicismi

Legge normale → $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Legge normale standard → $\frac{1}{2\pi} e^{-\frac{1}{2}z^2}$

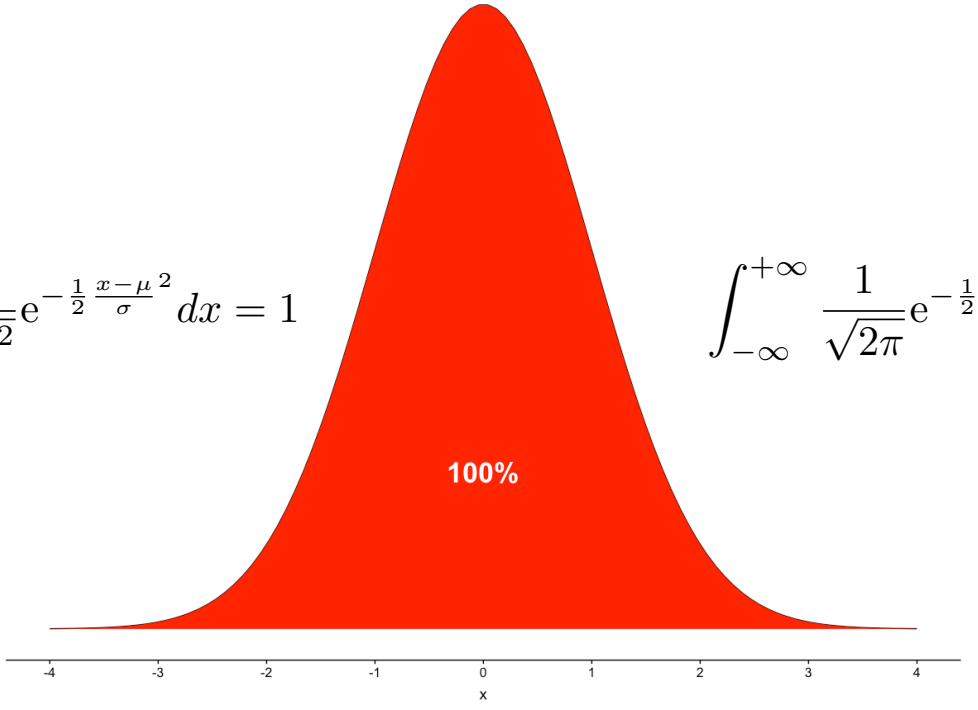
Parametri Z:

$\mu_Z = 0$ | $\sigma_Z^2 = \sigma_Z = 1$

Distribuzione normale: tecnicismi

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x-\mu}{\sigma}^2} dx = 1$$

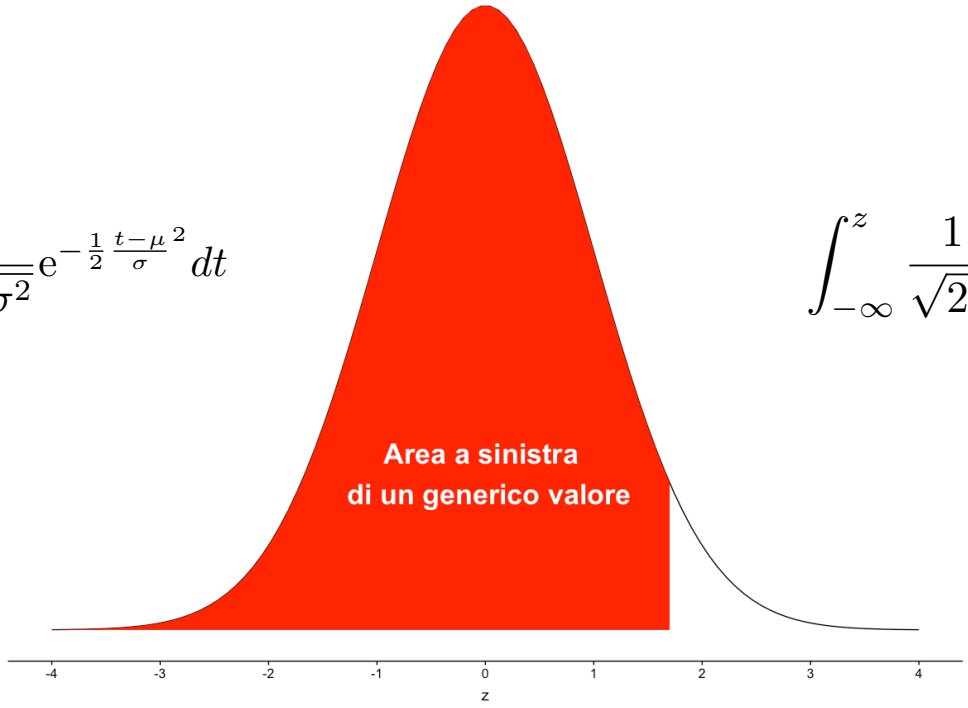
$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$$



