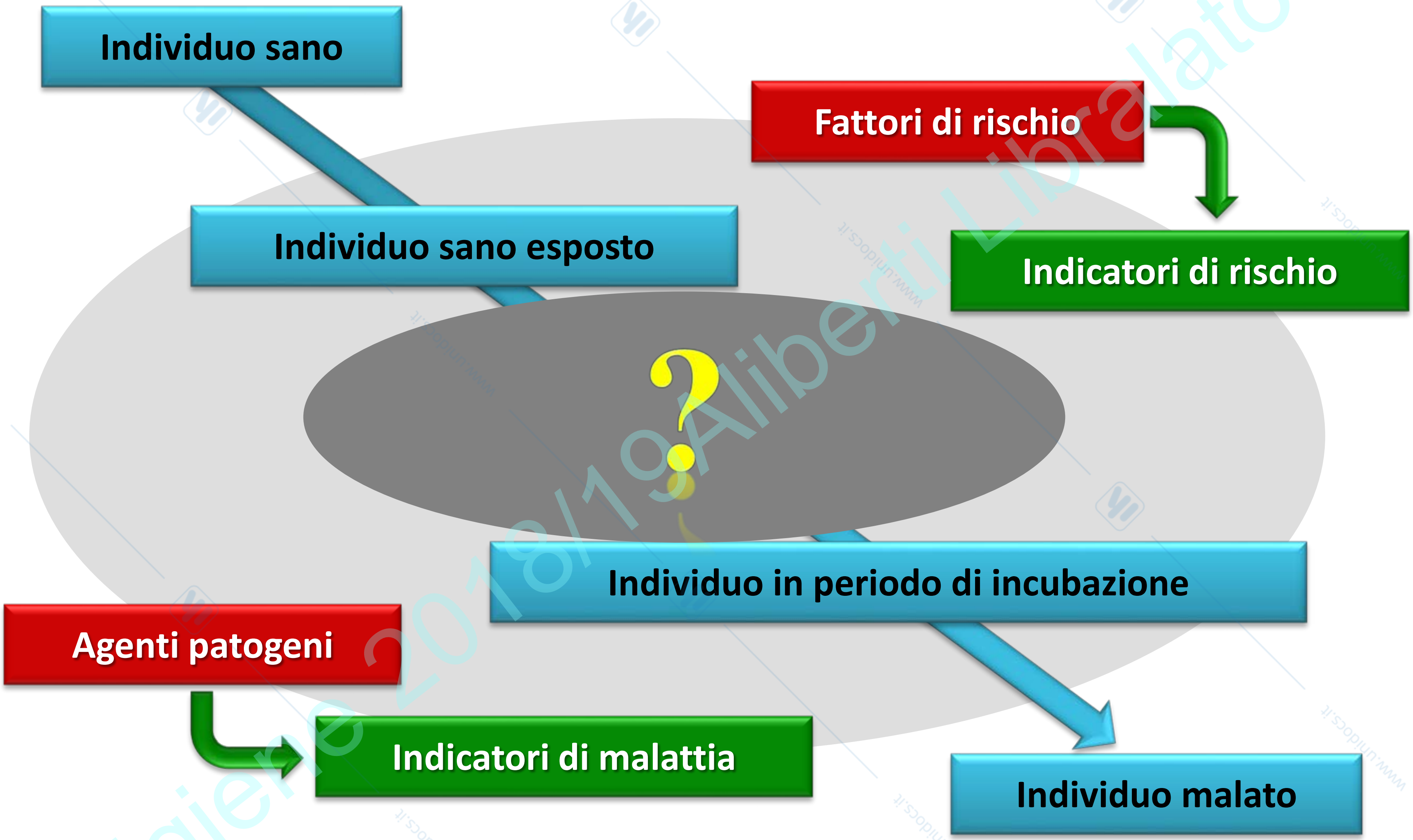
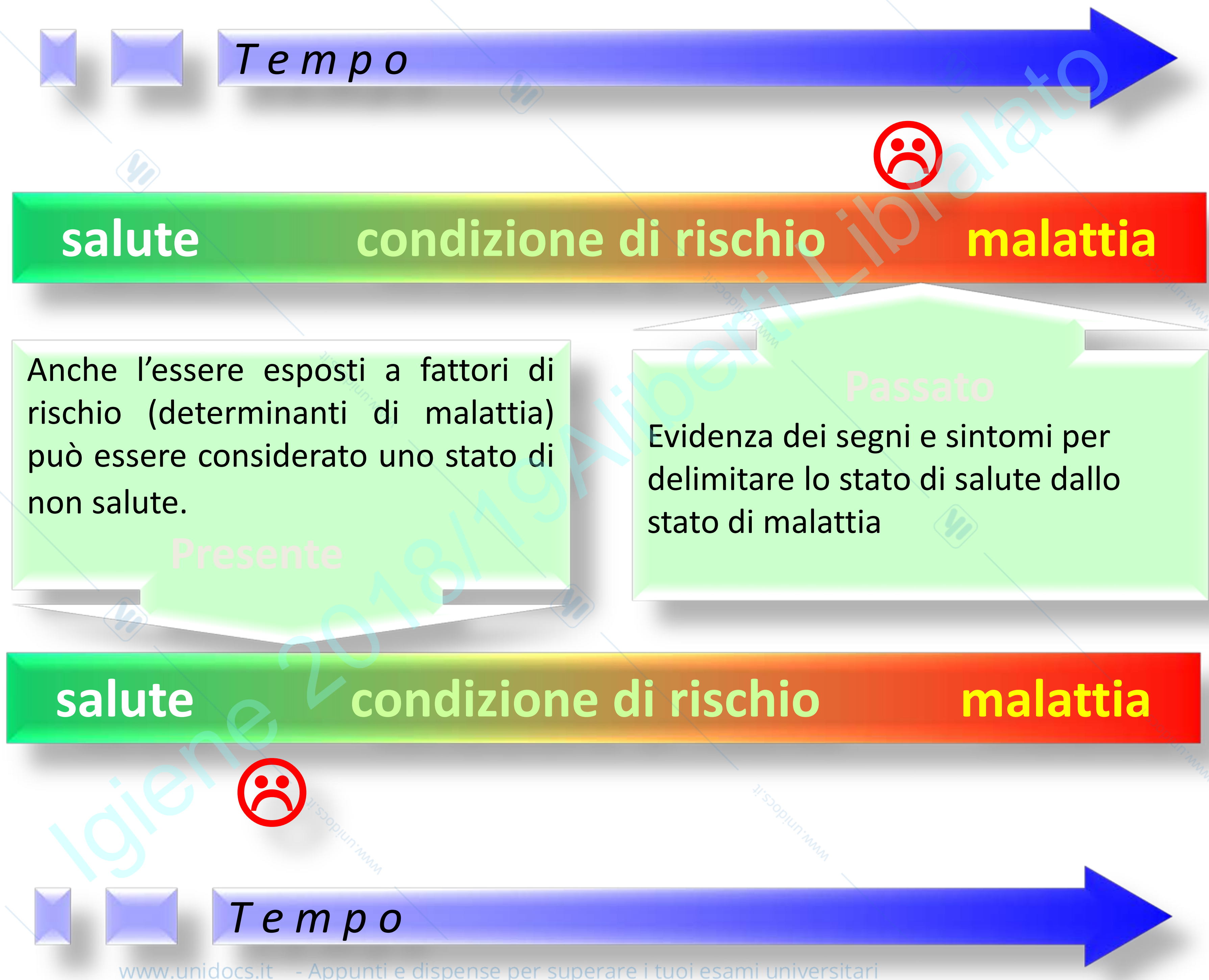
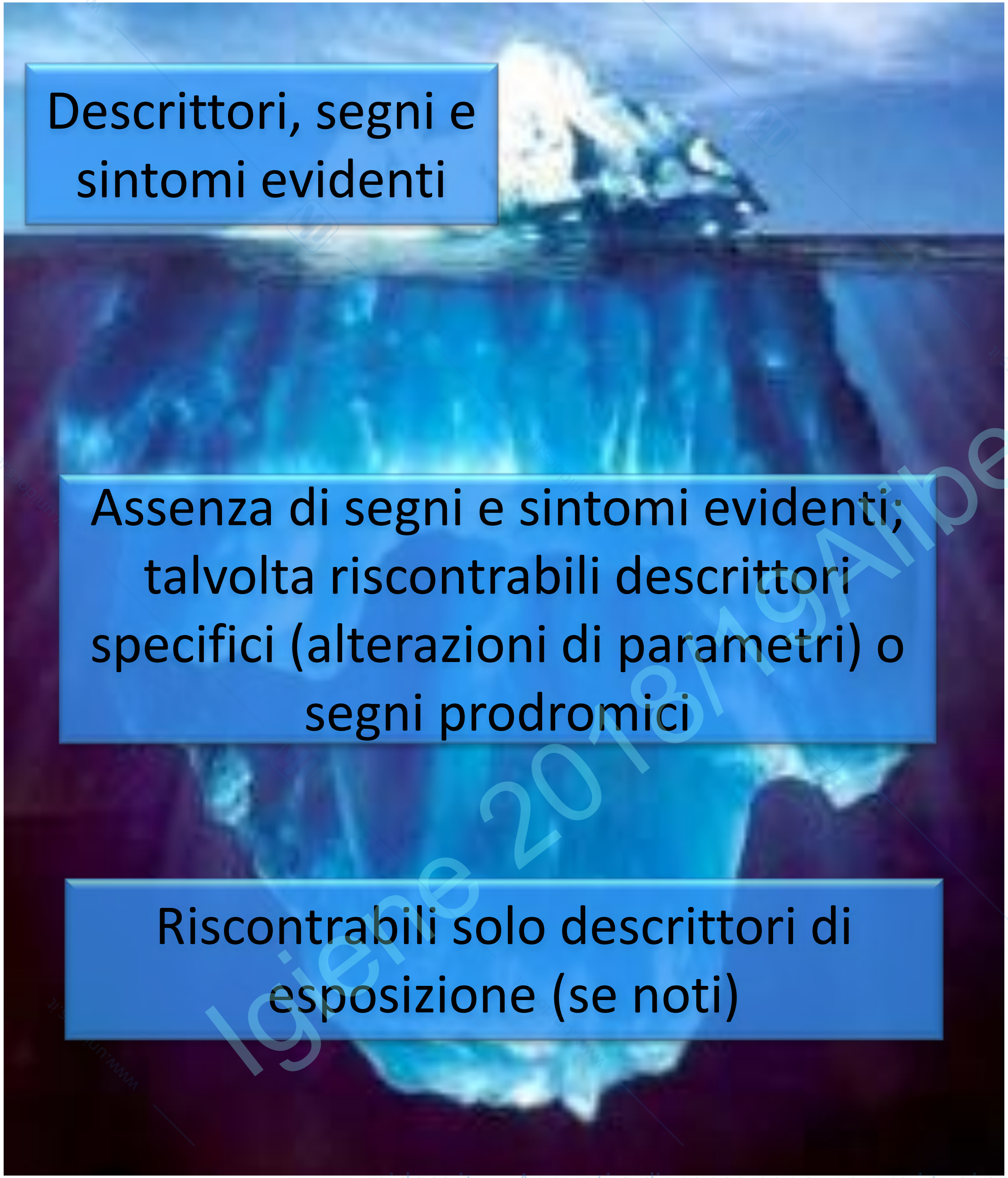
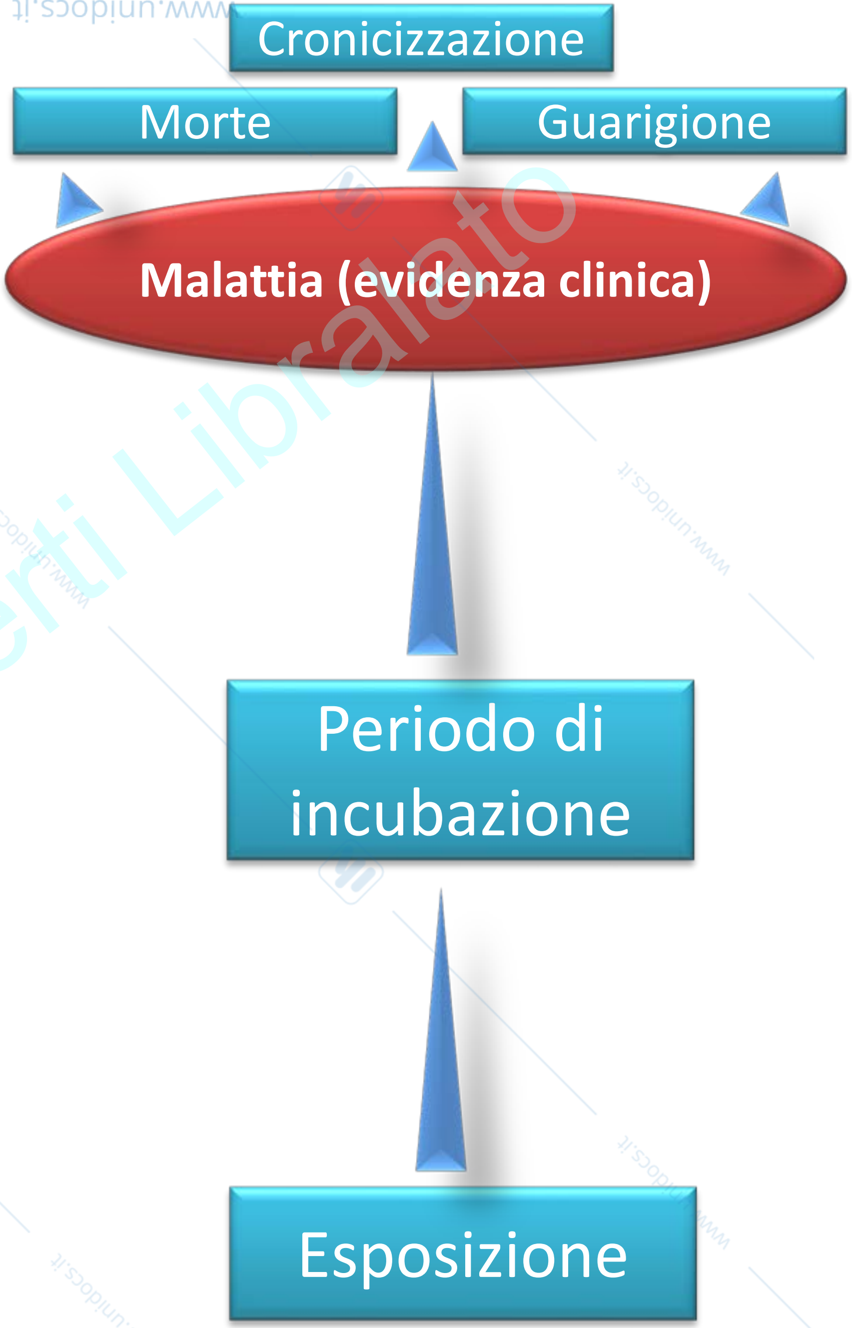

