

Progetto Formativo Aziendale  
ORGANIZZAZIONE E UTILIZZO DI DATASET DI DATI ANAMNESTICI E DESITI  
DIAGNOSTICI: RACCOLTA ED ARCHIVIAZIONE DEI DATI  
Roma, 7-8 giugno 2016

## Pronto uso del dataset: tabelle descrittive di base

Marcello Sala  
Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana  
Osservatorio Epidemiologico Veterinario Regione Lazio  
Roma

# Pronto uso del dataset

La pronta usabilità del dataset dipende dalla qualità delle procedure adottate nelle fasi di

- Progettazione
- Raccolta dati
- Inserimento
- Alimentazione
- Verifica
- Codifica/ricodifica/riclassificazione

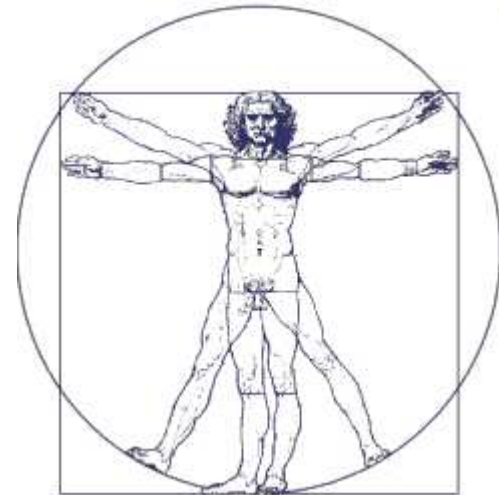
# Pronto uso del dataset

Dipende dal “fisico” del dataset prodotto



Panzaset

Ci vuole un personal  
trainer!

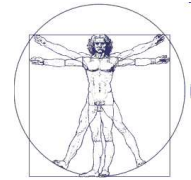


Dataset

Pronto all'uso

# Pronto uso del dataset

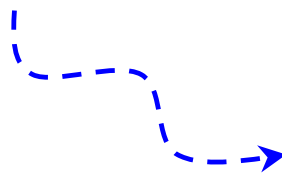
Ammettiamo che il dataset sia in forma



Quali sono gli utilizzi immediati?

Valutazione **descrittiva**  
di base dei dati e dei  
risultati

Valutazione **analitica**  
dei risultati



# Valutazione descrittiva di base

- Consente di produrre la descrizione della base dati relativamente alle caratteristiche della popolazione, del campione o del gruppo sottoposto ad indagine (record)
- Ulteriore strumento di controllo/verifica dei dati
- Fornisce una prima idea dei risultati e della loro distribuzione in base alle diverse caratteristiche (campi) di popolazione, campione o gruppo considerati

# Valutazione descrittiva di base

- Consiste nella “esplorazione” dei dati
  - Tabelle di frequenza dei valori di un campo
  - Tabelle di contingenza a due (o più) entrate

# Tabelle di frequenza

- Viene calcolata la frequenza assoluta e percentuale dei valori contenuti in un campo del dataset
- ..ossia si conteggia quante volte un valore è presente in quel campo

# Tabelle di frequenza

Es.  
tabella  
SESSO

Progressivo campione	matricola	Indirizzo produttivo	sexso	ANNO NASCITA	esito	lt/die	n° parti	TRATTAMENTI	vaccinazione
1	00000001	latte	F	2012	pos	28	1	MAI	vaccino 1
2	00000002	carne	F	2010	neg	24	2	QUALCHE VOLTA	nessuna
3	00000003	latte	F	2006	neg	20	3	MAI	nessuna
4	00000004	carne	F	2016	neg	26	3	QUALCHE VOLTA	nessuna
5	00000005	latte	M	2012	pos	32	1	QUALCHE VOLTA	vaccino 1
6	00000006	carne	F	2010	pos	18	2	SEMPRE	vaccino 1
7	00000007	carne	F	2014	pos	38	1	SEMPRE	nessuna
8	00000008	carne	MISSING	2013	neg	51	1	MAI	nessuna
9	00000009	latte	F	2012	neg	180	2	QUALCHE VOLTA	vaccino 1
10	00000010	misto	F	1992	neg	19	1	SEMPRE	nessuna