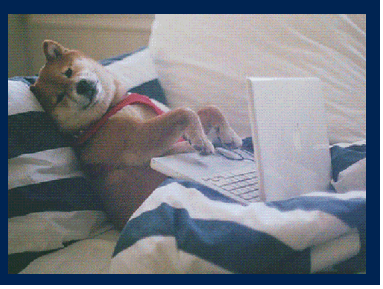
Distribuzione normale: tecnicismi

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2}\frac{t-\mu}{\sigma}^2} dt$$

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}t^2} dt$$



Al lavoro...



Distribuzione normale: tavole

Restituiscono, per i valori di z tra -3.9 = -3.90 e +3.99:

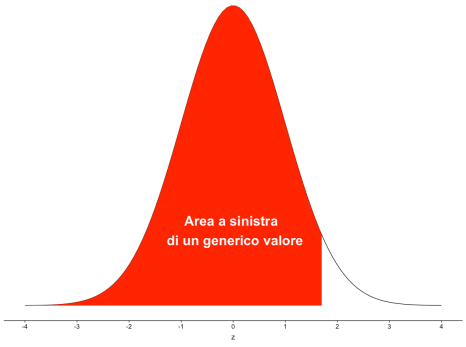
$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}t^2} dt$$

area a sinistra della z di interesse (funzione di ripartizione)

NOTA

- Alcune tavole contengono solo la coda destra (da -3.90 a 0.00)
- Altre tavole contengono l'area tra 0 e la z di interesse
- Altre ancora contengono sia l'area a sinistra che quella a destra della z di interesse

Area a sinistra di un valore z



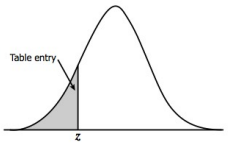


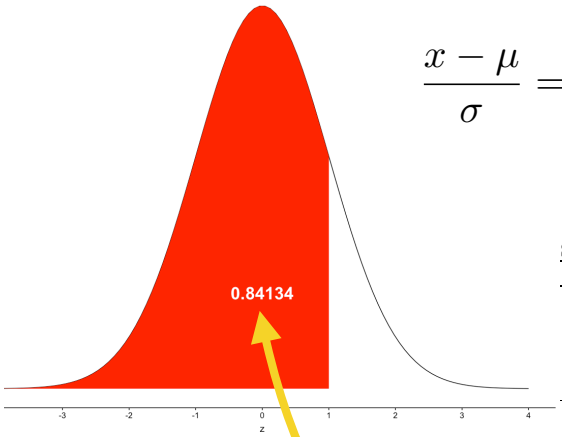
Table entry for z is the area under the standard normal curve to the left of z.

