Variabili aleatorie continue: la normale

Giovanni M. Marchetti

Statistica Capitolo 6 — Corso di Laurea in Economia 2015-16

Distinzione

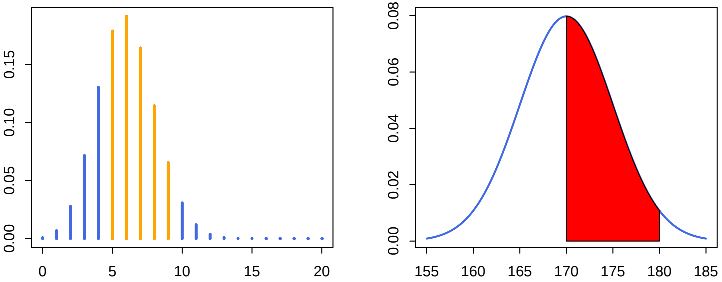
Le variabili aleatorie possono essere

1. discrete
2. continue

Variabili continue

- Misure come consumi, fatturati, rendimenti, durate etc. vengono pensate come grandezze continue, non discrete.
- Una variabile continua ha come modalità teoriche i numeri reali $\mathbb{R}$
- La probabilità è spalmata su $\mathbb{R}$ e non assegnata a singole modalità discrete.

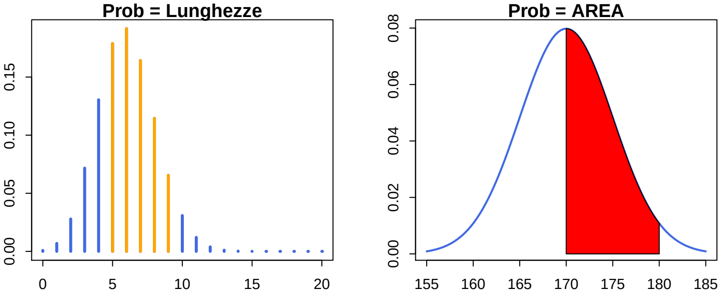
Confronta



Ricordate la differenza tra diagrammi a barre e istogrammi

Densità di probabilità

- Una variabile aleatoria discreta X definita dalla funzione di massa di probabilità che assegna ad ogni x una probabilità $p(x)$
- Una variabile aleatoria continua X è definita dalla sua funzione di densità di probabilità



Densità di probabilità

Una densità di probabilità di una variabile continua X è una funzione $f(x)$ tale che

1. $f(x) \geq 0$ Tutta sopra la linea di orizzonte
2. per ogni intervallo di estremi a e b la probabilità che X cada in $[a, b]$ è l'area sotto la densità tra a e b:

$$\Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

3. In particolare l'area complessiva sotto $f(x)$ è 1:

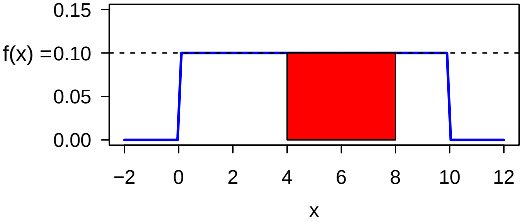
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

La v.a. uniforme su un intervallo $[a, b]$ ha una densità costante

$$f(x) = \frac{1}{b-a} \qquad x \in [a, b]$$

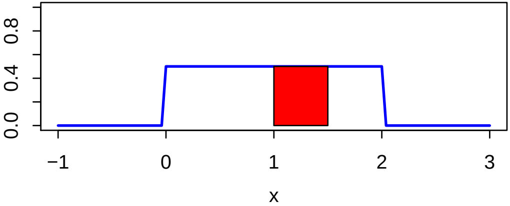
e zero altrimenti.

La probabilità che X appartenga a $[x_1, x_2]$ è proporzionale alla lunghezza $x_2 - x_1$.



$$P(4 \leq X \leq 8) = k \cdot 2 = (1/10) \cdot 2 = 0.2$$

Un ingegnere sa che su un oleodotto lungo 2 km possono capitare casualmente delle rotture in un punto qualsiasi X. Se X ha distribuzione uniforme qual è la probabilità che la rottura capiti tra il km 1 e il km 1.5?