Individuo sano
Individuo sano esposto
Fattori di rischio
Indicatori di rischio
?
Individuo in periodo di incubazione
Agenti patogeni
Indicatori di malattia
Individuo malato


Tempo
salute
condizione di rischio
malattia
Passato
Evidenza dei segni e sintomi per delimitare lo stato di salute dallo stato di malattia
Anche l’essere esposti a fattori di rischio (determinanti di malattia) può essere considerato uno stato di non salute.
Presente
salute
condizione di rischio
malattia
Tempo


Descrittori, segni e sintomi evidenti
Assenza di segni e sintomi evidenti; talvolta riscontrabili descrittori specifici (alterazioni di parametri) o segni prodromici
Riscontrabili solo descrittori di esposizione (se noti)


Cronicizzazione
Morte
Guarigione
Malattia (evidenza clinica)
Periodo di incubazione
Esposizione

έπί δήμοσ λόγος

epì	démos	logos
(attorno)	(popolazione)	(discussione, studio)

Studio della evoluzione delle malattie nelle popolazioni

EPIDEMIOLOGIA : POPOLAZIONE = PATOGENESI : INDIVIDUO

Epidemiologia

popolazioni

Gruppi di individui accomunati da una o più caratteristiche

frequenza

Quanti sono i casi

Quando compare

Dove compare

distribuzione

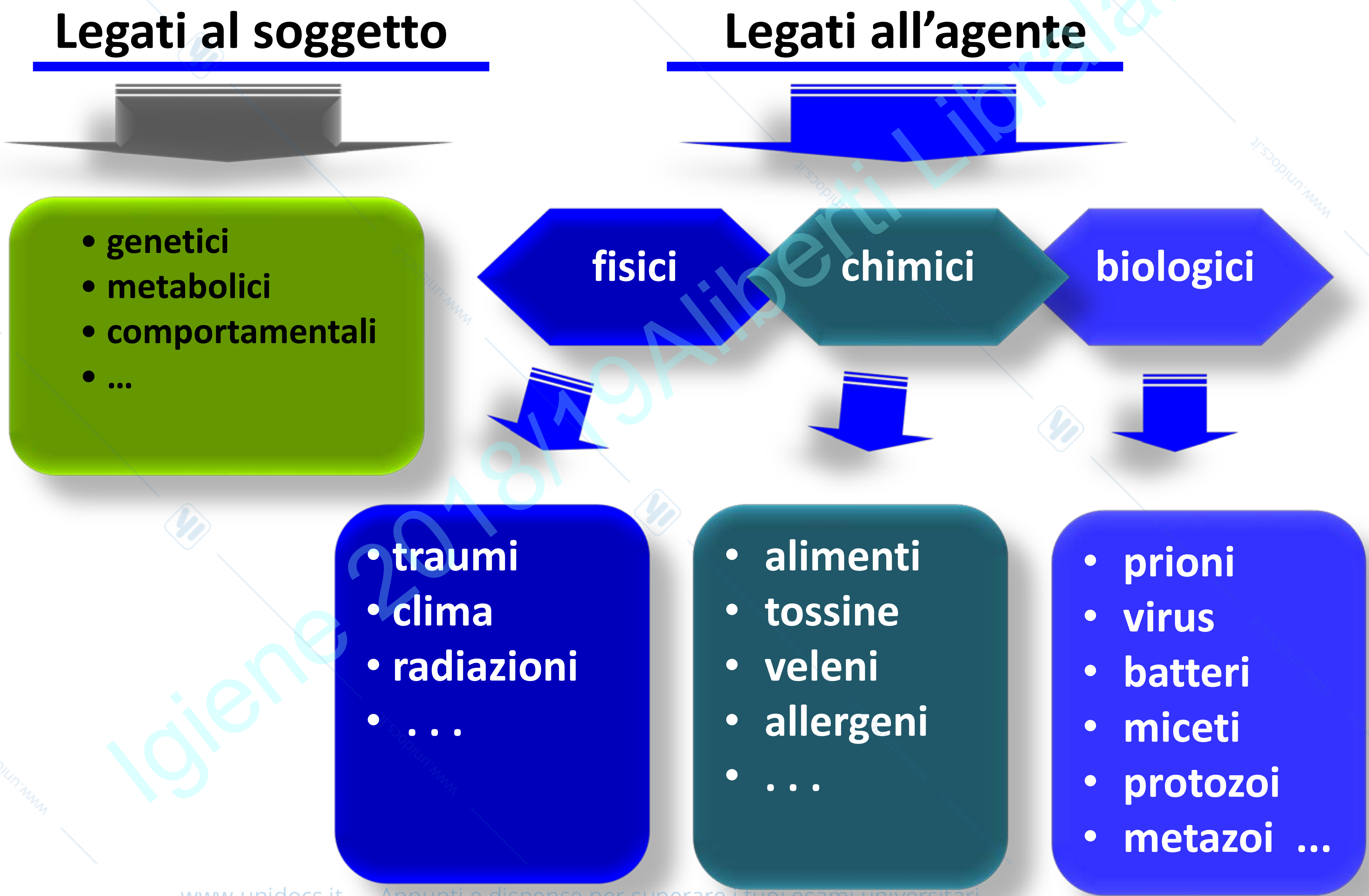
salute/malattia

Si studiano i malati ed i sani

determinanti

Quali le cause [fattori o determinanti (maggiori-minori)]