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
.	.	.	.	.	.	.	.	.	.	.
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Area a sinistra di un valore z

Percentuale di rispondenti con un punteggio minore di 115?



$$\frac{x-\mu}{\sigma} = \frac{115-100}{15} = +1 = +1.00$$

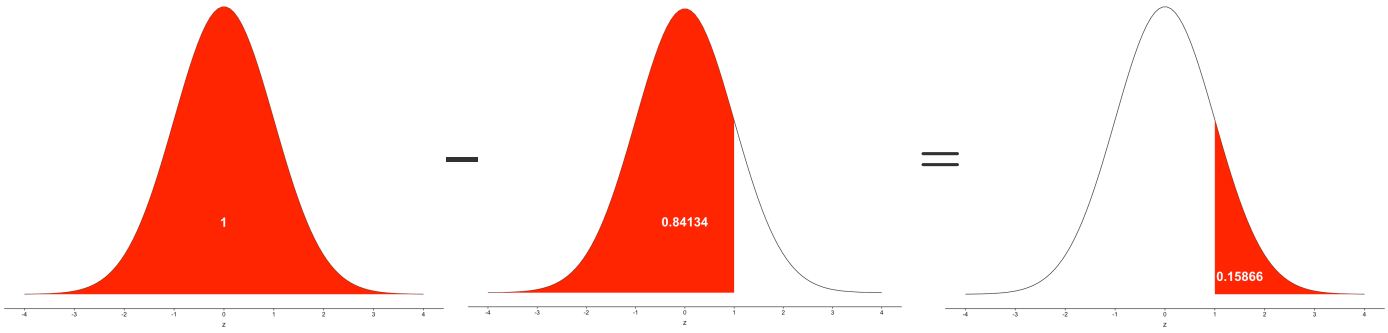
STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298

Area a destra di un valore z

Percentuale di rispondenti con un punteggio maggiore di 115?