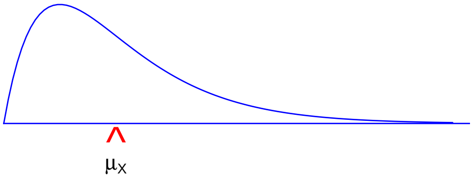
$$P(1 \leq X \leq 1.5) = k \cdot 0.5 = (1/2) \cdot 1/2 = 0.25.$$

Valore atteso

In una qualsiasi variabile continua il valore atteso è il baricentro della superficie sotto la funzione di densità

$$E(X) = \mu_X = \int_{\mathbb{R}} x \cdot f(x) \mathrm{d}x$$

Nota la somiglianza con $E(X) = \sum_x xf(x)$ per le variabili discrete.



Varianza di una variabile continua

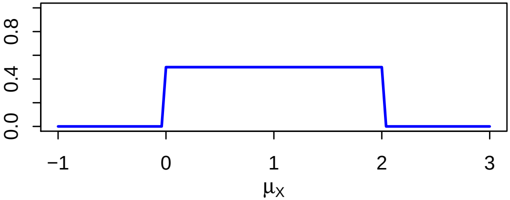
La varianza è

$$\sigma_X^2 = \mathrm{var}(X) = E[(X - \mu_X)^2] = \int_{\mathbb{R}} (x - \mu_X)^2 f(x) \mathrm{d}x$$

Valore atteso e varianza di una variabile uniforme

Se $X \sim \mathcal{U}(a, b)$

$$E(X) = (a + b)/2, \qquad \text{var}(X) = (b - a)^2/12$$

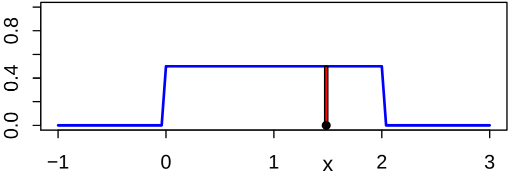


Eventi con probabilità 0

La probabilità che il guasto capiti esattamente in un punto x per definizione è 0

$$\Pr(x \leq X \leq x + h) \to 0$$

Quindi $\Pr(X = x) = 0$ ma non corrisponde a un evento impossibile.



Variabili aleatorie normali

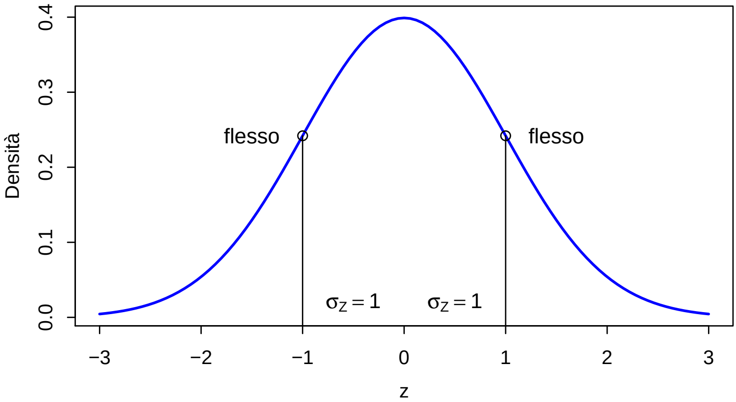
Cominciamo da qualla più semplice: la normale standard.

- Le modalità sono tutti i numeri reali $Z \in (-\infty, +\infty)$
- La densità è

$$f(z) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-z^2/2}$$

  dove $\pi = 3.1415\ldots$ ed $\mathrm{e} = 2.7182\ldots$

Normale standard

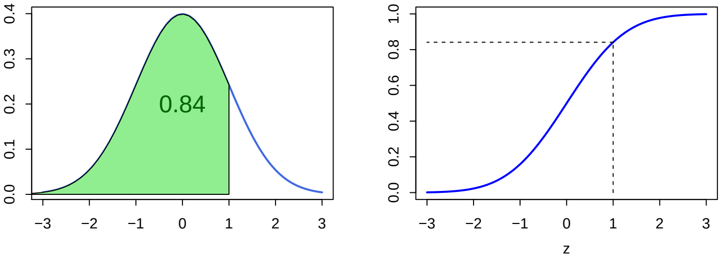


$$\mu_Z = 0, \qquad \sigma_Z = 1$$

Funzione di ripartizione

La funzione di ripartizione della normale standard è

$$\Pr(Z \leq z) = \Pr(-\infty \leq Z \leq z) = \text{area fino a } z$$



$$F(1) = P(Z \leq 1) = 0.84.$$

La funzione di ripartizione è tabulata (Tavola 1 del libro) da $z = 0$ a $z = 3.99$.

```
   z   F(z)
 0.00 0.5000
 0.01 0.5040
 0.02 0.5080
 0.03 0.5120
 0.04 0.5160
 0.05 0.5199
```