Determinanti primari



Determinanti secondari

Endogeni

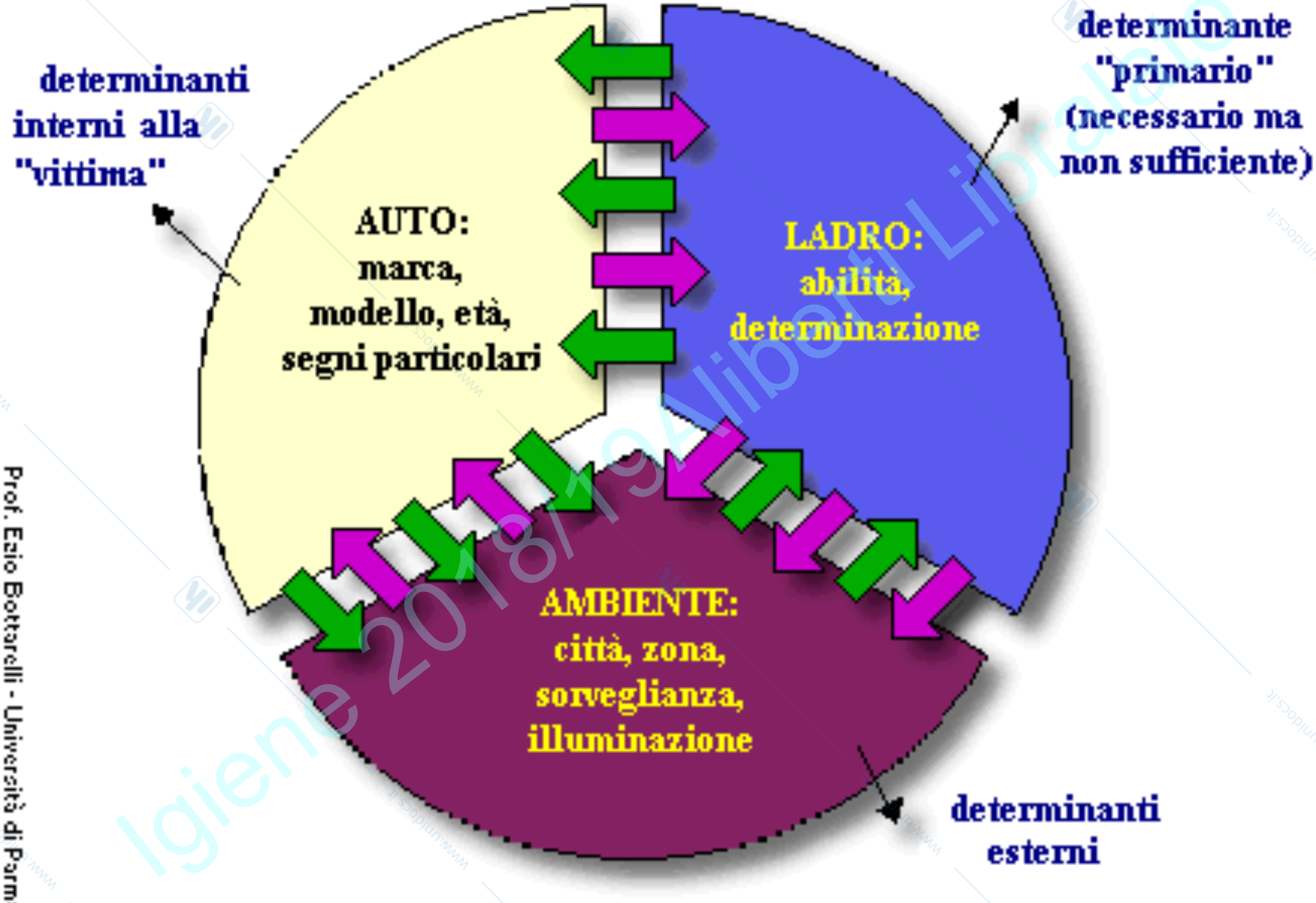
- genetici
- età
- sesso
- stato nutrizionale
- stato ormonale
- stato immunitario
- stato funzionale
- comportamento
- …

Esogeni

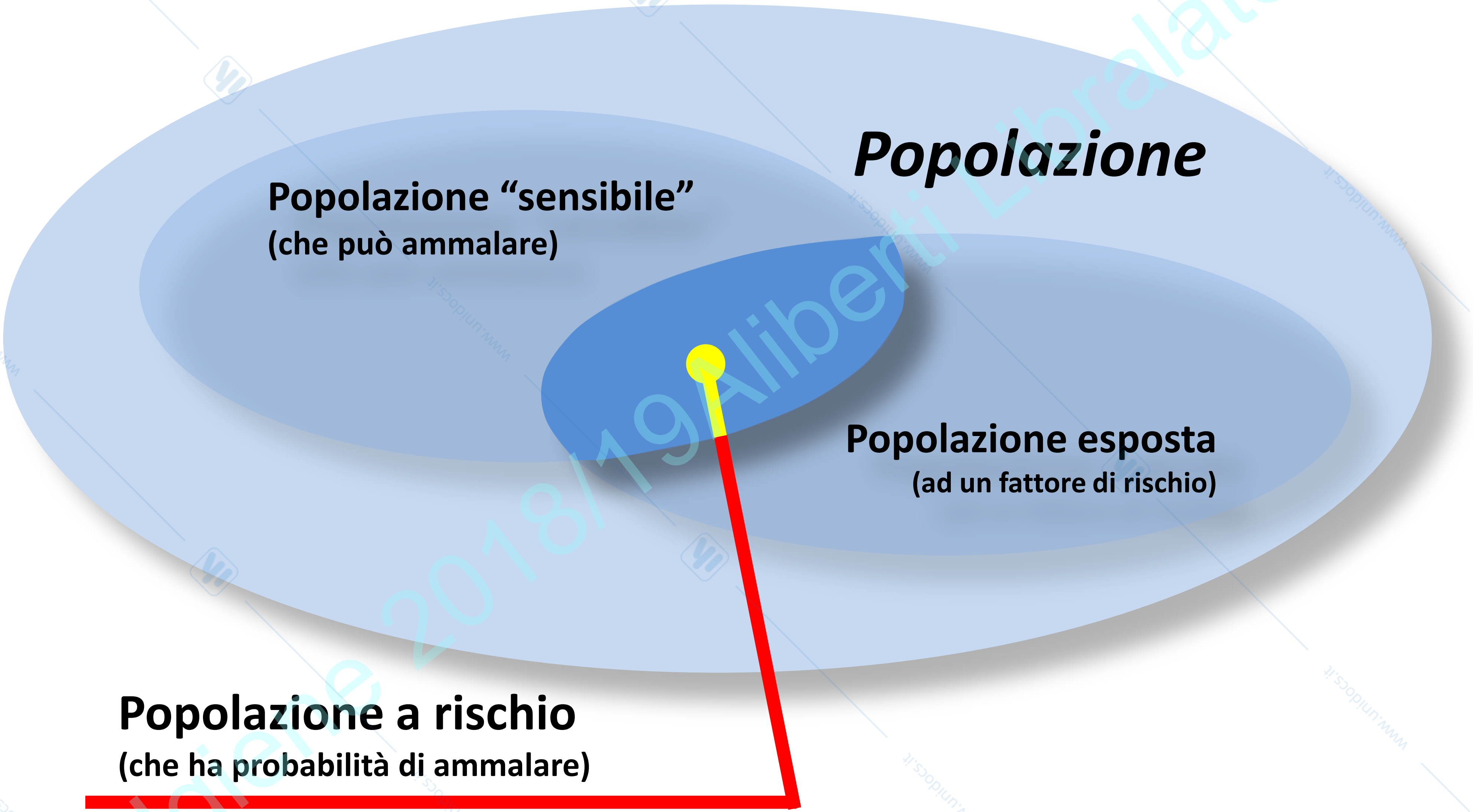
- ambiente
- clima
- stress
- guerre
- carestie
- …

La "triade" epidemiologica

Schema di alcuni possibili determinanti di furto d'auto



Prof. Ezio Bottarelli - Università di Parma


Popolazione
Popolazione "sensibile"
(che può ammalare)
Popolazione esposta
(ad un fattore di rischio)
Popolazione a rischio
(che ha probabilità di ammalare)

Tasso: proporzione che permette di definire la variazione di una grandezza rispetto ad un'altra

Nel caso di una osservazione epidemiologica:

numero di eventi osservati : popolazione ove l'evento è osservabile = Tasso : 10^n

$$\text{Tasso} = \frac{\text{numero di eventi osservati}}{\text{popolazione ove l'evento è osservabile}} \times 10^n$$

Tasso: esempio

Nella città di Napoli sono registrati casi di influenza

numero di eventi osservati = 152

popolazione ove l'evento è osservabile (tutta la popolazione tranne i vaccinati) = 961.257

$$152 : 961.257 = \text{Tasso} : 10^n \text{ (nel caso specifico 100.000)}$$

$$\text{Tasso} = \frac{152 \times 100.000}{961.257} = 15,81\ (10^5)$$

$$\text{Tasso} = \frac{\text{numero di eventi osservati}}{\text{numero di elementi nell'universo di osservazione ove è verificabile l'evento osservato}} \times 10^{n}$$

Fattore moltiplicativo che permette di rapportare l'evento ad un numero di elementi dell'universo (popolazione) di entità pari a un multiplo di 10 (x%, y ‰, z‱, w%ooo, …)

Tasso di prevalenza = $\dfrac{\text{numero di eventi osservati}}{\text{numero di elementi nell'universo di osservazione ove è verificabile l'evento osservato}} \times 10^{n}$

Fattore moltiplicativo che permette di rapportare l'evento ad un numero di elementi dell'universo (popolazione) di entità pari a un multiplo di 10 (x%, y ‰, z‱, w%ooo, …)

il tasso rilevato è puntuale, cioè si riferisce al momento dell'osservazione

Tasso di prevalenza periodale =

$$\frac{\text{numero di eventi osservati nel tempo} - \Delta t \text{ (periodo di osservazione)}}{\text{numero di elementi nell'universo di osservazione ove è verificabile l'evento osservato}} \times 10^{n}$$