$$\frac{x-\mu}{\sigma} = \frac{115-100}{15} = +1 = +1.00$$

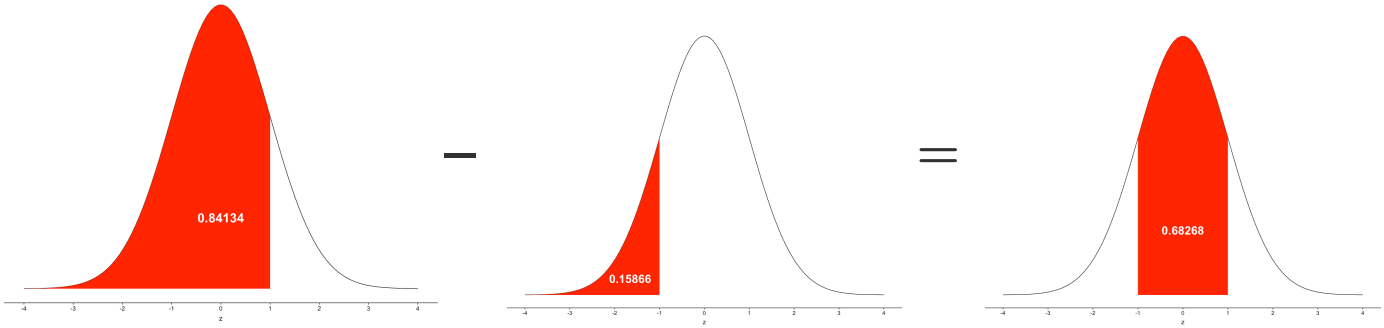


Area tra due valori z_1 e z_2

Percentuale di rispondenti con un punteggio tra 85 e 115

$$\frac{x-\mu}{\sigma} = \frac{85-100}{15} = -1 = -1.00$$

$$\frac{x-\mu}{\sigma} = \frac{115-100}{15} = +1 = +1.00$$

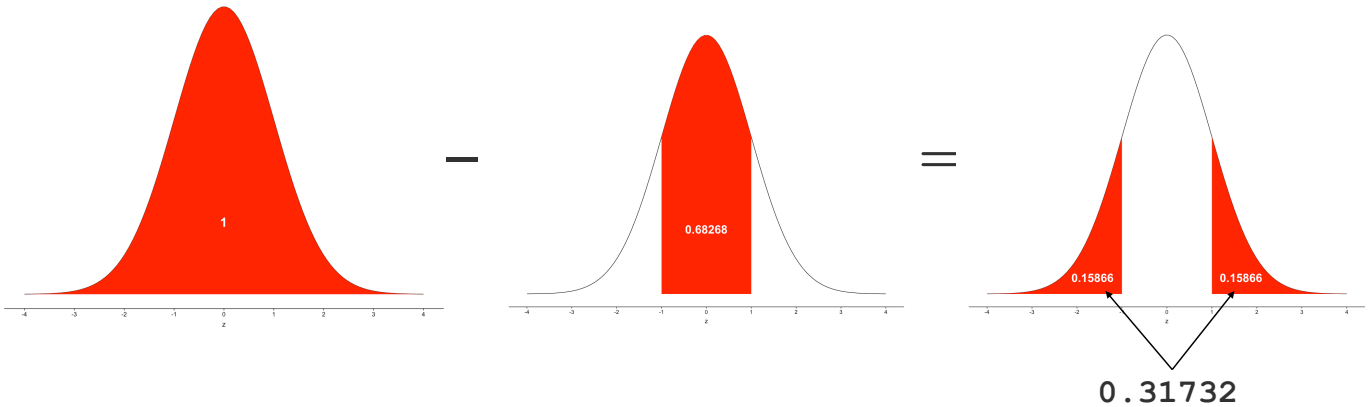


Area esterna a due valori z_1 e z_2

Percentuale di rispondenti con un punteggio minore di 85 o maggiori di 115

$$\frac{x-\mu}{\sigma} = \frac{85-100}{15} = -1 = -1.00$$

$$\frac{x-\mu}{\sigma} = \frac{115-100}{15} = +1 = +1.00$$



Casi particolari…

Casi particolari (z non disponibili)

Percentuale di rispondenti con un punteggio minore di 35

$$\frac{x-\mu}{\sigma} = \frac{35-100}{15} = -4.33$$ Area circa 0

Percentuale di rispondenti con un punteggio maggiore di 35

$$\frac{x-\mu}{\sigma} = \frac{35-100}{15} = -4.33$$ Area circa 1

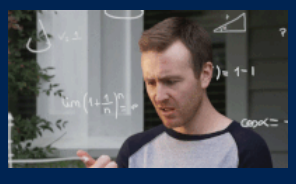
Percentuale di rispondenti con un punteggio minore di 160

$$\frac{x-\mu}{\sigma} = \frac{160-100}{15} = +4.00$$ Area circa 1

Percentuale di rispondenti con un punteggio maggiore di 160

$$\frac{x-\mu}{\sigma} = \frac{160-100}{15} = +4.00$$ Area circa 0

Cambiamo punto di vista?



Problema inverso: i quantili

Punteggio superato solo dal 10% dei rispondenti → 90° percentile

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189

90° percentile z è circa 1.285

$$z = \frac{x-\mu}{\sigma} = 1.285 = \frac{x-100}{15} \implies x = \mu + z \times \sigma = 100 + 1.285 \times 15 = 119.275$$

Problema inverso: i quantili

Punteggio superato dal 25% dei rispondenti

primo quartile

sulla Z è circa -0.657

$$z = \frac{x - \mu}{\sigma} = -0.675$$

$$\frac{x - 100}{15} = -0.675$$

$$q_1(x) = \mu + z \times \sigma = 100 - 0.675 \times 15 = 89.875$$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760

Dettagli "pratici"…

Distribuzione normale: proprietà (2)