⋮

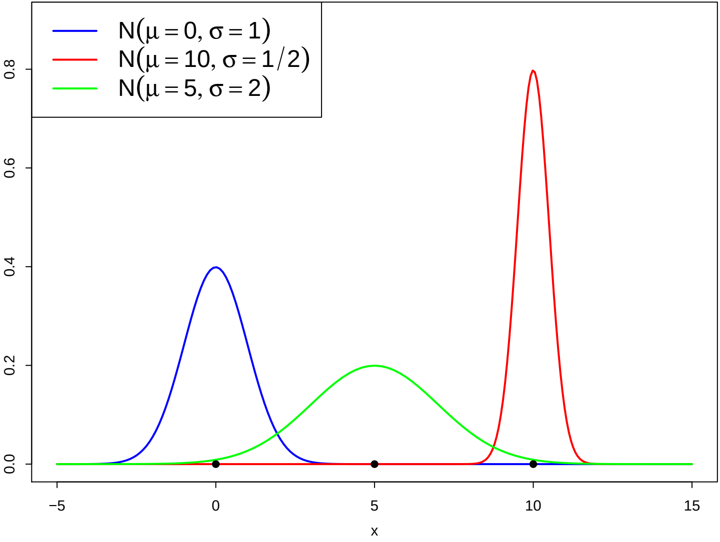
La variabile normale generale

Se si trasforma la normale standard Z riscalandola e traslandola in

$$X = \sigma Z + \mu \qquad \sigma > 0 \text{ e } \mu \text{ qualsiasi}$$

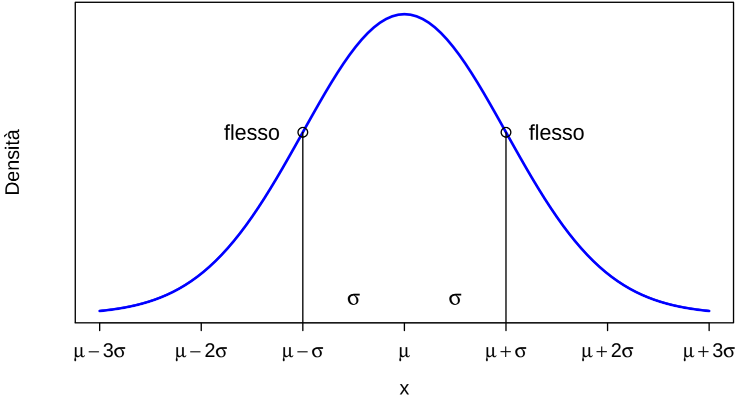
si ottiene una variabile normale generale $X \sim \mathcal{N}(\mu, \sigma^2)$. dove

- μ è il valore atteso di X
- σ è la deviazione standard di X
- σ^2 è la varianza di X



$N(\mu = 0, \sigma = 1)$
$N(\mu = 10, \sigma = 1/2)$
$N(\mu = 5, \sigma = 2)$
0.0
0.2
0.4
0.6
0.8
−5
0
5
10
15
x

Normale generale



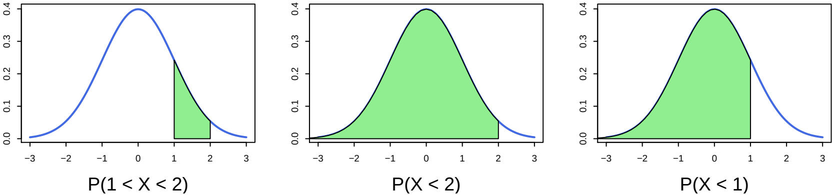
$$E(X) = \mu, \qquad \mathrm{var}(X) = \sigma^2.$$

Due regole per la probabilità di intervalli per la normale

Regola 1

La probabilità di un intervallo $[a, b]$ è

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a)$$



I segni di uguaglianza non hanno importanza.