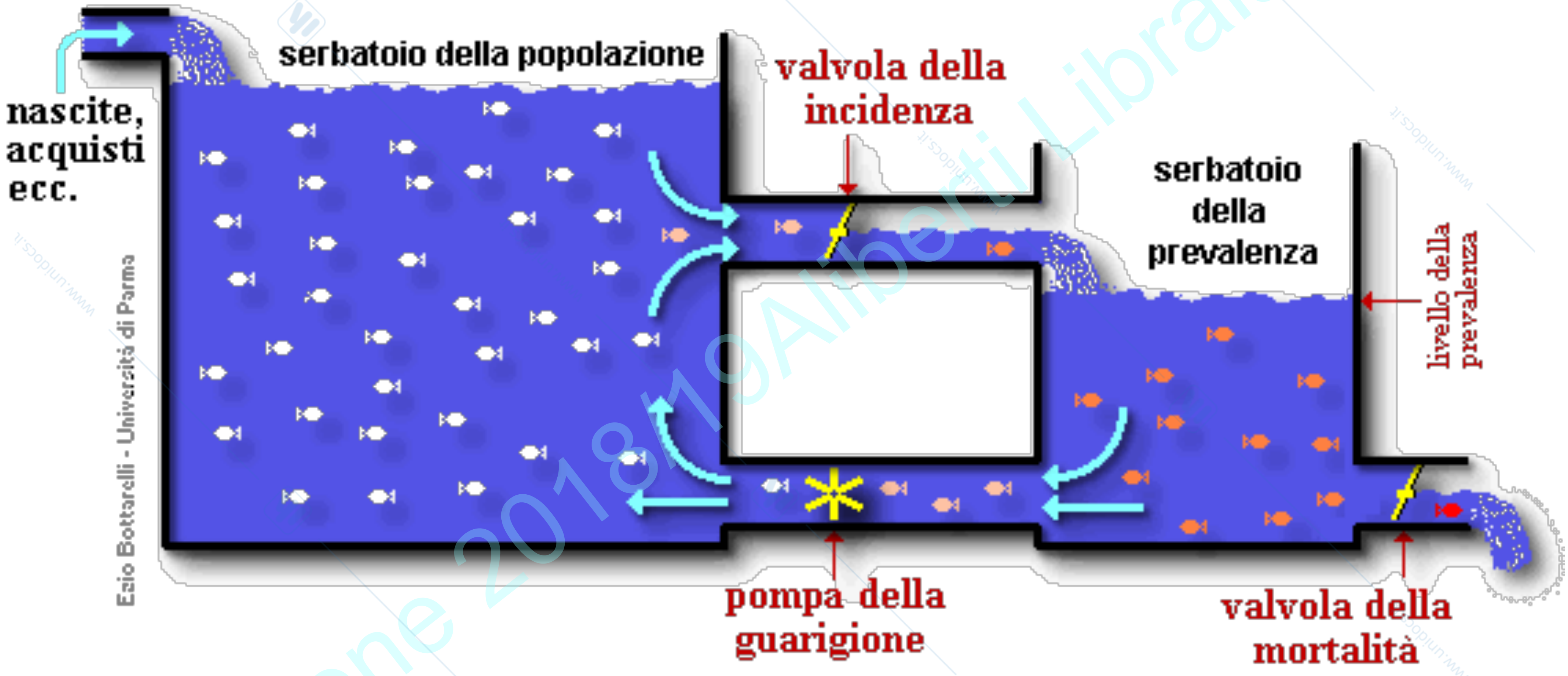
Fattore moltiplicativo che permette di rapportare l'evento ad un numero di elementi dell'universo (popolazione) di entità pari a un multiplo di 10 (x%, y ‰, z%oo, w%ooo, …)

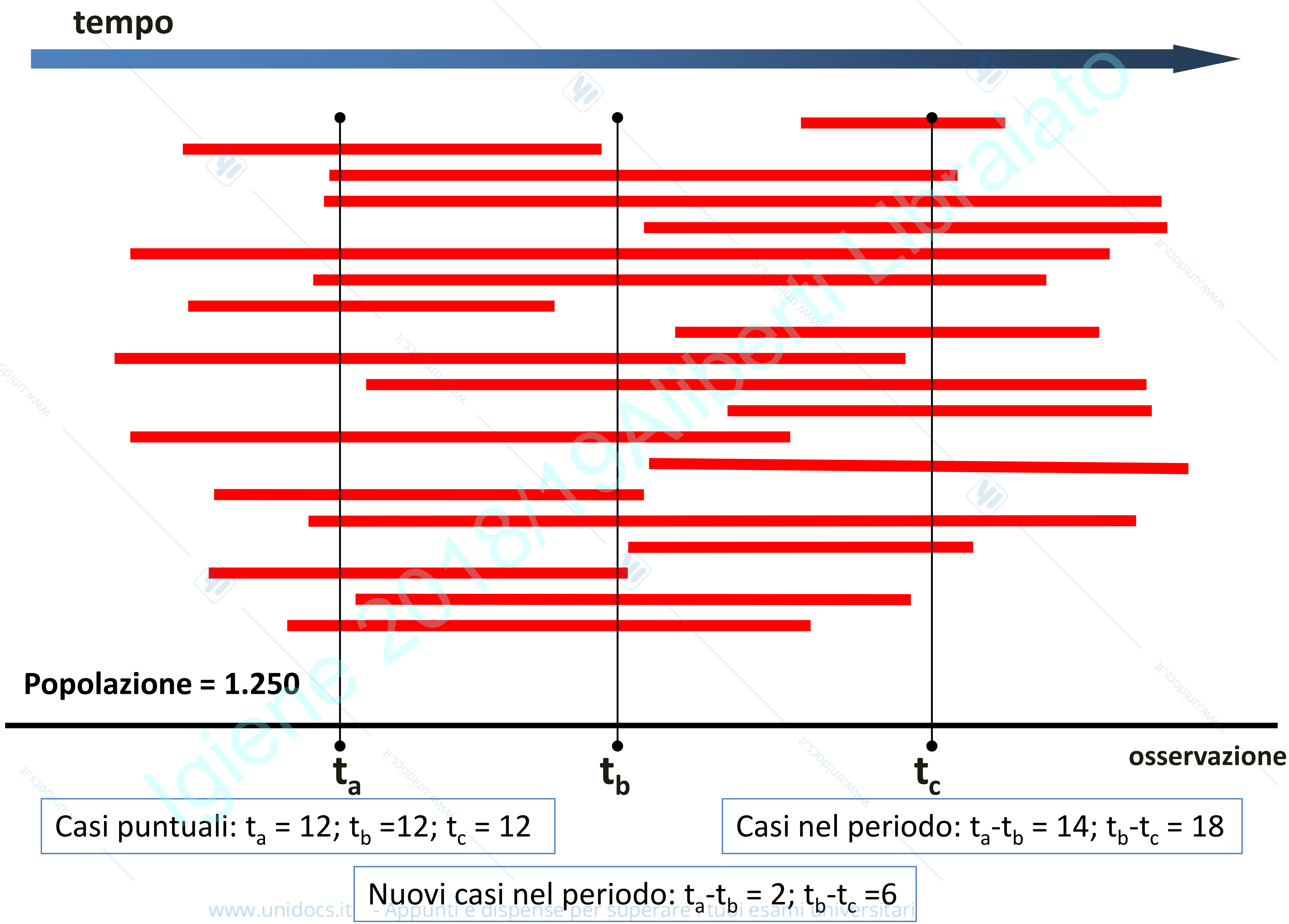
il tasso rilevato è puntuale, ma si riferisce al periodo dell'osservazione

Tasso di incidenza = $\dfrac{\text{numero di } \textbf{nuovi} \text{ eventi osservati nel tempo} - \Delta t \text{ (periodo di osservazione)}}{\text{numero di elementi nell'universo di osservazione ove è verificabile l'evento osservato}} \times 10^{n}$

Fattore moltiplicativo che permette di rapportare l'evento ad un numero di elementi dell'universo (popolazione) di entità pari a un multiplo di 10 (x%, y ‰, z%oo, w%ooo, …)

il tasso rilevato è riferito ad un periodo di osservazione ove si rilevano i nuovi casi nella popolazione


nascite, acquisti ecc.
serbatoio della popolazione
valvola della incidenza
serbatoio della prevalenza
livello della prevalenza
pompa della guarigione
valvola della mortalità
Ezio Bottarelli - Università di Parma


tempo
Popolazione = 1.250
t_a
t_b
t_c
osservazione
Casi puntuali: t_a = 12; t_b =12; t_c = 12
Casi nel periodo: t_a-t_b = 14; t_b-t_c = 18
Nuovi casi nel periodo: t_a-t_b = 2; t_b-t_c =6

Prevalenza puntuale (t_a) $\quad \mathbf{Tp_a} = \dfrac{12}{1250} \times 1000 = \mathbf{9{,}6\ ‰}$

Prevalenza puntuale (t_b) $\quad \mathbf{Tp_b} = \dfrac{12}{1250} \times 1000 = \mathbf{9{,}6\ ‰}$

Prevalenza puntuale (t_c) $\quad \mathbf{Tp_c} = \dfrac{12}{1250} \times 1000 = \mathbf{9{,}6\ ‰}$

Prevalenza periodale $(t_a - t_b\ ;\ t_b - t_c)$

$$Tp_{(t_a - t_b)} = \frac{14}{1250} \times 1000 = \mathbf{11{,}2\ ‰}$$

$$Tp_{(t_b - t_c)} = \frac{18}{1250} \times 1000 = \mathbf{14{,}4\ ‰}$$

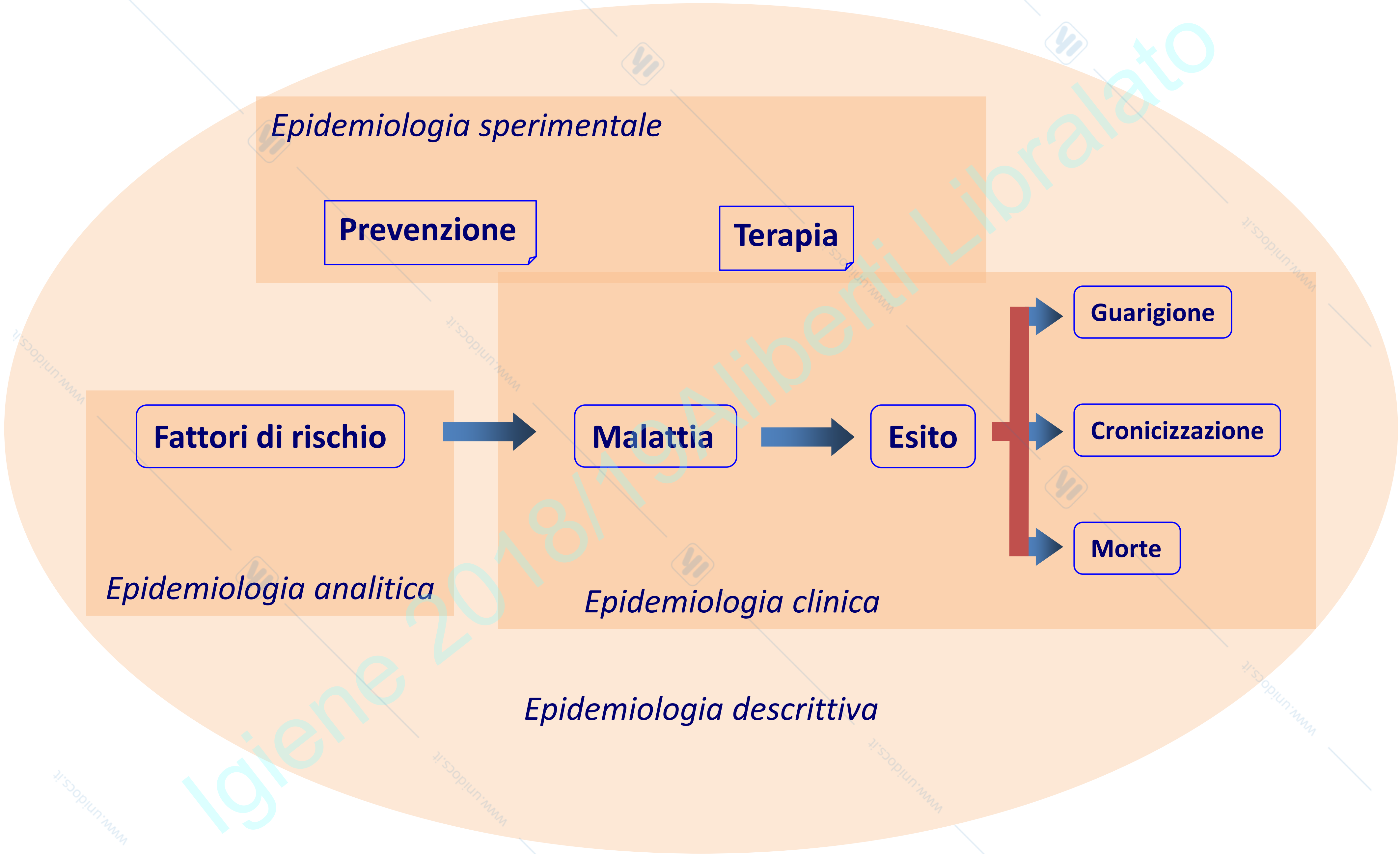
Variazione % (b-a)/a*100 = (14,4-11,2)/11,2 *100= 28,6