Qualunque combinazione lineare di variabili normali è ancora normale

$$X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \ldots, n$$

$$W = \sum_{i=1}^{n} a_i X_i \sim N(\mu_W, \sigma_W)$$

proprietà riproduttiva

Distribuzione normale: proprietà (2)

Qualunque combinazione lineare di variabili normali è ancora normale

$$X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \ldots, n$$

$$W = \sum_{i=1}^{n} a_i X_i \sim N(\mu_W, \sigma_W)$$

pesi della combinazione lineare

somma $\quad a_i = 1 \qquad \forall i = 1, \ldots, n$

media $\quad a_i = \frac{1}{n} \qquad \forall i = 1, \ldots, n$

proprietà riproduttiva

Distribuzione normale: proprietà (2)

Qualunque combinazione lineare di variabili normali è ancora normale

$$X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \ldots, n$$

$$W = \sum_{i=1}^{n} a_i \, X_i \sim N(\mu_W, \sigma_W)$$

$$\sum_{i=1}^{n} a_i \, \mu_i$$

proprietà riproduttiva

Distribuzione normale: proprietà (2)

Qualunque combinazione lineare di variabili normali è ancora normale

$$X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \ldots, n$$

$$W = \sum_{i=1}^{n} a_i \, X_i \sim N(\mu_W, \sigma_W)$$

$$\sum_{i=1}^{n} a_i \, \mu_i$$

proprietà riproduttiva

somma $\sum_{i=1}^{n} \mu_i$

media $\frac{1}{n} \sum_{i=1}^{n} \mu_i$

Distribuzione normale: proprietà (2)

Qualunque combinazione lineare di variabili normali è ancora normale

$$X_i \sim N(\mu_i, \sigma_i) \qquad i = 1, \ldots, n$$

$$W = \sum_{i=1}^{n} a_i X_i \sim N(\mu_W, \sigma_W)$$

somma: $\sum_{i=1}^{n} \sigma_i^2$

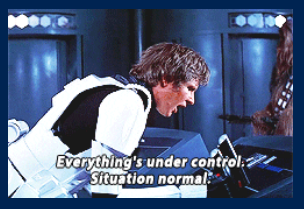
media: $\frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2$

$$\sum_{i=1}^{n} a_i^2 \sigma_i^2$$

solo se le Xi sono indipendenti

proprietà riproduttiva

Normalità?