Regola 2

Se $X \sim N(\mu, \sigma^2)$ è una normale qualsiasi la sua probabilità cumulata $\Pr(X \leq x)$ si può calcolare usando le tavole della normale standard perché

$$\Pr(X \leq x) = \Pr\left(Z \leq \frac{x-\mu}{\sigma}\right)$$

dove

- Si standardizza $X \sim \mathcal{N}(\mu, \sigma^2)$ che diventa $Z \sim \mathcal{N}(0, 1)$
- Si standardizza x che diventa $z = \dfrac{x-\mu}{\sigma}$

Esempio

Supponiamo che il punteggio X allo scritto sia normale con media $\mu = 19$ e deviazione standard $\sigma = 2$, qual è la probabilità che uno studente prenda un voto $21 \leq X \leq 23$?

Risposta: $\Pr(X \leq 23) - \Pr(X \leq 21)$

dove

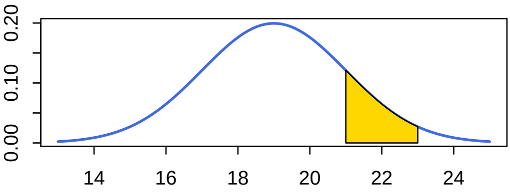
- $\Pr(X \leq 23) = \Pr[Z \leq (23 - 19)/2] = \Pr(Z \leq 2)$
- $\Pr(X \leq 21) = \Pr[Z \leq (21 - 19)/2] = \Pr(Z \leq 1)$

Quindi $\Pr(21 \leq X \leq 23) = \Pr(Z \leq 2) - \Pr(Z \leq 1)$.

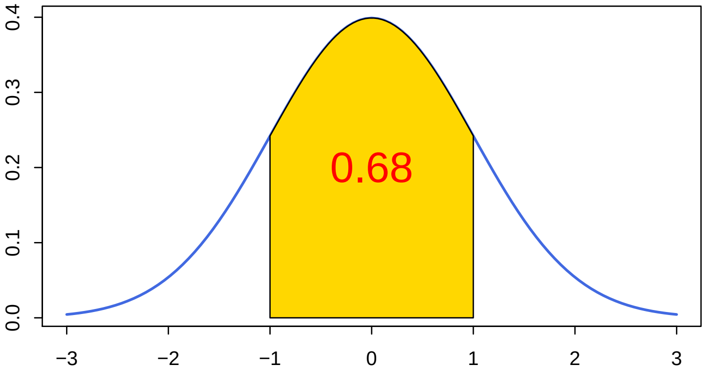
Si va sulla Tavola 1 e si calcola

$$\Pr(Z \leq 2) - \Pr(Z \leq 1) = 0.9772 - 0.8413 = 0.1359.$$

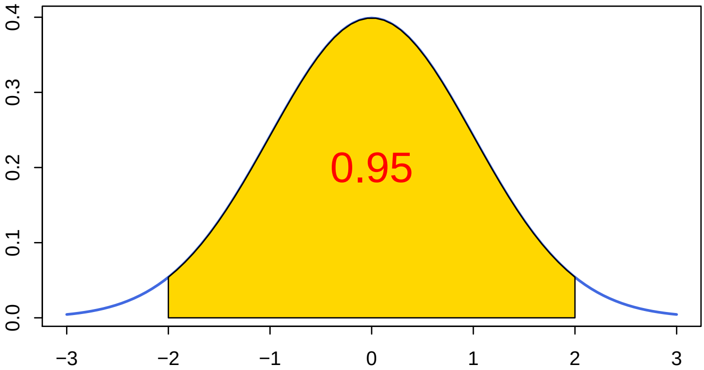
z	F(z)	z	F(z)
0.98	0.8365	1.98	0.9761
0.99	0.8389	1.99	0.9767
1.00	0.8413	2.00	0.9772
1.01	0.8438	2.01	0.9778
1.02	0.8461	2.02	0.9783



