

Lez. 1c

# Rischiosità

# Rischiosità

	Infarto miocardico		
Gruppo	Sì	No	
Placebo	189	10845	11034
Aspirina	104	10933	11037

Fonte: Preliminary Report: Findings from the Aspirin Component on the Ongoing Physicians' Health Study. *N. Engl. J. Med.*, 318, 262-264 (1988)

*The Physicians' Health Study is a randomized, double-blind, placebo-controlled trial testing two primary-prevention hypotheses:*

- (1) whether 325 mg of aspirin (as Bufferin, supplied by Bristol-Myers Products) taken every other day reduces mortality from cardiovascular disease, and*
- (2) whether 50 mg of beta carotene (as Lurotin, supplied by BASF) taken on alternate days decreases the incidence of cancer*

# Rischiosità

	Infarto miocardico		
Gruppo	Sì	No	
Placebo	189	10845	11034
Aspirina	104	10933	11037

Fonte: Preliminary Report: Findings from the Aspirin Component on the Ongoing Physicians'Health Study. *N. Engl. J. Med.*, 318, 262-264 (1988)

*11,037 physicians were assigned **at random** to receive aspirin and 11,034 to receive aspirin placebo*


(cioè, per ognuna delle 22071 persone ho tirato una moneta equilibrata: T->Aspirina, C->Placebo; ne è venuto fuori 11037 volte T e 11034 volte C)


Studio **prospettico**.

# Rischiosità

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289		
Aspirina (T)	0.00942285		


$$p_C = \frac{189}{11034}$$



$$p_T = \frac{104}{11037}$$

$|p_T - p_C| = 0.00770602$ ; **riduzione assoluta di rischio**

# Rischiosità

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289		
Aspirina (T)	0.00942285		


$$p_C = \frac{189}{11034}$$

$$p_T = \frac{104}{11037}$$


$|p_T - p_C| = 0.00770602$ ; **riduzione assoluta di rischio**

$p_T = 0.4500$ ;  $p_C = 0.4577 \Rightarrow |p_T - p_C| = 0.0077!!!!$

# Rischiosità

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289		
Aspirina (T)	0.00942285		


$$p_C = \frac{189}{11034}$$

$$p_T = \frac{104}{11037}$$

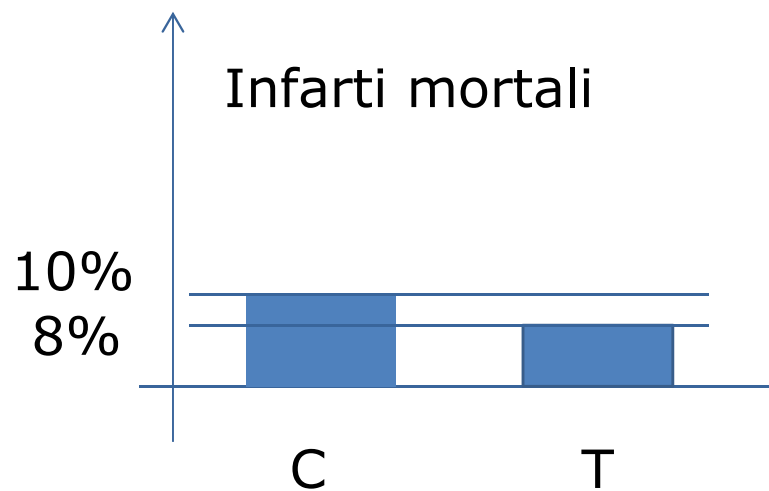
$|p_T - p_C| = 0.00770602$ ; **riduzione assoluta di rischio**

$$\frac{p_T}{p_C} = \frac{P(IM|T)}{P(IM|C)} = 0.550$$

**rischio relativo**

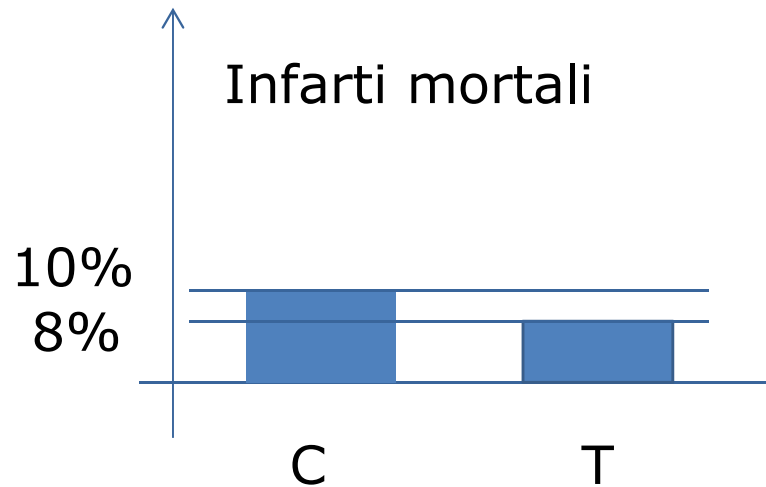
**Nel campione**  $p_T \approx 0.55p_C$  cioè l'aspirina sembra dimezzare il rischio di infarto miocardico.

# Odds ratio



	Decesso		
Gruppo	No	Sì	
T	92	8	100
C	90	10	100

# Odds ratio



	Decesso		
Gruppo	No	Sì	
T	92	8	100
C	90	10	100

*Odds*

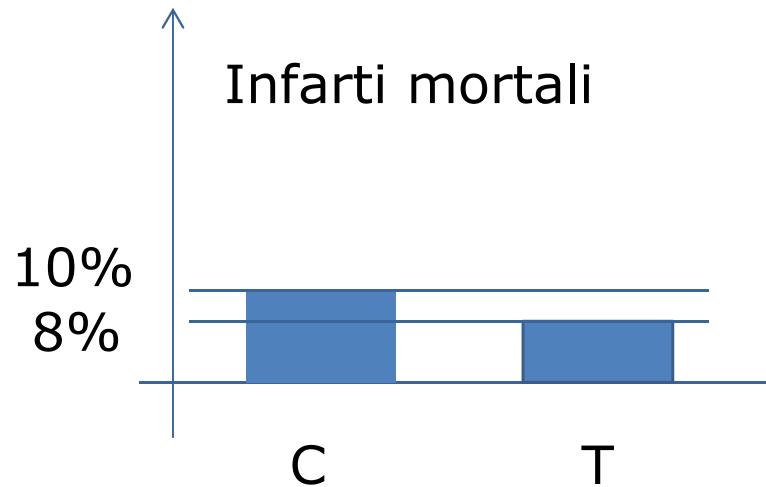
Quota a favore della sopravvivenza in T:

$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$$

**in T, per 1 che muore 11.5 persone sopravvivono**



# Odds ratio



	Decesso		
Gruppo	No	Sì	
T	92	8	100
C	90	10	100

**Odds**

Quota a favore della sopravvivenza in T:

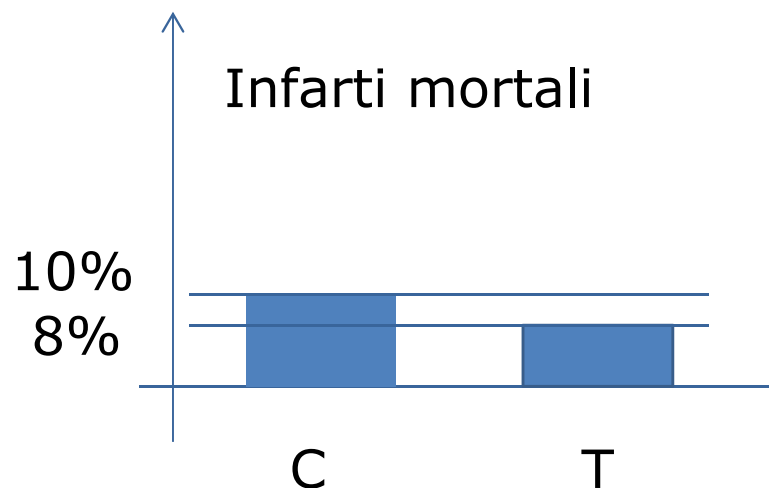
$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$$