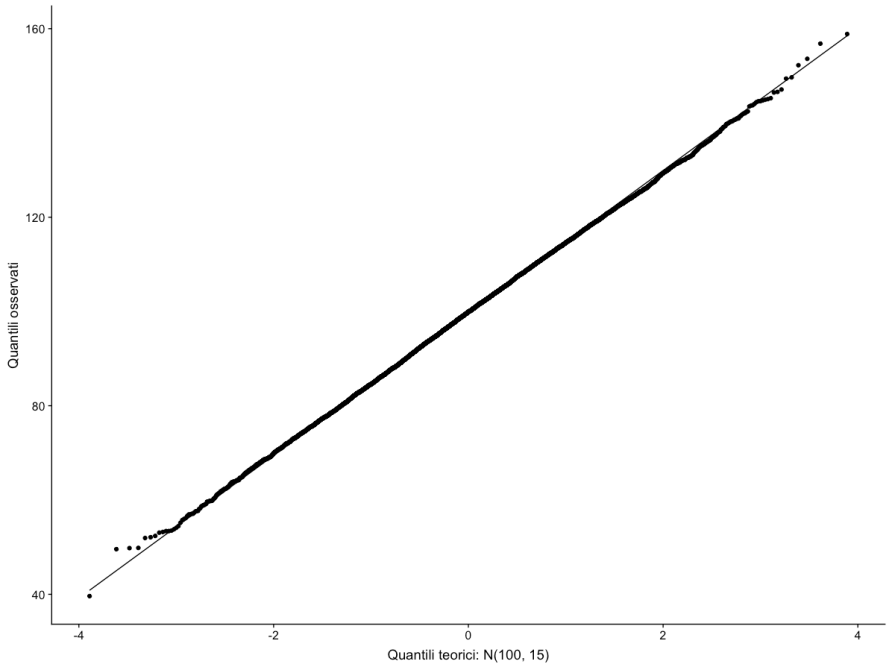


Regole empiriche per la normalità

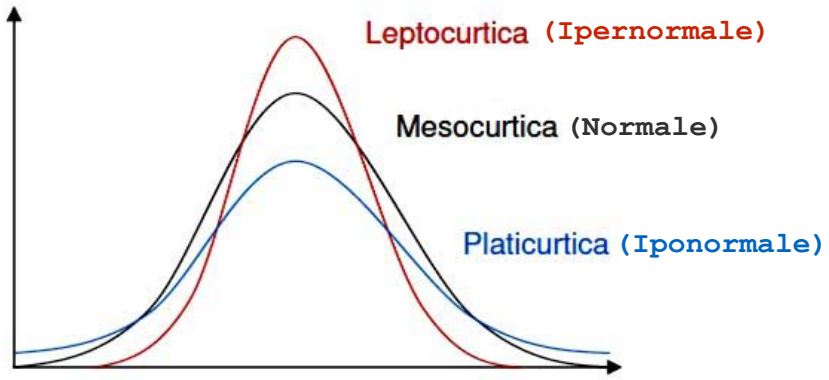
- Confronto tra media, mediana e mediana
- Indice di asimmetria
- Confronto tra la distanze dei due quartili opposti dalla mediana
- Verifica delle aree in intervalli simmetrici attorno alla media (calcolo dei corrispondenti quantili)
- Andamento del boxplot
- Andamento dell'istogramma ma soprattutto del Q-Q plot (vedi pagina seguente)
- Indice di curtosi (vedi due pagine dopo)
- Indice di adattamento (chi-quadro e dintorni: vedi esercizio numerico)

Q-Q plot

I grafici quantili-quantili mettono a confronto i quantili di due distribuzioni



C(K)urtosi



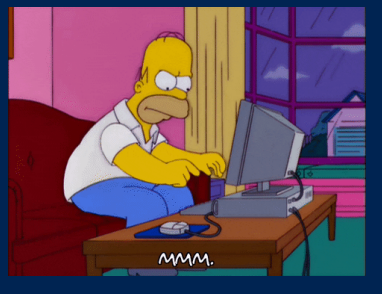
$$\sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right)^4$$

da confrontare con il valore 3

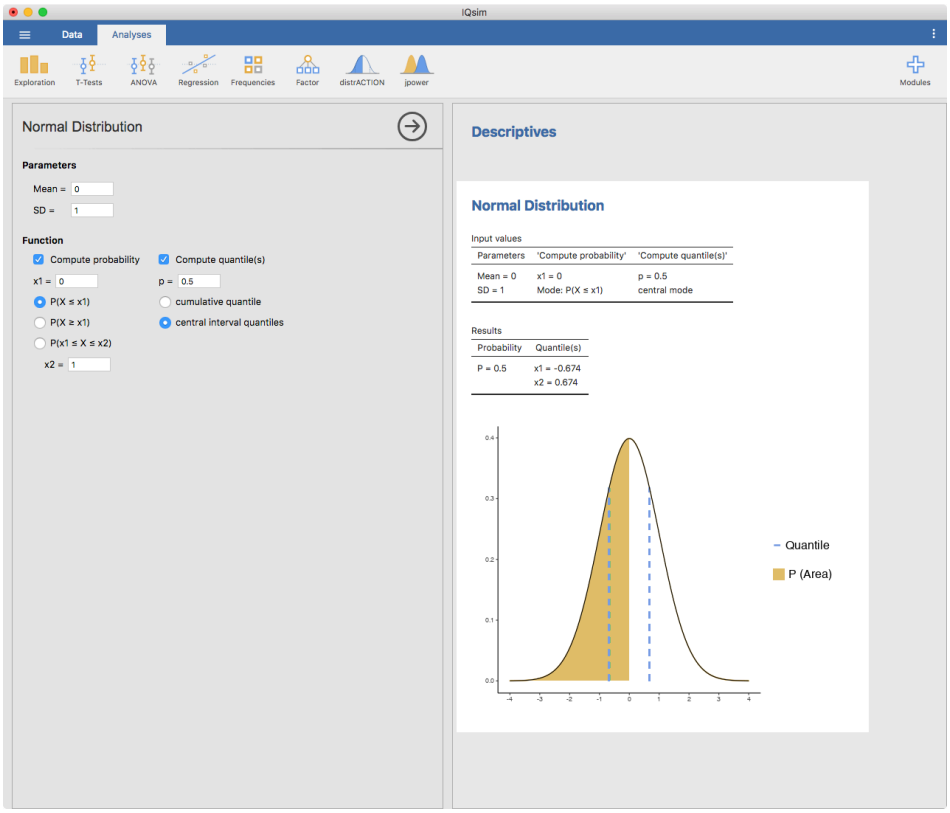
$$\sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right)^4 - 3$$