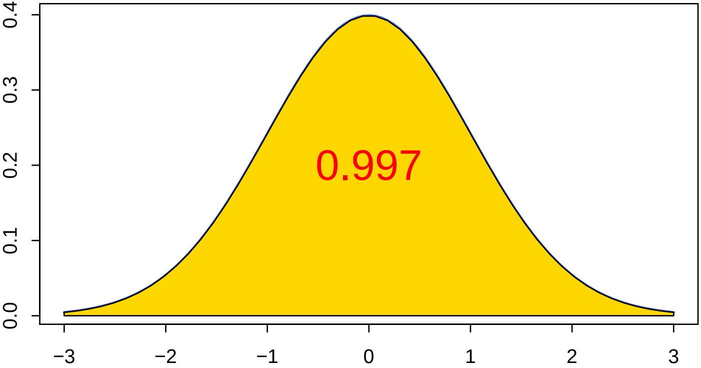
Intervallo $\Pr(-1 \leq Z \leq 1)$



Intervallo $\Pr(-2 \leq Z \leq 2)$



Intervallo $\Pr(-3 \leq Z \leq 3)$



Intorni della media

Consideriamo le probabilità degli intorni della media di raggio un multiplo di σ

$$I_z = (\mu - z\,\sigma, \mu + z\,\sigma)$$

- z può anche essere un numero con la virgola
- In particolare studiamo I_1, I_2, I_3.

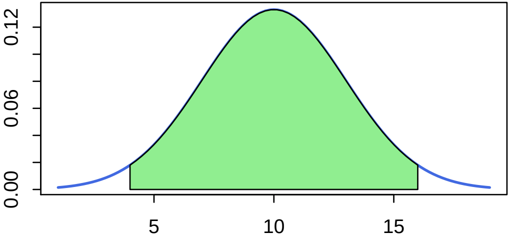
$$\Pr(I_z) = \Pr(\mu - z\,\sigma \leq X \leq \mu + z\,\sigma)$$

è la probabilità che X differisca dalla media per z deviazioni standard.

Esempio

$X \sim \mathcal{N}(\mu = 10, \sigma = 3)$

$$\Pr(I_2) = \Pr(10 - 6 \le X \le 10 + 6)$$



è un intorno con centro μ e raggio 2 σ.

Regola empirica

Per la normale esattamente, e per una distribuzione di forma campanulare simmetrica, approssimativamente, le probabilità $\Pr(I_k)$ sono

$$\Pr(I_1) \simeq 0.68, \quad \Pr(I_2) \simeq 0.95, \quad \Pr(I_3) \simeq 0.99$$

Disuguaglianza di Chebyshev

Per una qualunque variabile aleatoria la probabilità che X differisca dalla media per z deviazioni standard, è sempre maggiore di $1 - \frac{1}{z^2}$:

$$\Pr(I_z) \geq 1 - \frac{1}{z^2}$$

quindi

$$\Pr(I_1) \geq 0 \text{ (banale)}, \quad \Pr(I_2) \geq \frac{3}{4}, \quad \Pr(I_3) \geq \frac{8}{9}$$

Esercizio

Una variabile X ha una distribuzione con media 250 e deviazione standard 20. Dare indicazioni sulla probabilità: $\Pr(210 \leq X \leq 290)$

- L'intervallo ha come punto centrale la media: $(210 + 290)/2 = 250$
- Il raggio è $(290 - 210)/2 = 40$. Quindi l'intervallo è 250 ± 40.
- Quante deviazioni standard è il raggio? $z = 40/\sigma = 40/20 = 2$
- Chebyshev indica allora che

$$\Pr(210 \leq X \leq 290) \geq 1 - 1/(z^2) = 1 - 1/(2^2) = 0.75.$$

z non deve essere intero per forza

Una variabile X ha una distribuzione con media 250 e deviazione standard 20. Dare indicazioni sulla probabilità: $\Pr(220 \le X \le 280)$.

- L'intervallo ha come punto centrale la media: $(220 + 280)/2 = 250$
- Il raggio è $(280 - 220)/2 = 30$. Quindi l'intervallo è 250 ± 30.
- Quante deviazioni standard è il raggio? $z = 30/\sigma = 30/20 = 1.5$
- Chebyshev indica allora che

$$\Pr(220 \le X \le 280) \ge 1 - 1/(z^2) = 1 - 1/(1.5)^2 = 0.55.$$

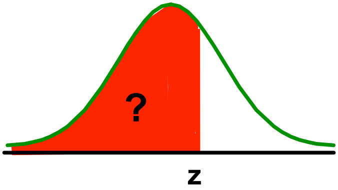
Quartili e quantili della normale

- Il primo quartile detto anche quantile di ordine 0.25 ed è il valore k tale che $\Pr(X \leq k) = 0.25$
- Il secondo quartile è la mediana ed è il valore k tale che $\Pr(X \leq k) = 0.50$: ovviamente $k = \mu$.
- Il terzo quartile detto anche quantile di ordine 0.75 ed è il valore k tale che $\Pr(X \leq k) = 0.25$

Si può definire altresì il quantile di ordine p come il valore k tale che $\Pr(X \leq x) = p$.