Quota a favore della sopravvivenza in C:

$$\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.90}{0.10} = 9$$

**in C, per 1 che muore 9 persone sopravvivono**

# Odds ratio



	Decesso		
Gruppo	No	Sì	
T	92	8	100
C	90	10	100

**Odds**

Quota a favore della sopravvivenza in T:

$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$$

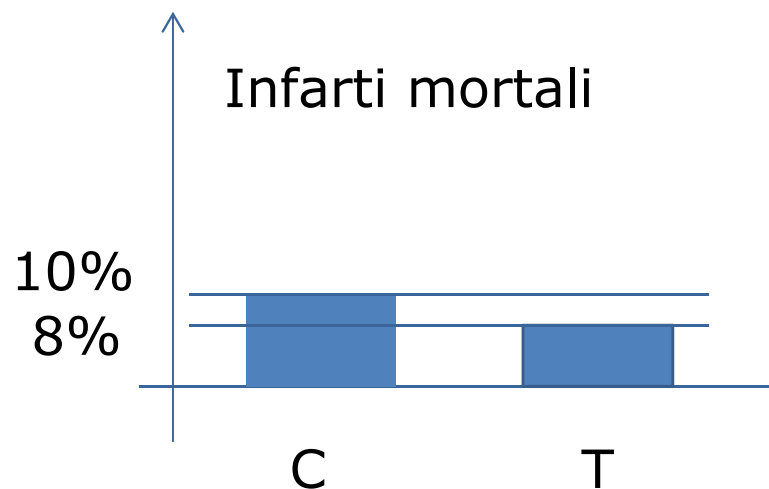
Quota a favore della sopravvivenza in C:

$$\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.90}{0.10} = 9$$

$$OR = \frac{\text{odds}|T}{\text{odds}|C} = \frac{11.5}{9} = \frac{92/8}{90/10} = 1.278$$

**Odds ratio**

# Odds ratio



	Decesso		
Gruppo	No	Sì	
<b>T</b>	46	4	<b>50</b>
<b>C</b>	135	15	<b>150</b>

Quota a favore della sopravvivenza in T:

$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$$

Quota a favore della sopravvivenza in C:

$$\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.90}{0.10} = 9$$

$$OR = \frac{\text{odds}|T}{\text{odds}|C} = \frac{11.5}{9} = \frac{46/4}{135/15} = 1.278$$

Odds ratio

# Odds ratio

## Es. 2

Quota a favore della sopravvivenza in T:

$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$$

Quota a favore della sopravvivenza in C:

$$\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.90}{0.10} = 9$$

$$OR = \frac{11,5}{9} = 1,278$$

quanto più è lontano da 1  
tanto maggiore è la  
differenza tra T e C

## Es. 1

Quota a favore della sopravvivenza in T:

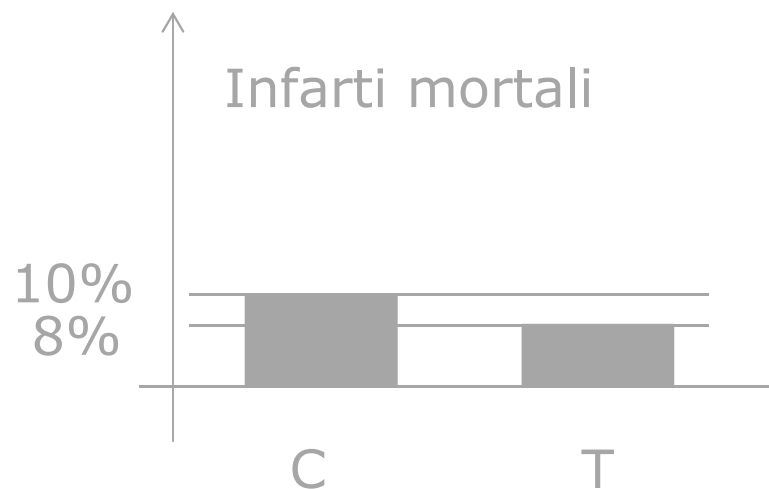
$$\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.75}{0.25} = 3$$

Quota a favore della sopravvivenza in C:

$$\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.25}{0.75} = \frac{1}{3}$$

$$OR = \frac{3}{1/3} = 9$$

# Odds ratio



	Decesso		
Gruppo	No	Sì	
<b>T</b>	46	4	<b>50</b>
<b>C</b>	135	15	<b>150</b>

Quota a favore della sopravvivenza (No) in T:  $\frac{P(\text{sopravv}|T)}{P(\text{morire}|T)} = \frac{0.92}{0.08} = 11.5$

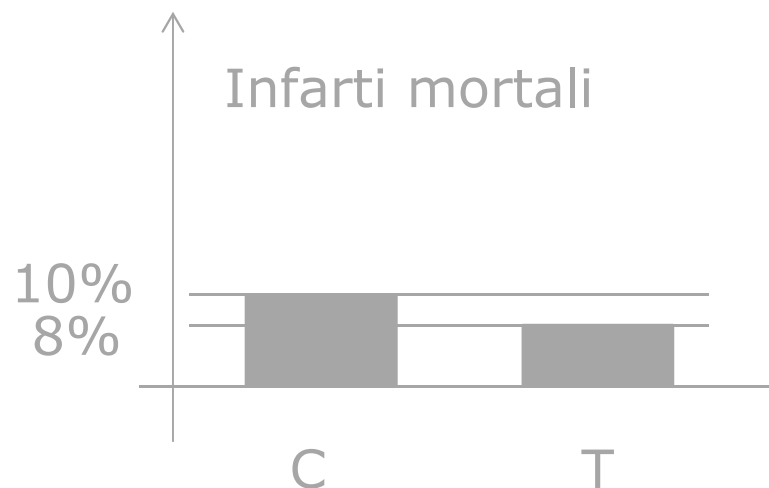
Quota a favore della sopravvivenza (No) in C:  $\frac{P(\text{sopravv}|C)}{P(\text{morire}|C)} = \frac{0.90}{0.10} = 9$

$$OR = \frac{\text{odds}(No)|T}{\text{odds}(No)|C} = \frac{11.5}{9} = \frac{46/4}{135/15} = \frac{46 \times 15}{135 \times 4} = 1.278$$

Odds ratio

# Odds ratio

Scambiamo righe con colonne



	Gruppo		
Decesso	T	C	
No	46	135	<b>181</b>
Sì	4	15	<b>29</b>
	50	150	

Quota a favore di T in No:

$$\frac{P(T|No)}{P(C|No)} = \frac{46/181}{135/181} = \frac{46}{135} = 0.341$$

Quota a favore di T in Sì:

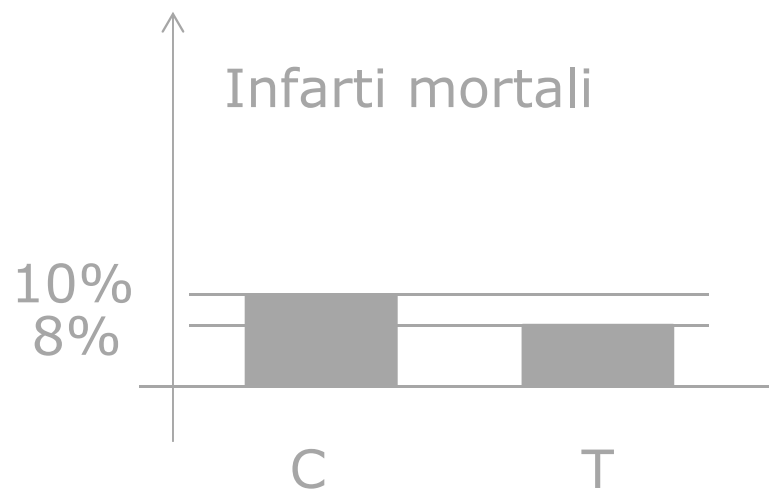
$$\frac{P(T|Sì)}{P(C|Sì)} = \frac{4/29}{15/29} = \frac{4}{15} = 0.267$$

$$OR = \frac{odds(T)|No}{odds(T)|Sì} = \frac{0.341}{0.267} = \frac{46/135}{4/15} = \frac{46 \times 15}{4 \times 135} = 1.278$$

**Odds ratio  
invariato!**

# Odds ratio

Scambiamo righe con colonne



	Gruppo		
Decesso	T	C	
No	46	135	<b>181</b>
Sì	4	15	<b>29</b>
	50	150	

$$RR = \frac{P(T|No)}{P(T|Sì)} = \frac{46/181}{4/29} = 1.842 \neq \frac{P(No|T)}{P(No|C)} = \frac{0.92}{0.90} = 1.022$$

	Decesso		
Gruppo	No	Sì	
T	46	4	<b>50</b>
C	135	15	<b>150</b>

# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

$p$  = probabilità di IM

$$p_C = \frac{189}{11034}$$

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289	0.9828711	
Aspirina (T)	0.00942285	0.9905772	

$$p_T = \frac{104}{11037}$$

rischio relativo:

$$\frac{P(IM|T)}{P(IM|C)} = \frac{p_T}{p_C} = 0.550$$



# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

$p$  = probabilità di IM

$$p_C = \frac{189}{11034}$$

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289	0.9828711	
Aspirina (T)	0.00942285	0.9905772	

$$p_T = \frac{104}{11037}$$

rischio relativo:

$$\frac{P(IM|T)}{P(IM|C)} = \frac{p_T}{p_C} = 0.550$$

**Odds ratio:**

$$\frac{p_T/(1-p_T)}{p_C/(1-p_C)} = 0.546$$

# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

$p$  = probabilità di IM

$$p_C = \frac{189}{11034}$$

$$p_T = \frac{104}{11037}$$

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	0.0171289	0.9828711	
Aspirina (T)	0.00942285	0.9905772	

rischio relativo:

$$\frac{P(IM|T)}{P(IM|C)} = \frac{p_T}{p_C} = 0.550$$

**Odds ratio:**

$$\frac{p_T/(1-p_T)}{p_C/(1-p_C)} = 0.546 = \frac{p_T}{p_C} \times \frac{1-p_C}{1-p_T} \approx \frac{p_T}{p_C}$$

$$= 0.992 \approx 1$$

# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

Studio **randomizzato**: i 22 071 medici partecipanti alla sperimentazione sono stati suddivisi, con una assegnazione casuale, tra C e T. Si è trattato di uno studio **prospettico**, cioè pianificato e poi attuato e monitorato nel tempo.

(<http://www.nejm.org/doi/full/10.1056/NEJM198907203210301#t=article>)

# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

Studio **randomizzato**: i 22 071 medici partecipanti alla sperimentazione sono stati suddivisi, con una assegnazione casuale, tra C e T. Si è trattato di uno studio **prospettico**, cioè pianificato e poi attuato e monitorato nel tempo.

(<http://www.nejm.org/doi/full/10.1056/NEJM198907203210301#t=article>)

Gruppo	IM	Controllo
Fumatore	172	173
Non F.	90	346
	<b>262</b>	<b>≈262×2</b>

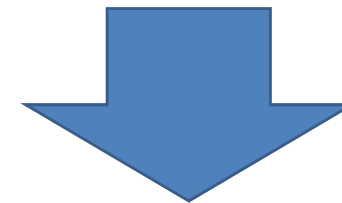
Studio **retrospettivo**, cioè dati dal passato, di tipo **caso-controllo**. Ogni *caso* accoppiato con 2 *controlli* ⇒ **totali di colonna fissate.**

# Caso-controllo

Gruppo	IM	Controllo
Fumatore (T)	172	173
Non F. (C)	90	346
	<b>262</b>	<b>≈262×2</b>

Fonte: A. Gramenzi et al. (1989) *J. Epidemiol. and Commun. Health*, 43: 214-217

Ogni *caso* (262 donne di età <69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) *accoppiato con 2 controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**



Le 781( **≈262×3**) u.s. **non** "sono state scelte a caso, e tra di loro è saltato fuori che 262 hanno avuto IM"

$$\frac{262}{262 \times 3} = \frac{1}{3}$$

**non** è una stima della frazione di infartuati nella popolazione

# Caso-controllo

Gruppo	IM	Controllo
Fumatore (T)	172	173
Non F. (C)	90	346
	<b>262</b>	<b>≈262×2</b>

Ogni *caso* (262 donne di età <69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) *accoppiato con 2 controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

Fonte: A. Gramenzi et al. (1989) *J. Epidemiol. and Commun. Health*, 43: 214-217

Delle u.s. **so già se IM o no**, e "misuro" se F o no (nell'esempio di prima, essendo prospettico io delle u.s. a priori non so nulla, poi assegno T o C a caso e poi misuro IM (e F) o no: è esattamente l'opposto).

# Caso-controllo

Gruppo	IM	Controllo
Fumatore (T)	172	173
Non F. (C)	90	346
	<b>262</b>	<b>≈262×2</b>

Ogni *caso* (262 donne di età <69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) *accoppiato con 2 controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

Fonte: A. Gramenzi et al. (1989) *J. Epidemiol. and Commun. Health*, 43: 214-217

Delle u.s. **so già se IM o no**, e "misuro" se F o no (nell'esempio di prima, essendo prospettico io delle u.s. a priori non so nulla, poi assegno T o C a caso e poi misuro IM (e F) o no: è esattamente l'opposto).

$$P(F|IM) = \frac{172}{262} = 0.656$$

$$P(IM|F) = \frac{172}{172 + 173}$$



**no  
RR!**

# Caso-controllo

Gruppo	IM	Controllo
Fumatore (T)	172	173
Non F. (C)	90	346
	<b>262</b>	<b>≈262×2</b>

Ogni *caso* (262 donne di (\*) età < 69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) *accoppiato con 2 controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

$$OR = \frac{172 \times 346}{90 \times 173} = 3.82$$

**posso calcolarlo** perchè prescinde da chi funge da variabile risposta ("misura")

$$OR = \frac{p_T}{p_C} \times \frac{1-p_C}{1-p_T} \approx \frac{p_T}{p_C} = RR \text{ se } p_T \text{ e } p_C \text{ piccole! (*)}$$



Delle u.s. **so già se IM o no**, e "misuro" se F o no (nell'esempio di prima, essendo prospettico io delle u.s. a priori non so nulla, poi assegno T o C a caso e poi misuro IM (e F) o no: è esattamente l'opposto).