Incidenza $(t_a - t_b\ ;\ t_b - t_c)$

$$Ti_{(t_a - t_b)} = \frac{2}{1250} \times 1000 = \mathbf{1{,}6\ ‰}$$

$$Ti_{(t_b - t_c)} = \frac{6}{1250} \times 1000 = \mathbf{4{,}8\ ‰}$$

Variazione % (b-a)/a*100 = (4,8-1,6)/1,6 *100= 200



Epidemiologia sperimentale
Prevenzione
Terapia
Fattori di rischio
Malattia
Esito
Guarigione
Cronicizzazione
Morte
Epidemiologia analitica
Epidemiologia clinica
Epidemiologia descrittiva

In epidmiologia

Popolazione

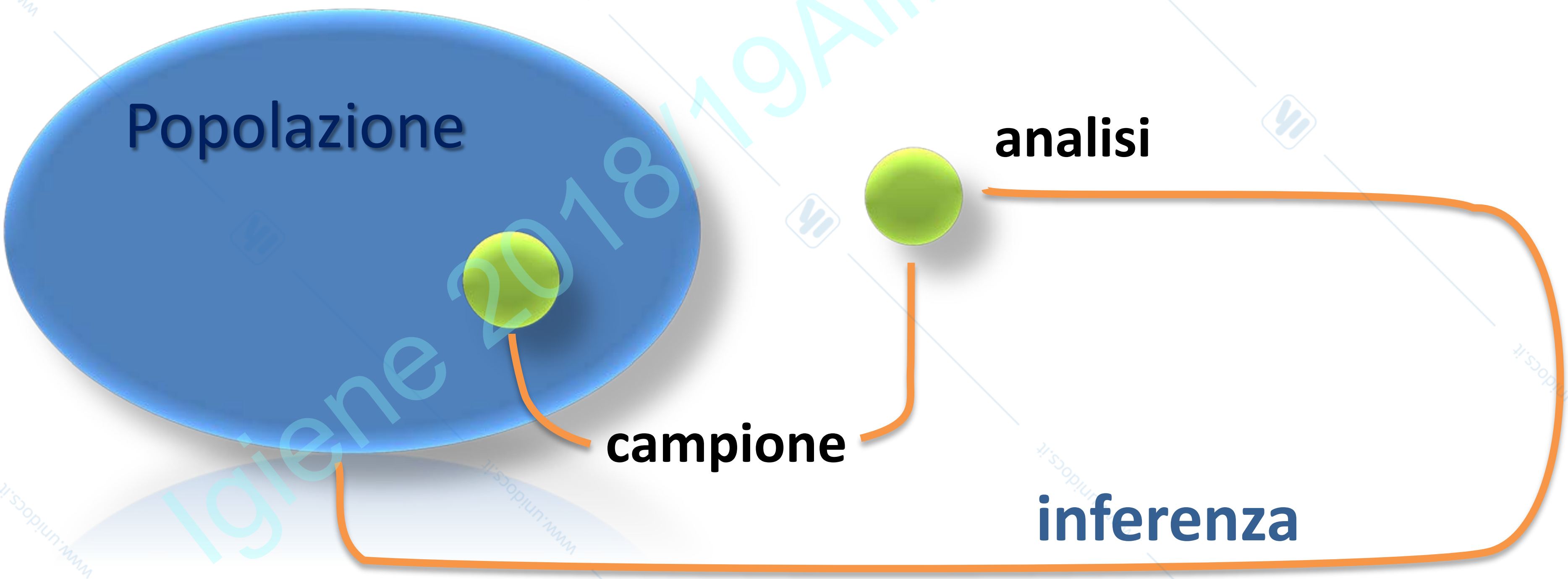
significa

Insieme di individui (o di unità) che hanno in comune una o più caratteristiche. È inteso anche quale "universo di osservazione"

Le osservazioni su una popolazione raramente coinvolgono tutta la popolazione; quasi sempre ci si basa su un "campione "

Perché si campiona?

- Raramente possono essere studiate tutte le unità che compongono la popolazione oggetto di studio
- L'analisi del campione, se correttamente prelevato, può permettere la generalizzazione dei risultati alla popolazione da cui proviene (inferenza)



Campionamento: metodi

Non probabilistico

Probabilistico

Ciascun elemento dell'insieme ha la stessa probabilità di far parte del campione (o quasi).

- Randomizzazione semplice
- Randomizzazione Sistematica
- Randomizzazione Stratificata
- Randomizzazione a cluster (grappolo)

Campionamento

Non probabilistico: viene effettuato allorquando non è possibile tecnicamente effettuare un campionamento probabilistico o quando non è cruciale avere campioni randomizzati per la tipologia di studio

Randomizzazione semplice: si identificano, numerandole le unità campionarie; poi viene effettuato il campionamento casuale seguendo numeri casuali generati da specifici software.

Randomizzazione sistematica: si identificano ordinandole le unità campionarie; viene poi generato un numero causale che identifica il primo campione; infine si procede al campionamento di altre unità intervallate dallo stesso numero casuale (da ciò la necessità che l'insieme da campionare sia ordinato.

Randomizzazione stratificata: è applicata ad indiemi di unità campionarie non omogenee. Vengono separate le unità campionarie in funzione di una caratteristica che le differenzia (sesso, peso, età, ecc.). In ciascuno dei sottoinsiemi si procede al campionamento randomizzato (semplice o stratificato)

Randomizzazione a cluster (a grappolo): si identificano sottoinsiemi, custers o grappoli, dell'universo di osservazione (famiglie, scuole, aree geografiche, paesi, ecc.); in modo randomizzato si scelgono i cluster e, sempre in modo randomico, i campioni nei clusters scelti. Il campionamento è anche definito come a doppio stadio.

Lo studio epidemiologico si basa su osservazioni andamenti, confronti

È quindi fondamentale conoscere i parametri che caratterizzano un insieme di dati ricavati da un'osservazione

Che sono:

Indici di tendenza centrale:
Media, Moda, Mediana …

Ed ancora

Indici di dispersione:
Devianza, Varianza, Deviazione Standar, Errore standard, …

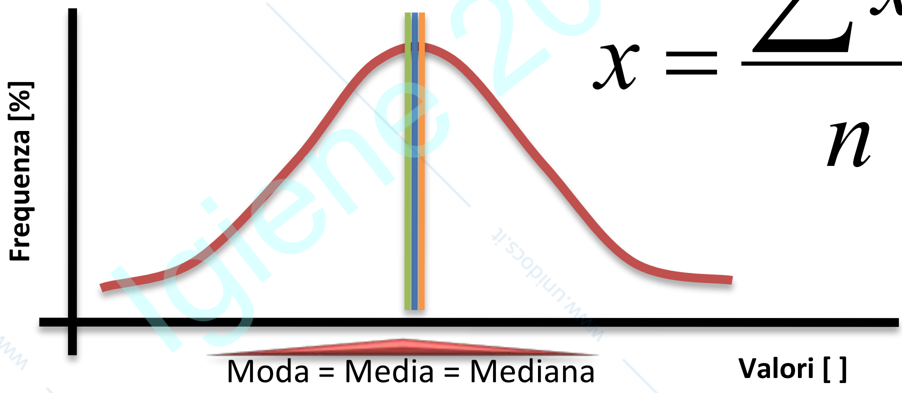
È altresì fondamentale, nei confronti, verificare la significatività di una osservazione

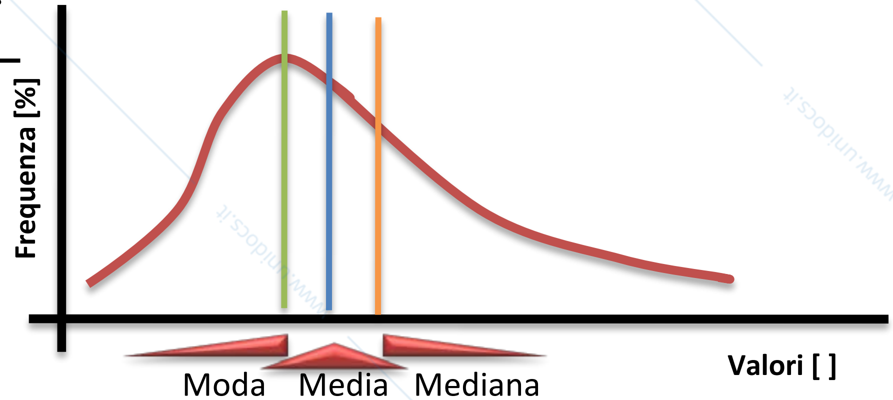
Tests statistici:
T.student, F Fischer, χ^2, …

Indici di tendenza centrale

- **Moda**: Valore più frequentemente riscontrato in una distribuzione;
- **Mediana**: valore centrale della distribuzione. Il numero di valori che lo precedono è pari a quello che lo seguono;
- **Media**: somma delle osservazioni diviso il loro numero

$$\bar{x} = \frac{\sum x_i}{n}$$





Indici di variabilità

S1	S2
13	13
14	18
13	8
11	12
12	17
14	9
13	11
15	14
14	16
11	12

Media

S1=S2=13

S1	S2
13-13=0	13-13=0
14-13=+1	18-13=+5
13-13=0	8-13=-5
11-13=-2	12-13=-1
12-13=-1	17-13=+4
14-13=+1	9-13=-4
13-13=0	11-13=-2
15-13=+2	14-13=+1
14-13=+1	16-13=+3
11-13=-2	12-13=-1