da confrontare con il valore 0

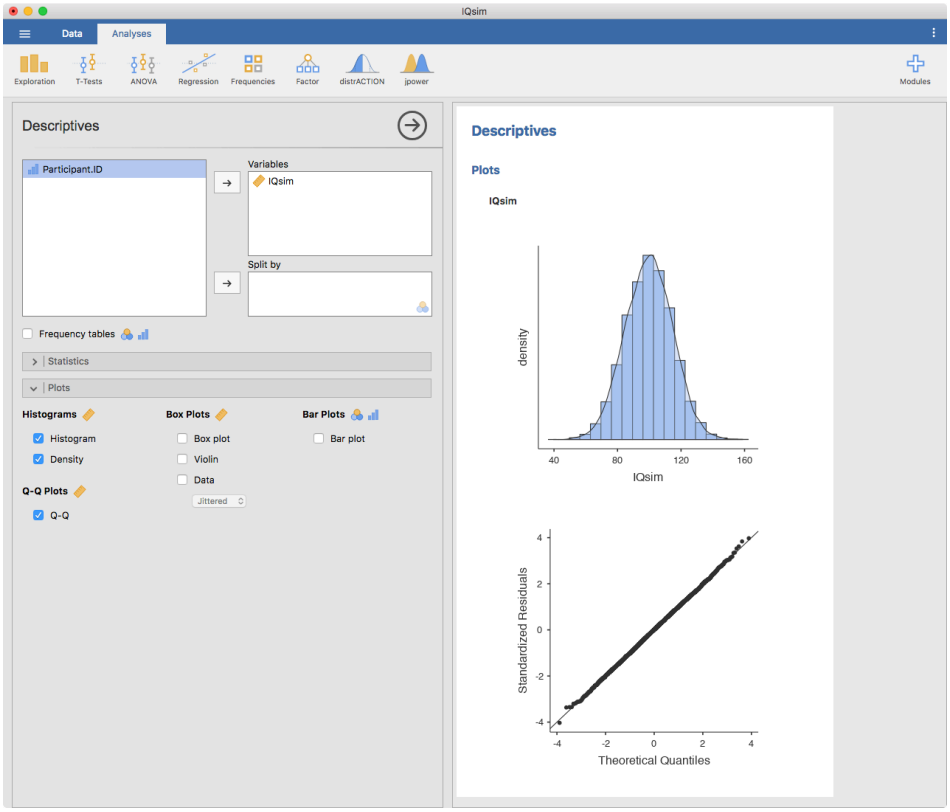
Al pc...



DistrACTION



Il caso dei dati sul QI



Il caso dei dati sul QI