# Odds ratio e rischio relativo

	Infarto miocardico		
Gruppo	Sì	No	
Placebo (C)	189	10845	11034
Aspirina (T)	104	10933	11037

**RR & OR**

Studio **randomizzato**: i 22 071 medici partecipanti alla sperimentazione sono stati suddivisi, con una assegnazione casuale, tra C e T. Si è trattato di uno studio **prospettivo**, cioè pianificato e poi attuato e monitorato nel tempo.

(<http://www.nejm.org/doi/full/10.1056/NEJM198907203210301#t=article>)

Gruppo	IM	Controllo
Fumatore	172	173
Non F.	90	346
	<b>262</b>	<b>≈262×2</b>

Studio **retrospettivo**, cioè dati dal passato, di tipo **caso-controllo**. Ogni *caso* accoppiato con 2 *controlli* ⇒ **totali di colonna fissate.**

**solo OR !**

# I tassi

$$\text{Tasso di mortalità globale (nel 2015)} = \frac{\text{numero di decessi}}{\text{totale della popolazione mondiale}} \times k$$

$$\text{Tasso di natalità globale (nel 2015)} = \frac{\text{numero di nati vivi}}{\text{totale della popolazione mondiale}} \times k$$

$$\text{Tasso di natalità generale (nel 2015)} = \frac{\text{numero di decessi}}{\text{num. di donne tra i 15 ed i 44 anni}} \times k$$

(v. sezione 3.7 per ulteriori definizioni)

# I tassi

$$\text{Tasso di mortalità globale (nel 2015)} = \frac{\text{numero di decessi}}{\text{totale della popolazione mondiale}} \times k$$

$$\text{Tasso di natalità globale (nel 2015)} = \frac{\text{numero di nati vivi}}{\text{totale della popolazione mondiale}} \times k$$

$$\text{Tasso di natalità generale (nel 2015)} = \frac{\text{numero di decessi}}{\text{num. di donne tra i 15 ed i 44 anni}} \times k$$

Dal sito dell'Unicef: Nel 2010, il tasso di mortalità dei decessi infantili (0-5 anni) si è attestato a **57 decessi ogni 1.000 nati vivi**.

$$\frac{\text{numero di decessi nei bambini fino a 5 anni}}{\text{num. totale di nati vivi}} = 0.057$$

$$k = 1000$$

# I tassi (su base annua)

$$\text{Tasso di } \mathbf{incidenza} = \frac{\text{numero di casi (nuovi) riportati (in un anno)}}{\text{totale della popolazione (in quell'anno)}} \times k$$

$$\text{Tasso di } \mathbf{prevalenza} = \frac{\text{numero di malati ad un certo anno}}{\text{totale della popolazione nell'anno}} \times k$$

Nuovi dell'anno + quelli  
già diagnosticati

Nel 2014 l'incidenza dell'HIV è stata di 6.1 casi per 100.000 residenti.  
(Fonte: Istituto Superiore di Sanità)

# Esercizio 7 p. 108

In un certo anno negli USA 2 416 000 decessi, di cui 700 142 per malattie cardiache.

- a) Determinare il tasso di mortalità per malattie cardiache ( $k = 100$ )
- b) Scegliendo a caso tre persone decedute, qual è la probabilità che nessuna sia morta per una malattia cardiaca?

# Esercizio 7 p. 108

In un certo anno negli USA 2 416 000 decessi, di cui 700 142 per malattie cardiache.

a) Determinare il tasso di mortalità per malattie cardiache ( $k = 100$ )

b) Scegliendo a caso tre persone decedute, qual è la probabilità che nessuna sia morta per una malattia cardiaca?

a) 
$$\frac{700\,142}{2\,416\,000} = 0.290 \quad \Rightarrow \text{tasso} = 29 \text{ ogni } 100 \text{ decessi}$$

b) Prob. che un morto scelto a caso sia deceduto per una m.c. è 0.29.

$$\Rightarrow (1 - 0.29)^3 = 0.358 \quad (\text{tutti e tre morti per altra causa})$$

# Esercizio 7 p. 108

In un certo anno negli USA 2 416 000 decessi, di cui 700 142 per malattie cardiache.

a) Determinare il tasso di mortalità per malattie cardiache ( $k = 100$ )

b) Scegliendo a caso tre persone decedute, qual è la probabilità che nessuna sia morta per una malattia cardiaca?

a)  $\frac{700\,142}{2\,416\,000} = 0.290 \Rightarrow \text{tasso} = 29 \text{ ogni } 100 \text{ decessi}$

3 estrazioni **con reimmissione**  
tra i deceduti

b) Prob. che un morto scelto a caso sia deceduto per una m.c. è 0.29.

$\Rightarrow (1 - 0.29)^3 = 0.358$  (tutti e tre morti per altra causa)

# Sul libro

Riduzione del rischio, (NNT), Rischio relativo e Odds Ratio: Cap. 3.6

Tassi: Cap. 3.7



# Statistica


Statistica descrittiva

**GRAFICI**

# Scopo e funzione

Conoscenza quantitativa di fenomeni collettivi

Gli elementi del collettivo si chiamano *unità statistiche* e formano una *popolazione*.




**Reale:** *“tutti gli studenti UniMI iscritti alla data....”*

**Ipotetica:** tutte le u.s. identificate da una lista di caratteristiche comuni. *“Tutti gli studenti UniMI” .... Tutte le merendine prodotte dalla tal ditta...*

# Scopo e funzione

Conoscenza quantitativa di fenomeni collettivi

Gli elementi del collettivo si chiamano *unità statistiche* e formano una *popolazione*.



**Reale:** *“tutti gli studenti UniMI iscritti alla data....”*

**Ipotetica:** tutte le u.s. identificate da una lista di caratteristiche comuni. *“Tutti gli studenti UniMI” .... Tutte le merendine prodotte dalla tal ditta...*

- Popolazione osservata per intero, es. *Censimento*.
- ***Campione rappresentativo***(\*): unità effettivamente sotto osservazione (*controllo di qualità, interviste...*).

# Scopo e funzione

Conoscenza quantitativa di fenomeni collettivi

Gli elementi del collettivo si chiamano *unità statistiche* e formano una *popolazione*.

**Reale:** *“tutti gli studenti UniMI iscritti alla data....”*

**Ipotetica:** tutte le u.s. identificate da una lista di caratteristiche comuni. *“Tutti gli studenti UniMI” .... Tutte le merendine prodotte dalla tal ditta...*

- Popolazione osservata per intero, es. *Censimento*.
- *Campione rappresentativo* (**DATI** amamente sotto osservazione (*controllo di qualità, interviste...*)).

# Scopo e funzione

Le caratteristiche che vengono osservate sulle unità statistiche sono chiamate *caratteri* o ***variabili statistiche***.