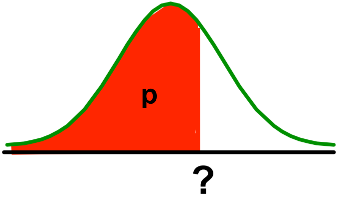
Probabilità cumulate e quantili

Problema diretto: dato un valore di *z* determinare la probabilità cumulata



Cerca la probabilità nella **2a** colonna F(z)

Problema inverso: dato un valore *p* della probabilità cumulata, determinare il valore *z* corrispondente



Cerca il valore nella **1a** colonna z

- Si usa la Tavola 1 all'inverso.

  Esempio: trovare il quantile 0.75: $\Pr(Z \leq k) = 0.75$

```
   z   F(z)
0.66 0.7454
0.67 0.7486
0.68 0.7517
0.69 0.7549
0.70 0.7580
```

- Si cerca il valore 0.75 nella colonna $F(z)$
- Sta tra 0.7486 e 0.7517, quindi $k = (0.67 + 0.68)/2 = 0.675$

La temperatura minima X ha distribuzione Normale con media $15C^{\circ}$ e deviazione standard $2C^{\circ}$. Qual è la temperatura k tale che $\Pr(X \leq k) = 0.80$?

- Standardizza:

$$\Pr(X \leq k) = \Pr\left(Z \leq \frac{k-15}{2}\right) = 0.8$$

- Trova il quantile 0.8 sulla Tavola 1

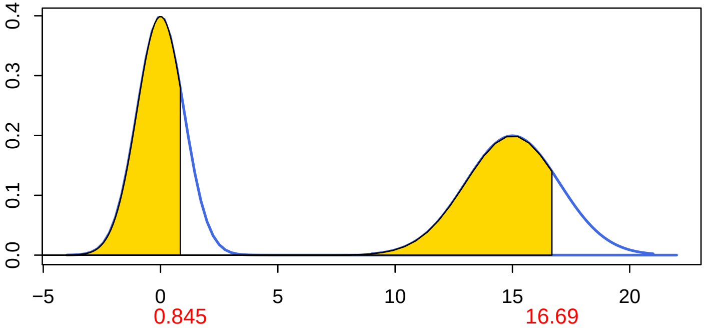
```
   z   F(z)
0.82 0.7939
0.83 0.7967
0.84 0.7995
0.85 0.8023
0.86 0.8051
```

(continua)

- Ottieni 0.845
- Trova k risolvendo l'equazione

$$\frac{k-15}{2} = 0.845$$

- Ottieni $k = 15 + 2 \cdot 0.845 = 16.69$.



Quantili superiori

Per $Z \sim N(0, 1)$ il determina il valore k tale che $P(Z > k) = 0.10$

- Prima occorre trasformare il problema cin modo che riguardi le probabilità cumulate:

  $P(Z > k) = 0.10$ è equivalente a $P(Z \leq k) = 1-0.1 = 0.90$

- Quindi uso le tavole all'inverso

```
   z   F(z)
1.26 0.8962
1.27 0.8980
1.28 0.8997
1.29 0.9015
1.30 0.9032
```

e trovo che il quantile 0.9 è 1.285 che è la risposta.

Le previsioni sulla domanda di un prodotto sono una variabile normale X con media 1200 e deviazione standard 100.

– Qual è il valore delle vendite k che ha probabilità 0.10 di essere superato?

- Se $X \sim N(\mu = 1200, \sigma = 100)$ sono le vendite si deve calcolare il valore k tale che $\Pr(X > k) = 0.1$.
- Standardizzo ambo i membri:

$$\Pr\left(Z > \frac{k - 1200}{100}\right) = 0.1$$

- Trasformo in una probabilità cumulata

$$\Pr\left(Z \leq \frac{k - 1200}{100}\right) = 1 - 0.1 = 0.9$$

(continua)

- Sulle tavole trovo che il quantile 0.9 è 1.285
- Risolvo l'equazione

$$\frac{k - 1200}{100} = 1.285$$

ottenendo $k = 1200 + 1.285 \cdot 100 = 1328.5$.