$\sum$ **(x-media)**

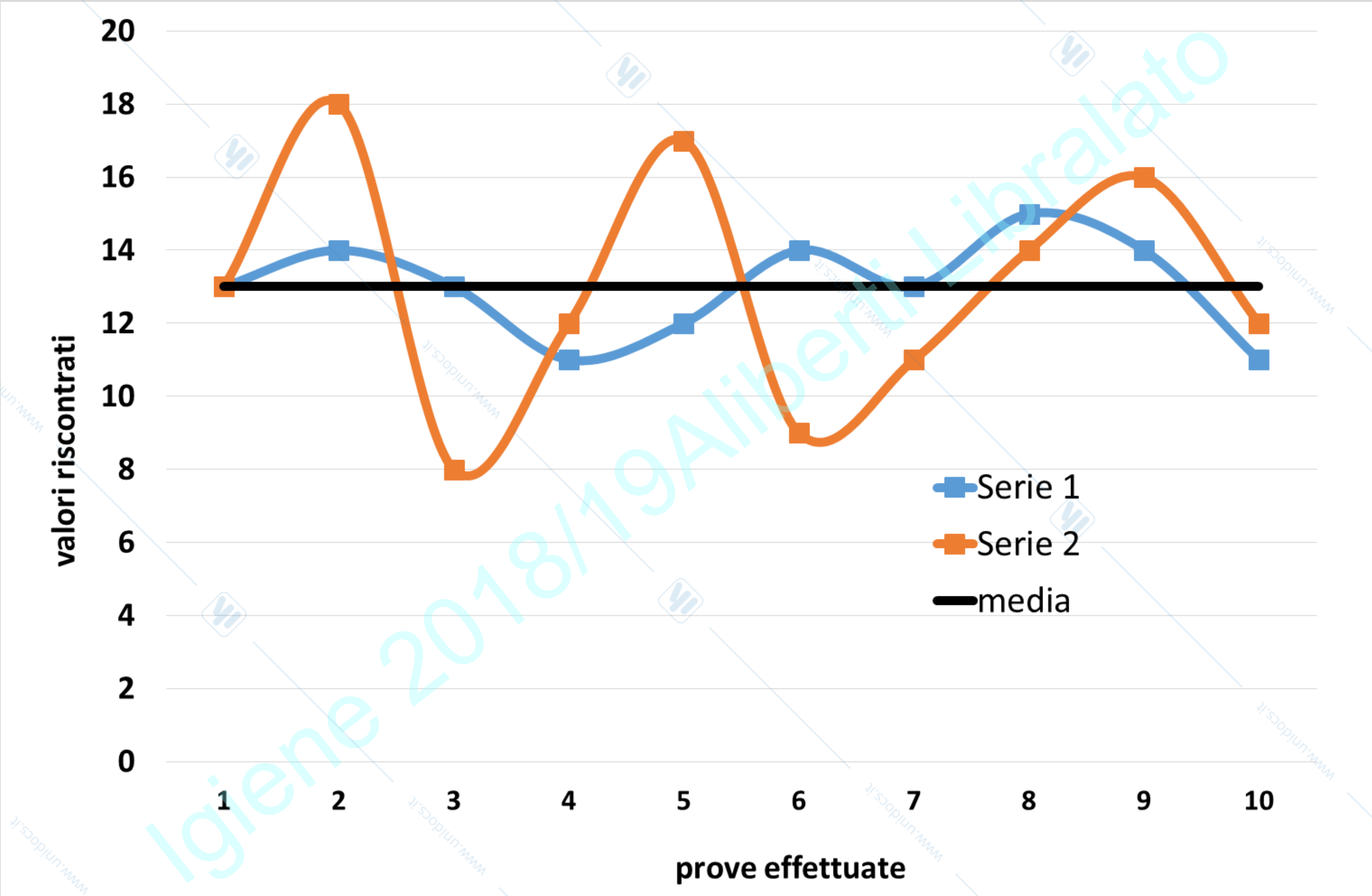
S1=S2=0

S1	S2
$0^2=0$	$0^2=0$
$+1^2=1$	$+5^2=25$
$0^2=0$	$-5^2=25$
$-2^2=4$	$-1^2=1$
$-1^2=1$	$+4^2=16$
$+1^2=1$	$-4^2=16$
$0^2=0$	$-2^2=4$
$+2^2=4$	$+1^2=1$
$+1^2=1$	$+3^2=9$
$-2^2=4$	$-1^2=1$

$\sum$ **(x-media)^2**

S1=16 S2=98

Un buon indice di dispersione è rappresentato dalla sommatorie delle differenze tra ciascun valore della serie e la media della medesima (scarto); è ovvio che la sommatoria è pari a zero (gli scarti positivi e negativi si equivalgono) perciò si eleva ciascun scarto al quadrato per trasformare gli scarti negativi in positivi. Si ottiene così la Devianza (sommatoria degli scarti quadratici)



valori riscontrati
20
18
16
14
12
10
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
prove effettuate
Serie 1
Serie 2
media

Indici di variabilità

- **Devianza (**scarto quadratico): dispersione del valore rispetto alla media.

$$D = \sum (x_i - \bar{x})^2$$

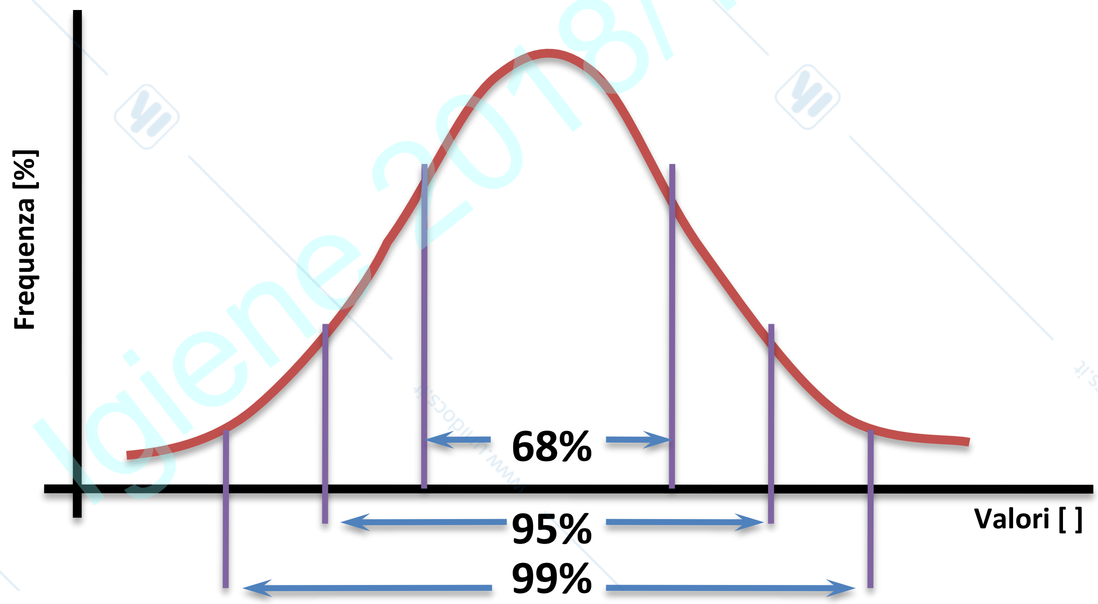
- **Varianza**: detta anche scarto quadratico medio; è la media degli scarti;

$$V = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Indici di variabilità

- **Deviazione standard (σ) :** è la radice quadrata della Varianza (riporta quindi i valori a dati reali visto che i parametri precedentemente calcolati erano elevati al quadrato;

$$Ds(\sigma) = \sqrt{\frac{\sum (xi - \bar{x})^2}{n-1}}$$



Nella distribuzione normale, in prima approssimazione, la proporzione dei campioni rispetto alla media ± la Deviazione standard è:

x ± 1 σ = 68%
x ± 2 σ = 95%
x ± 3 σ = 99%

Indici di variabilità

Errore standard: la deviazione standard indica la dispersione dei valori intorno alla media del campione estratto dalla popolazione; l'errore standard quantifica il grado di certezza col quale la media a la variabilità calcolata stima la vera media e la variazione nella popolazione da cui il campione è estratto. La stima è tanto più precisa quanto più la variabilità è piccola (Ds minima) e la numerosità del campione è elevato (n).

$$Es = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}}{\sqrt{n}}$$

Indici di variabilità

Limiti fiduciali: esprimono l'ampiezza della distribuzione (di quanto può variare la media Dell'universo di osservazione)

$$Lf = \bar{x} \pm (t \times Es)$$

t = coefficiente del livello di probabilità parametrato in funzione dei gradi di libertà riferito ad una distribuzione normale

G.L.	p=0,05 Liv.95%	p=0,01 Liv.99%
1	12,706	63,657
2	4,303	9,925
3	3,182	5,841
4	2,776	4,604
5	2,571	4,032
6	2,447	3,707
7	2,365	3,499
8	2,306	3,550
9	2,262	3,250
10	2,228	3,169
11	2,201	3,106
12	2,179	3,055
13	2,160	3,012
14	2,145	2,977
15	2,131	2,947
16	2,120	2,921
17	2,110	2,898
18	2,101	2,878
19	2,093	2,861
20	2,086	2,845
21	2,080	2,831
22	2,074	2,819
23	2,069	2,807
24	2,064	2,797
25	2,060	2,787
26	2,056	2,779
27	2,052	2,771
28	2,048	2,763
29	2,045	2,756
30	2,042	2,750
40	2,021	2,704
60	2,000	2,660
120	1,980	2,617
∞	1,960	2,576

Distribuzione T

	Trattamento		
Risultato	A	B	Totali (a+b)
Positivo	60	2	62
Negativo	37	11	48
Totali (x+y)	97	13	110

Osservati

Calcolo degli attesi:
Nel complesso nei due trattamenti (x) si è ottenuto 60+2=62 successi su 110 eventi; la percentuale di successo totale è quindi 62:110=x:100, da cui x=62x100/110=56,4%
Per cui gli attesi positivi nel trattamento (A) saranno il 56,4% di 97, cioè 54,7 (55), mentre su 13 positivi nel trattamento (B) saranno il 56,4% di 13, cioè 7,33 (7).
Per differenza si calcolano i negativi, rispettivamente per (A) 97-55=42 e per (B) 13-7=6

	Trattamento		
Risultato	A	B	Totali (a+b)
Positivo	55	7	62
Negativo	42	6	48
Totali (x+y)	97	13	110