Variabile	Tipologia	Commento	Confronti
<b>Qualitativa o categorica</b>	<b>Nominale</b> (o sconnesso)	Ex.: Sesso, professione	= , ≠
	<b>Ordinale</b>	Esiste un ordinamento delle <b>modalità</b> : «sereno, poco nuvoloso, coperto»	= , ≠, <
<b>Quantitativa</b>	<b>Discreto</b>	“numero di ...” Ex. età	= , ≠, <
	<b>Continuo</b>	Ex. Altezza, età	= , ≠, <

# Scopo e funzione

**STATISTICA DESCRITTIVA: *organizzare* e *riassumere*** le informazioni su un fenomeno osservato in modo da semplificarne l'utilizzo



Paesi dell'Unione Europea	Omicidi volontari		Lesioni gravi		Furti		Delitti in materia di stupefacenti	
	Totale	% Minori	Totale	% Minori	Totale	% Minori	Totale	% Minori
Italia	1.974	3,49	27.862	2,51	72.058	9,05	48.320	2,79
Austria	126	5,60	122	13,10	33.020	18,90	21.085	8,40
Belgio	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Danimarca	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Finlandia	121	9,00	2.224	10,00	66.835	16,00	16.368	7,00
Francia	1.726	7,50	83.891	16,10	259.701	33,80	92.615	17,00
Germania	2.891	8,25	136.456	20,70	736.980	20,00	202.281	17,00
Grecia*	278	2,52	7.891	0,75	11.749	9,27	11.557	2,88
Irlanda	1	0,00	414	2,08	2.159	8,00	443	2,00
Lussemburgo	57	8,77	490	9,18	1.878	30,09	1.469	15,79
Paesi Bassi	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Portogallo	163	n.d.	932	1,90	71.332	4,30	7.947	1,80
Regno Unito (Inghilterra e Galles)	553	7,00	1.815	18,00	229.031	31,00	85.045	15,00
Regno Unito (Scozia)	83	12,00	1.015	9,10	18.583	9,20	5.054	4,30
Spagna	1.372	4,30	7.219	n.d.	95.998	19,89	17.380	6,47
Svezia	420	19,00	11.750	31,00	29.486	31,00	12.954	22,00

**STATISTICA INFERENZIALE: *generalizzare*** le informazioni, ricavando proprietà e leggi generali sulla base di dati rilevati solamente sul **campione rappresentativo**.

# Ci vuole un bel campione

per fare Statistica

Campione "volontario": tutti quelli che spontaneamente rispondono al sondaggio indetto oggi da La Provincia di Cremona on-line

Campione "casuale": tutti coloro che sono stati scelti **a caso**, cioè con la stessa probabilità di chiunque altro nella popolazione di riferimento di essere scelti.

# Ci vuole un bel campione

per fare Statistica

Si possono trarre conclusioni riguardanti il campione soltanto.

Campione "volontario": tutti coloro che rispondono al sondaggio indetto oggi da La Provincia di Cremona on-line

Potremo trarre conclusioni per l'intera popolazione di riferimento.

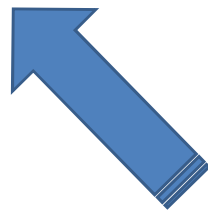
Campione "**casuale**": tutti coloro che sono stati scelti **a caso**, cioè con la stessa probabilità di chiunque altro nella popolazione di riferimento di essere scelti.



# Ci vuole un bel campione

per fare Statistica

Sistematico  
Di convenienza  
Stratificato  
A grappoli



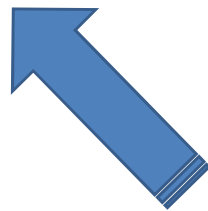
Campione "**casuale**": tutti coloro che sono stati scelti **a caso**, cioè con la stessa probabilità di chiunque altro nella popolazione di riferimento di essere scelti.

# Ci vuole un bel campione

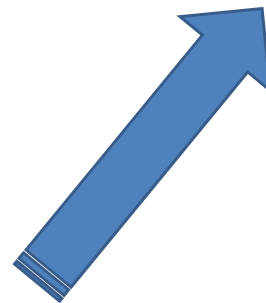
per fare Statistica

Errore di campionamento: frutto delle fluttuazioni casuali nei campioni.

Sistematico  
Di convenienza  
Stratificato  
A grappoli



Errore non dovuto al campionamento



Campione "**casuale**": tutti coloro che sono stati scelti **a caso**, cioè con la stessa probabilità di chiunque altro nella popolazione di riferimento di essere scelti.

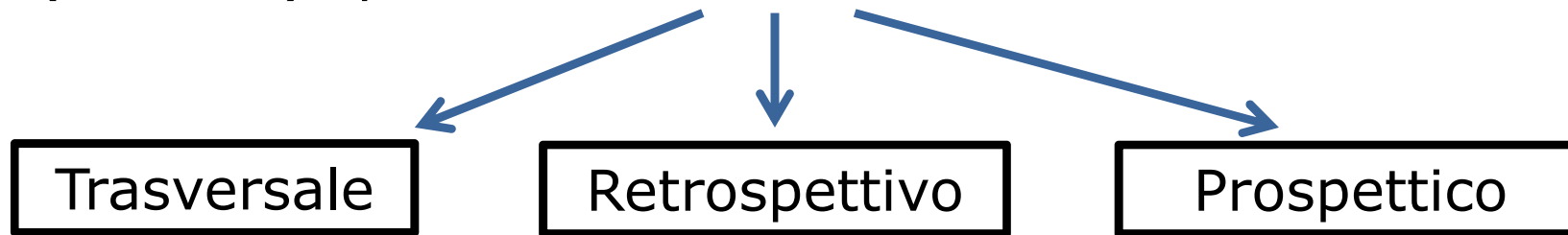
# Diversi tipi di studi

**Osservazionale:** osserviamo e misuriamo caratteri (variabili) specifiche

**Pianificato:** le unità statistiche vengono osservate e misurate dopo aver subito un trattamento

# Diversi tipi di studi

**Osservazionale:** osserviamo e misuriamo caratteri (variabili) specifiche



**Pianificato:** le unità statistiche vengono osservate e misurate dopo aver subito un trattamento

# Un esempio semplice

	<b>Età (y)</b>	<b>Peso (kg)</b>	<b>Altezza (m)</b>	<b>Sesso</b>	<b>Causa di morte</b>
1	79	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	51	80.1	1.91	M	Inc.
4	83	56.7	1.72	M	Nat.
5	72	75.0	1.81	M	M.C.
6	60	58.3	1.68	F	Tum.

# Dati categorici

Che sia nominale o ordinale (“sereno, poco nuvoloso, coperto, pioggia”), non ci sono molte elaborazioni possibili:

- Frequenze **assolute**

Genere	$n_i$		
F	2		
M	4		
<b>tot.</b>	<b>6</b>		

# Dati categorici

Che sia nominale o ordinale (“sereno, poco nuvoloso, coperto, pioggia”), non ci sono molte elaborazioni possibili:

- Frequenze assolute, **relative**

Genere	$n_i$	$f_i = n_i/n$	
F	2	2/6	
M	4	4/6	
<b>tot.</b>	<b>6</b>	<b>1</b>	

# Dati categorici

Che sia nominale o ordinale (“sereno, poco nuvoloso, coperto, pioggia”), non ci sono molte elaborazioni possibili:

- Frequenze assolute, relative e **percentuali**;

Genere	$n_i$	$f_i = n_i/n$	% = $100f_i$
F	2	$2/6=0.333$	33.3%
M	4	$4/6=0.667$	66.7%
<b>tot.</b>	<b>6</b>	<b>1</b>	<b>100%</b>

**distribuzioni di frequenze**



# Dati categorici

permettono di fare confronti  
tra gruppi di diversa numerosità

Che sia nominale o ordinale ("se  
nuvoloso, coperto, pioggia"), non sono molte  
elaborazioni possibili:

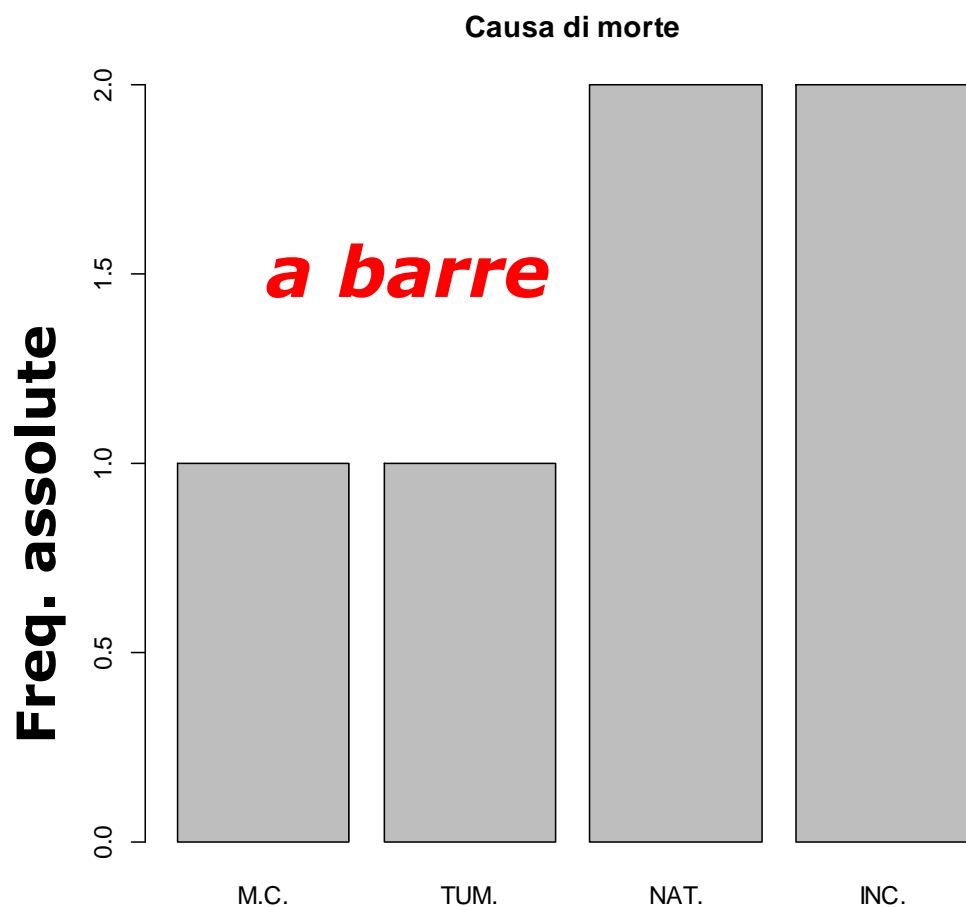
- Frequenze assolute, relative e **percentuali**;

Genere	$n_i$	$f_i = n_i/n$	% = $100f_i$
F	2	$2/6=0.333$	33.3%
M	4	$4/6=0.667$	66.7%
<b>tot.</b>	<b>6</b>	<b>1</b>	<b>100%</b>

**distribuzioni di frequenze**

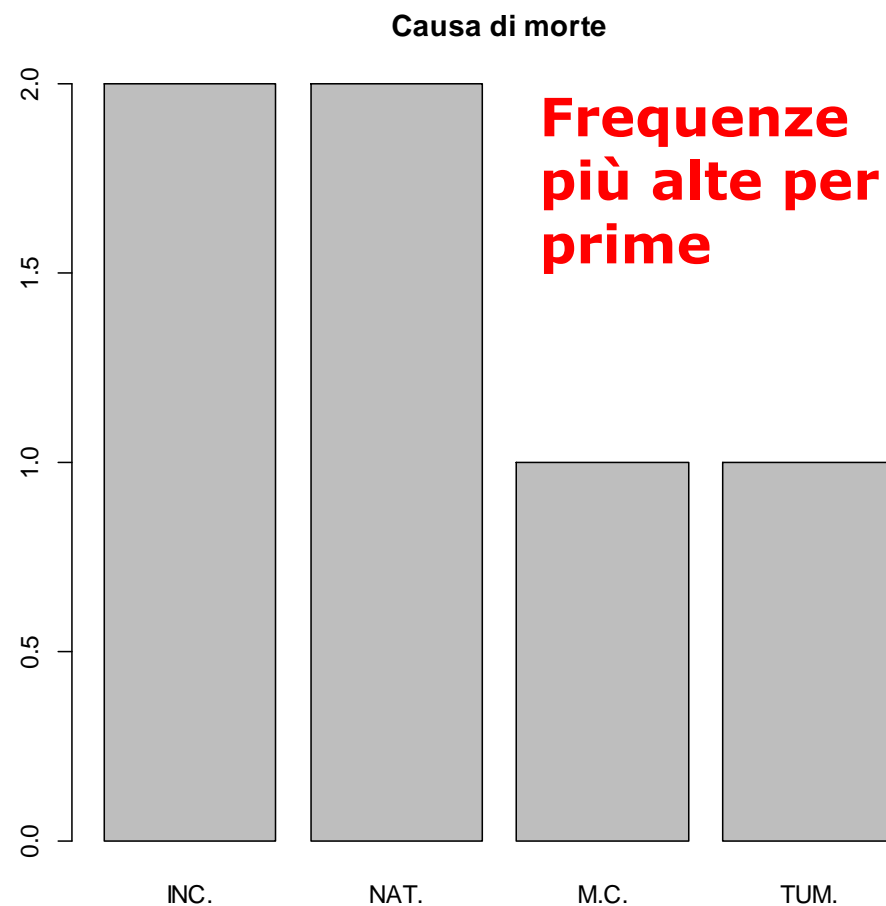
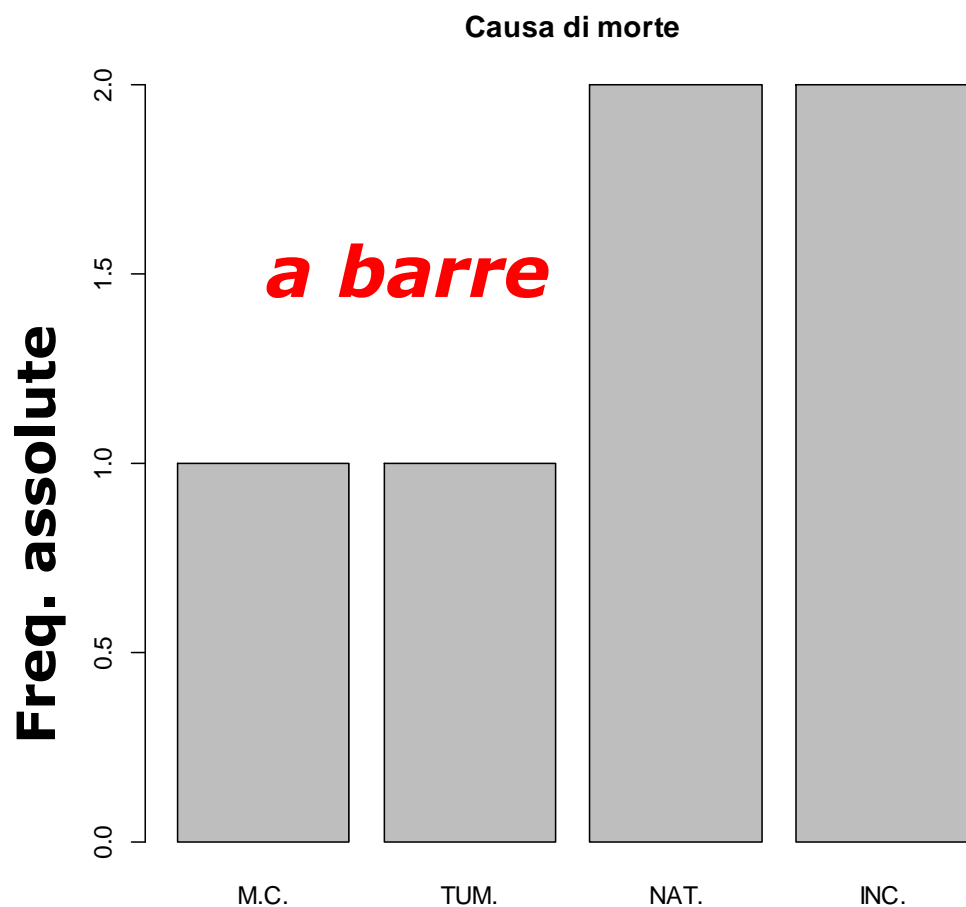
# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):



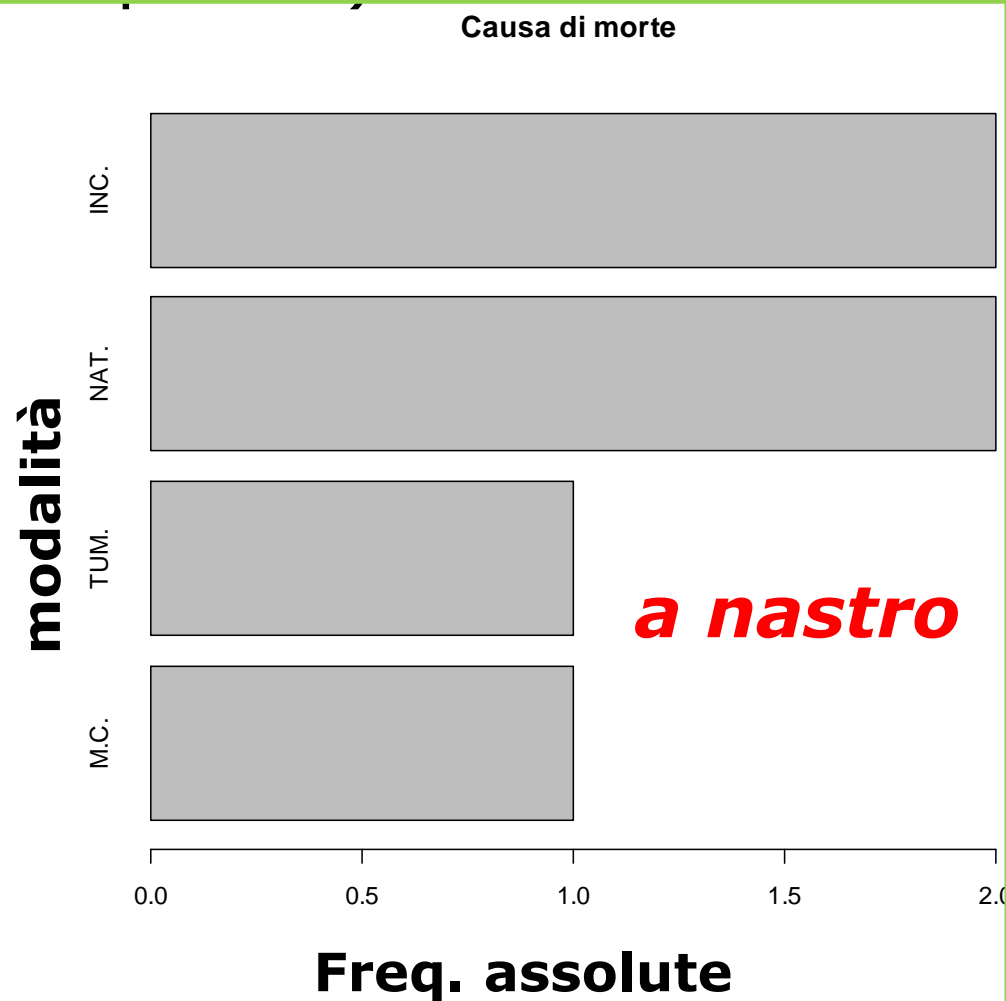
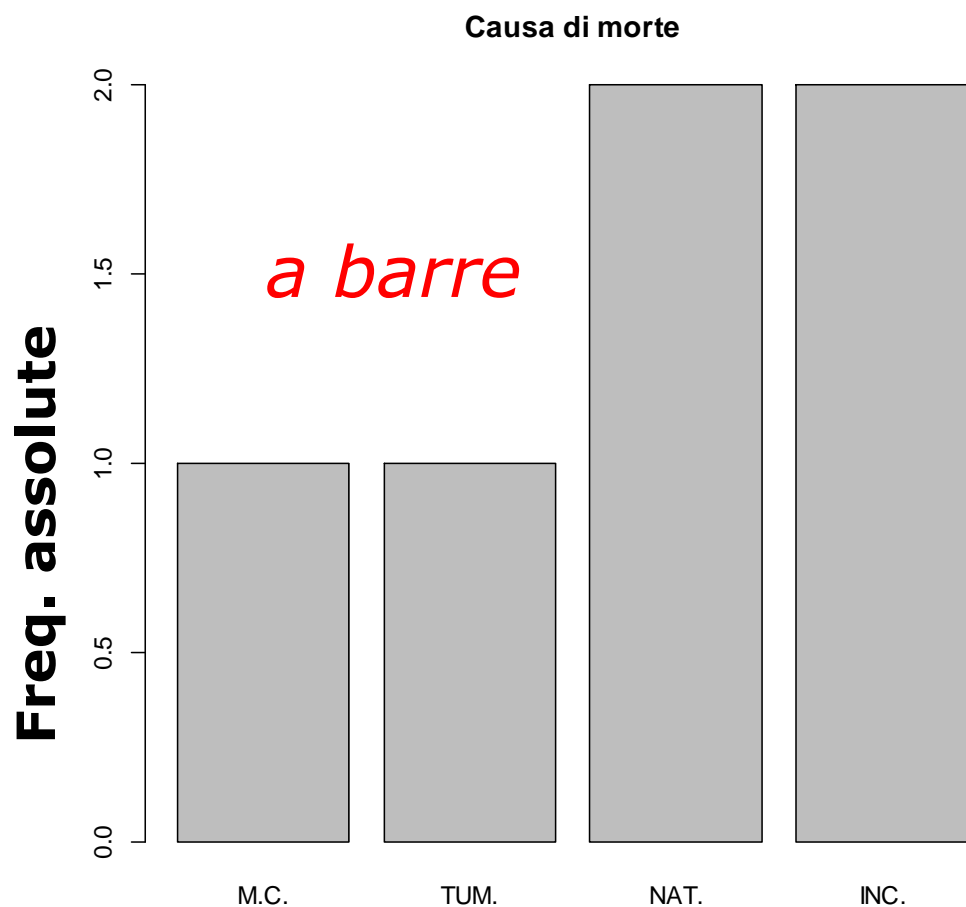
# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):