Attesi

Trattamento — Osservati

Risultato	A	B	Totali (a+b)
Positivo	60	2	62
Negativo	37	11	48
Totali (x+y)	97	13	110

Trattamento — Attesi

Risultato	A	B	Totali (a+b)
Positivo	55	7	62
Negativo	42	6	48
Totali (x+y)	97	13	110

Gradi di libertà	Pobabilità	
	P=0,05	P=0,01
1	3,841	6,235
2	5,991	9,210
3	7,815	11,345
4	9,488	13277
5	11,070	15,086
6	12,592	16,812
7	14,067	18,475
…	…	…

$$\chi^2 = \sum_{eventi=1}^{eventi=n} \frac{(osservati - attesi)^2}{attesi}$$

$$\chi^2 = \frac{(60-55)^2}{55} + \frac{(2-7)^2}{7} + \frac{(37-42)^2}{42} + \frac{(11-6)^2}{6} = 8,78$$

Significatività e regressione

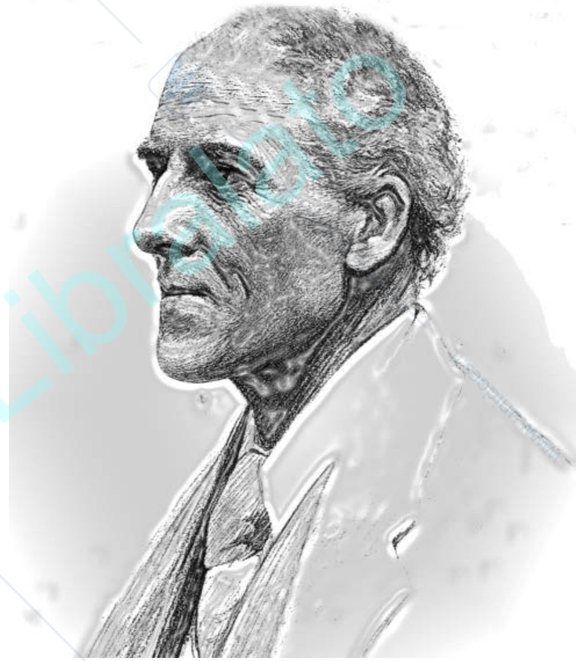
Test del T di student (confronto fra due gruppi);

t = differenza tra le medie/ errore standard della differenza tra medie

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\frac{\sum (x_1 - \bar{x}_1)^2}{(n_1 - 1)}}{n_1} + \frac{\frac{\sum (x_2 - \bar{x}_2)^2}{(n_2 - 1)}}{n_2}}}$$

Altri test (divisi in test parametrici e non parametrici) sono: ANOVA, test F, Test Z, ….

Karl Pearson (1857-1936),
nonostante il carattere "duro" ed estremista, ebbe, per i suoi meriti, una brillante carriera nelle istituzioni accademiche del tempo.
Tra i suoi numerosi studi vanno ricordati quelli sui modelli matematici e grafi descrittori dei fenomeni biologici che non hanno distribuzione normale (gaussiana) Nel 1900 egli inventò il **metodo del chi-quadrato**, che rappresenta la più antica procedura inferenziale tuttora in uso.

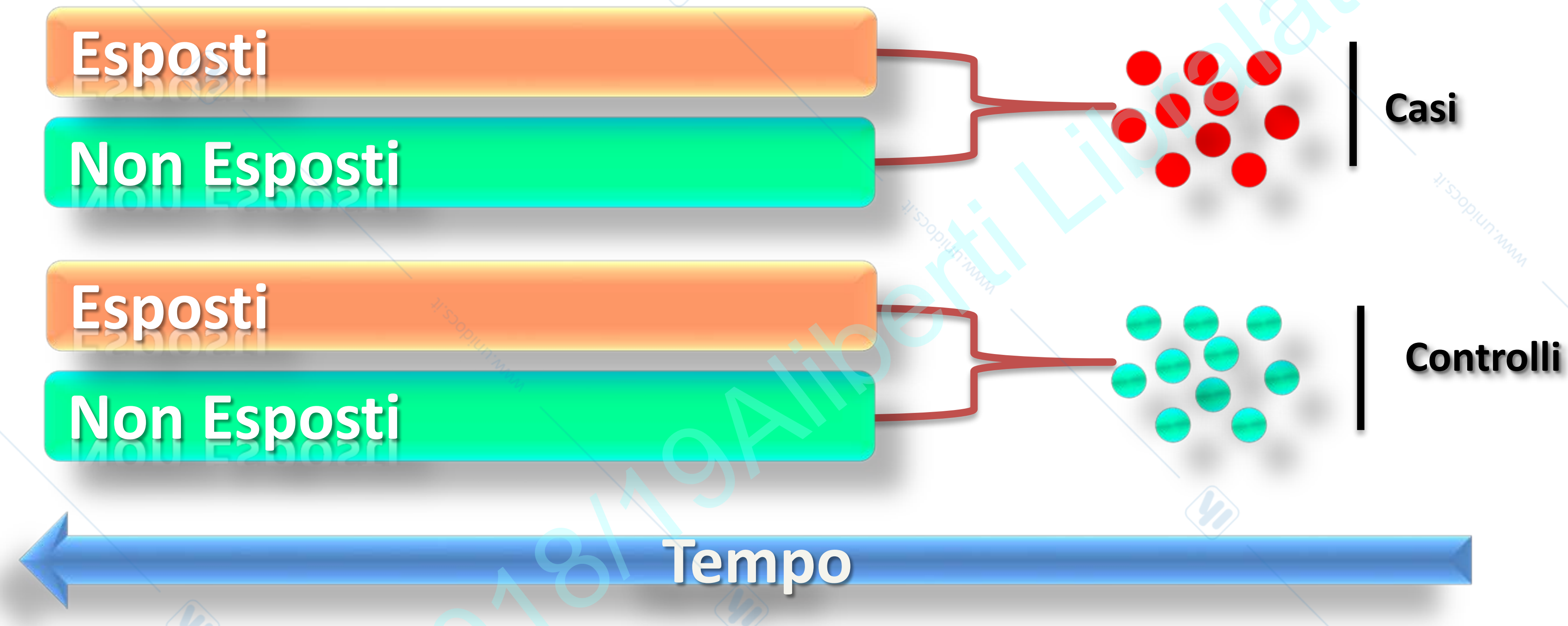


William S. Gosset (1876-1937),
operò nel reparto sperimentale della fabbrica di birra Guinnes ove sperimentò vari metodi statistici per raffrontare birre preparate in vario modo, le diverse materie prime impiegate (orzo, luppolo) ed i metodi di coltura e conservazione, trattamento.
Ideò il test **t** per valutare la migliore varietà di orzo per la produzione della birra scura (la famosa "stout"). Si narra che ottenne dall'azienda l'autorizzazione a pubblicare i suoi studi solo in forma anonima; perciò utilizzò lo pseudonimo "a student of statistics".
Il test **t**, infatti, è in suo onore detto **t di student**

Studio prospettivo (di coorte)

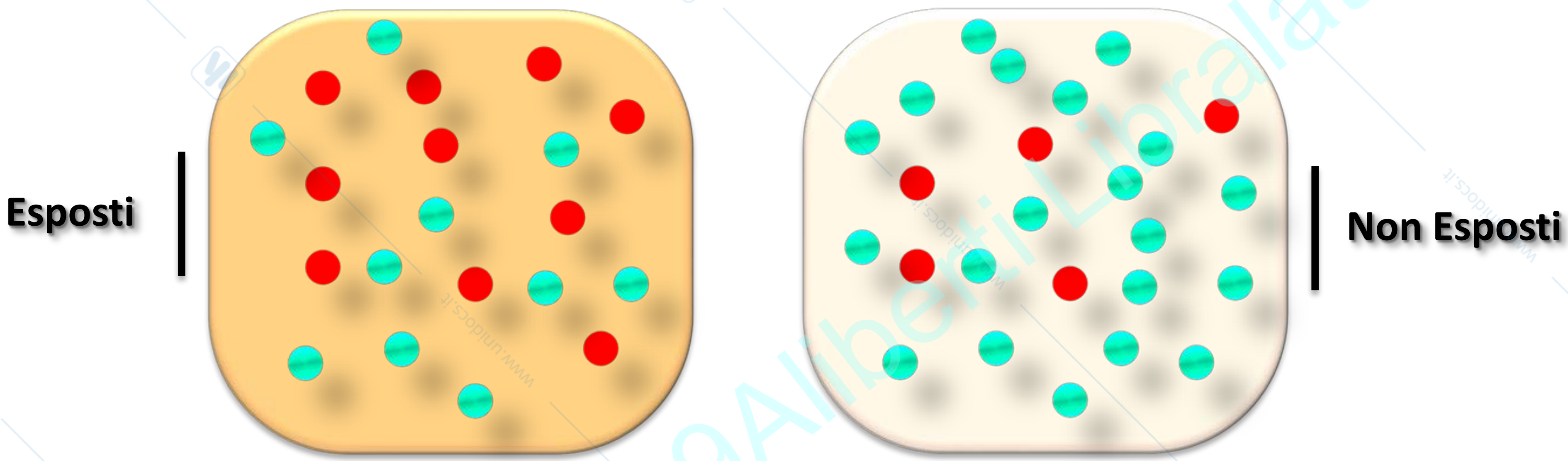
	Malati	Non Malati	Totale
Esposti	a	b	a+b
Non esposti	c	d	c+d
Totale	a+c	b+d	

Studio retrospettivo *(caso-controllo)*



	Malati	Non Malati	Totale
Esposti	a	b	a+b
Non esposti	c	d	c+d
Totale	a+c	b+d	

Studio trasversale



	Malati	Non Malati	Totale
Esposti	a	b	a+b
Non esposti	c	d	c+d
Totale	a+c	b+d	

$$Rischio\ relativo = \frac{Tasso\ Esposti}{Tasso\ Non\ Esposti}$$

$$R_R = \frac{T_{(E)}}{T_{(NE)}}$$

0	1	∞
Associazione negativa	Nessuna associazione	Associazione positiva

$$Rischio\ attribuibile = Tasso\ Esposti - Tassi\ Non\ Esposti$$

$$R_{attr.} = T_e - T_{ne}$$

Tasso di casi eliminabili qualora venisse annullato il fattore considerato

Cause di malattia: le regole di John Stuart Mill

1 Regola dell'accordo
se esiste un **fattore comune** nelle circostanze in cui si verifica una malattia, questo fattore può essere la causa della malattia

3 Regola della variazione concomitante
Se un fattore ed una malattia evidenziano una relazione **dose-effetto**, lo stesso può essere causa di malattia

2 Regola della differenza
Se le circostanze in cui si verifica la malattia sono simili a quelle in cui la stessa non si verifica **ad eccezione di un fattore**, questo può esserne la causa

4 Regola dell'analogia
Se la distribuzione di una malattia è **simile** a quella di un'altra ben nota, entrambe possono avere lo stesso fattore causale

5 Regola del residuo
Se un fattore giustifica solo **parte** dei casi di una malattia, altri fattori devono essere giustificati per i casi **residui**

John Stuart Mill

(Pentonville, 20 maggio 1806 – Avignone, 8 maggio 1873), filosofo ed economista.