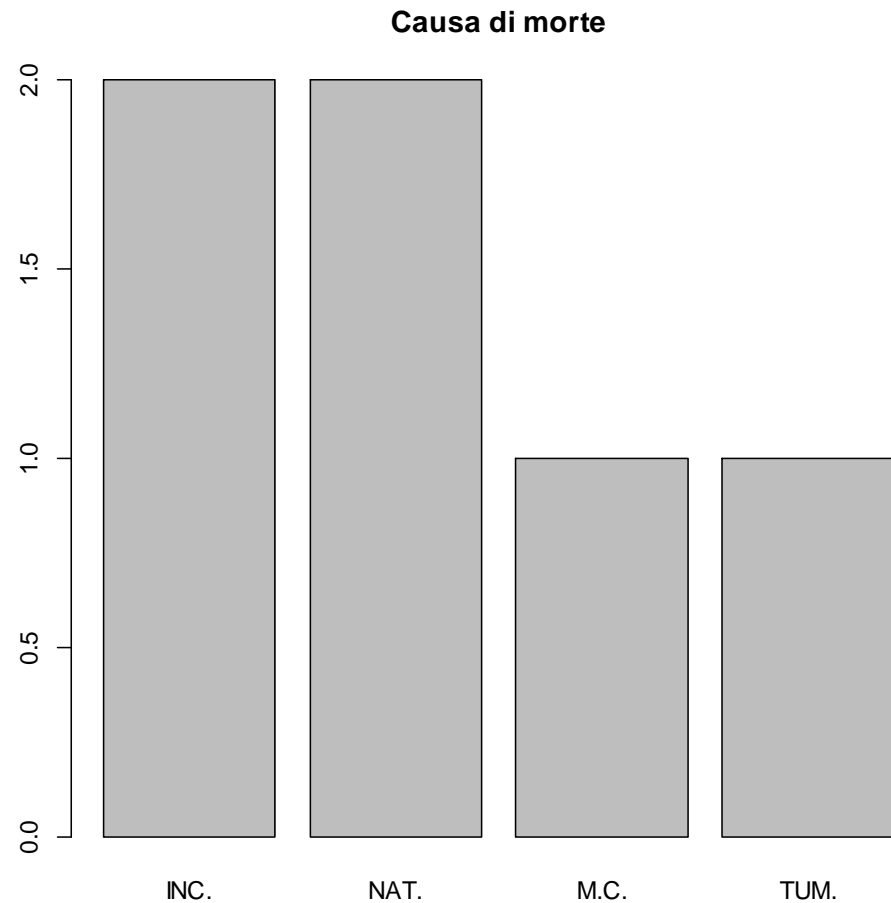
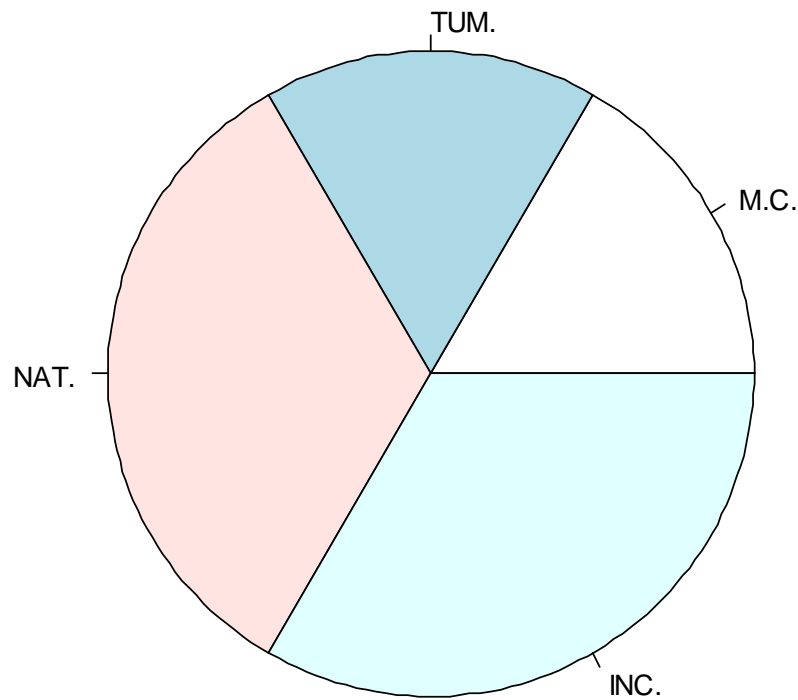


# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):



# Dati categorici



I diagrammi a torta sono una pessima rappresentazione, perché l'occhio è capace di giudicare bene le misure lineari, e meno bene le aree.

# Dati categorici

Voto all'esame: da 18 a 30L (variabile discreta)

# Dati categorici

Voto all'esame: da 18 a 30L

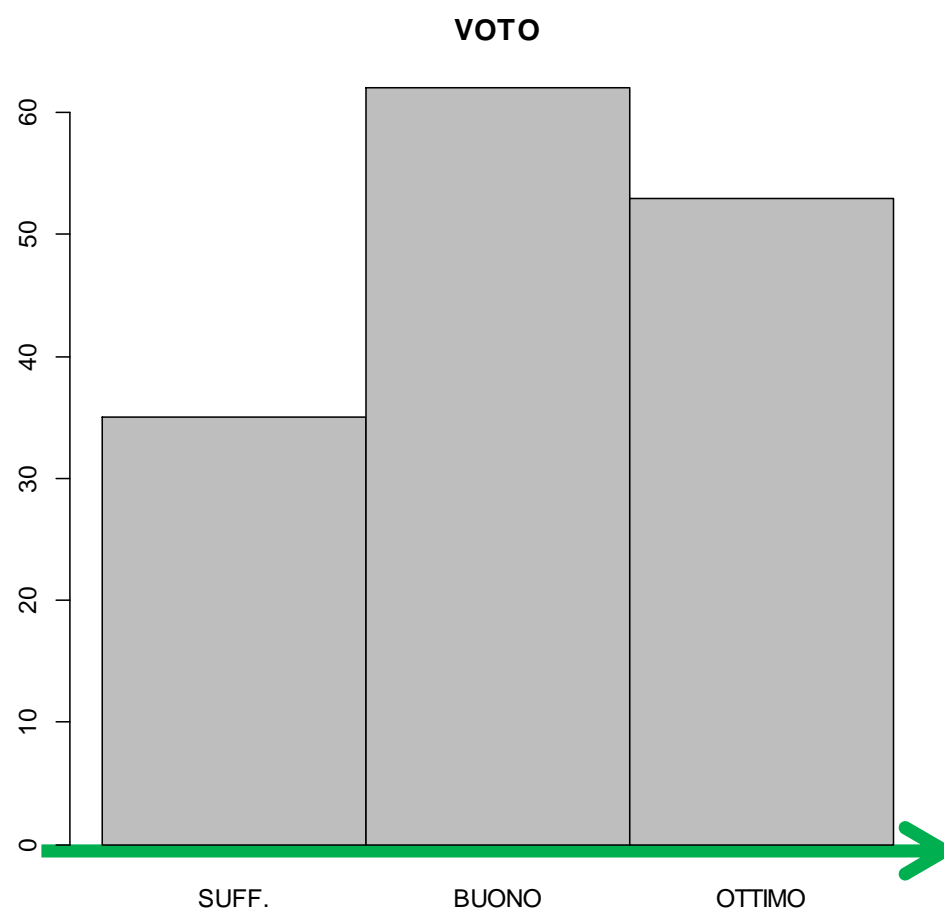
Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo

categorica **ordinale**



# Dati categorici

Voto all'esame: da 18 a 30L

Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo

Rendimento	$n_i$	$f_i = n_i/n$	% = $100f_i$
S	35	0.233	23.3
B	62	0.413	41.3
O	53	0.354	35.4
<b>tot.</b>	<b>150</b>	<b>1</b>	<b>100</b>



# Dati categorici


Voto all'esame: da 18 a 30L

Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo



Rendimento	$n_i$	$f_i = n_i/n$	% = $100f_i$	$P_i$ <b>cumulata</b>
S	35	0.233	<b>23.3</b>	<b>23.3%</b>
B	62	0.413	<b>41.3</b>	<b>64.6%</b>
O	53	0.354	<b>35.4</b>	<b>100.0%</b>
<b>tot.</b>	<b>150</b>	<b>1</b>	<b>100</b>	

# Dati categorici

Voto all'esame: da 18 a 30L

Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo



Rendimento	$n_i$	$f_i = n_i/n$	% = $100f_i$	$N_i$ cumulata
S	35	0.23̄	23.3̄	35
B	62	0.41	41.3̄	97
O	53	0.35	35.3̄	150
tot.	150	1	100	