I modelli matematici

Le basi

S - I (Sensibili - Infetti)

S - I - R (Sensibili - Infetti - Rimossi)

S - E - I - R (Sensibili - Esposti - Infetti - Rimossi)

Le tipologie

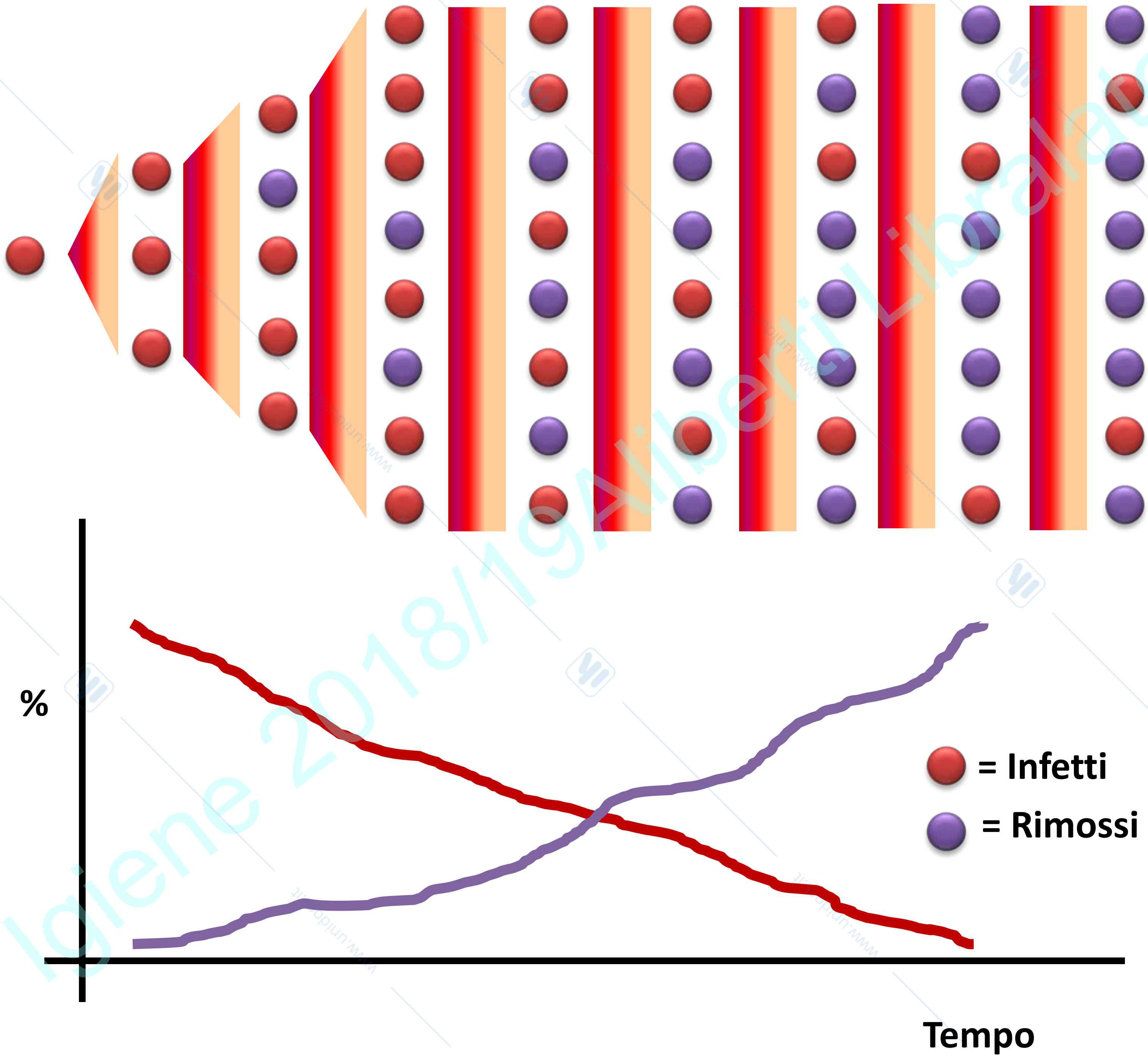
Modelli deterministici (risultati definiti)	Modelli stocastici (risultati probabili)

Modello di Reed e Frost

(Lowell Reed e Wade Hampton Frost)

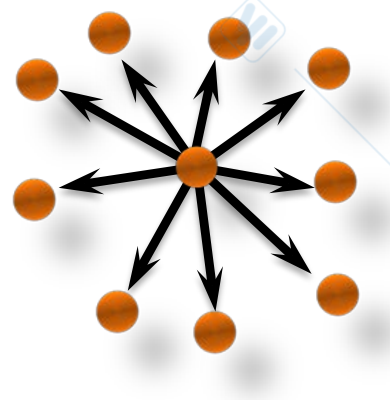
$$C_{t+1} = S_t \left(1 - q^{C_t}\right)$$

C = Numero di casi
t = Tempo
S = Sensibili
q = Probabilità di contatto non efficace la trasmissione: q = 1- p;
p = probabilità di contatto efficace per il contagio

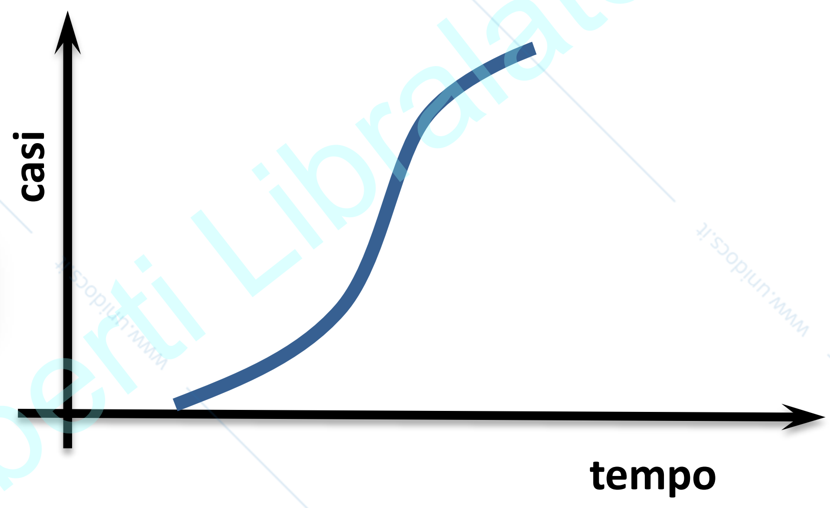


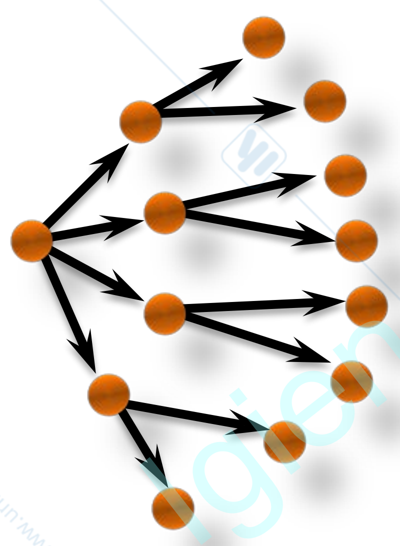
www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Epidemie

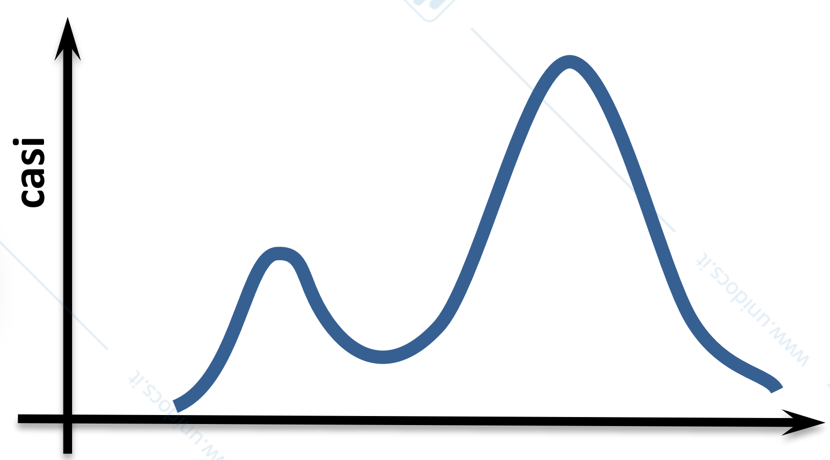


A sorgente comune





A propagazione



Tassi standardizzati: tassi resi omogenei rispetto ad una distribuzione “normalizzata” per fasce d’età o altra caratteristica della popolazione.

Permettono il confronto tra *universi di osservazione* diversi.

Esempio:

Popolazione A: tasso di incidenza della patologia x= 9‰

Popolazione B: tasso di incidenza della patologia x= 8‰

Sono diversi?

Popolazione A:

1 fascia d'età = 250 individui ➔ 1 caso
2 " " = 250 " ➔ 2 casi
3 " " = 500 " ➔ 6 casi

T o t a l e 1000 individui tasso: 9‰

1 fascia d'età = 250 individui ➔ 1 caso
2 " " = 500 " ➔ 4 casi
3 " " = 250 " ➔ 3 casi

T o t a l e 1000 individui tasso: 8‰

Popolazione B:

non c'è differenza !!!

Nell'esempio citato la differente distribuzione per fasce d'età nei due universi di osservazione è alla base della differenza riscontrata.
Da ciò la necessità dei tassi standardizzati.
In funzione delle caratteristiche della patologia in esame, più di una diversità tra gli universi considerati può influenzare gli studi; quali, ad esempio, il sesso, l'etnia, ecc..

Ovviamente nei casi in cui la malattia si distribuisce in modo diverso tra i sessi, le etnie, le età, ecc. delle popolazioni in cui si diffonde e le due popolazioni in esame sono tra loro incoerenti per i diversi fattori può essere necessario effettuare più standardizzazioni.