Sexso	N° capi	% capi
M	1	10
F	8	80
missing	1	10
Totale	10	

Descrive la  
distribuzione per  
sexso del gruppo

Individua valori inattesi  
(Completezza-omogeneità)

*..è anche strumento di controllo/verifica dei dati*



# Tabelle di frequenza

Progressivo campione	matricola	Indirizzo produttivo	sexso	ANNO NASCITA	esito	lt/die	n° parti	TRATTAMENTI	vaccinazione
1	00000001	latte	F	2012	pos	28	1	MAI	vaccino 1
2	00000002	carne	F	2010	neg	24	2	QUALCHE VOLTA	nessuna
3	00000003	latte	F	2006	neg	20	3	MAI	nessuna
4	00000004	carne	F	2016	neg	26	3	QUALCHE VOLTA	nessuna
5	00000005	latte	M	2012	pos	32	1	QUALCHE VOLTA	vaccino 1
6	00000006	carne	F	2010	pos	18	2	SEMPRE	vaccino 1
7	00000007	carne	F	2014	pos	38	1	SEMPRE	nessuna
8	00000008	carne	MISSING	2013	neg	51	1	MAI	nessuna
9	00000009	latte	F	2012	neg	180	2	QUALCHE VOLTA	vaccino 1
10	00000010	misto	F	1992	neg	19	1	SEMPRE	nessuna

Es. tabella  
trattamenti

Trattamenti	N° capi	% capi
MAI	3	30
QUALCHE VOLTA	4	40
SEMPRE	3	30
Totale	10	

Descrive la distribuzione  
per frequenza trattamento  
nel gruppo

# Tabelle di frequenza

Progressivo campione	matricola	Indirizzo produttivo	sezzo	ANNO NASCITA	esito	lt/die	n° parti	TRATTAMENTI	vaccinazione
1	00000001	A	0	2012	1	28	1	0	1
2	00000002	B	0	2010	0	24	2	1	0
3	00000003	A	0	2006	0	20	3	0	0
4	00000004	B	0	2016	0	26	3	1	0
5	00000005	A	1	2012	1	32	1	1	1
6	00000006	B	0	2010	1	18	2	2	1
7	00000007	B	0	2014	1	38	1	2	0
8	00000008	B	99	2013	0	51	1	0	0
9	00000009	A	0	2012	0	180	2	1	1
10	00000010	C	0	1992	0	19	1	2	0

Es. tabella  
trattamenti  
"codificata"

Trattamenti	N° capi	% capi
0	3	30
1	4	40
2	3	30
Totale	10	

Descrive la distribuzione per  
frequenza trattamento  
"codificata", nel gruppo

# Tabelle di frequenza

Progressivo campione	matricola	Indirizzo produttivo	sezzo	età anni	Clasetà
1	00000001	A	0	3	1
2	00000002	B	0	5	2
3	00000003	A	0	9	3
4	00000004	B	0	1	1
5	00000005	A	1	3	1
6	00000006	B	0	5	2
7	00000007	B	0	1	1
8	00000008	B	99	2	1
9	00000009	A	0	3	1
10	00000010	C	0	17	3

Es.  
tabella età  
&  
Classe età  
riclassificata

Descrive la distribuzione per  
età nel gruppo

Età anni	N° capi	% capi
1	2	20
2	1	10
3	3	30
5	2	20
9	1	10
17	1	10
Totale	10	

Classe_età	N° capi	% capi
1	6	60
2	2	20
3	2	20
Totale	10	

Livello informativo più  
esplicito e sintetico

Descrive la distribuzione per  
classe d'età nel gruppo

# Tabelle di frequenza

- Questi esempi oltre a rappresentare le modalità di esecuzione di semplici valutazioni descrittive dei **dati** introducono...
- Una prima idea delle frequenze delle variabili indipendenti o fattori di rischio o di esposizione che dovranno essere messi in relazione con gli esiti nella fase analitica dei dati p.d.

# Tabelle di frequenza esiti

- Le stesse modalità possono essere applicate anche al campo esito
- Più facile in caso di 2 o 3 livelli di esito
- Improponibile per esiti quantitativi rappresentati da variabili continue (es mg/Kg)

COD ASL	COD AZ	n° unità fisica	FORAGGI A RISCHIO SI/NO	MATRICE	ESITO ppb
O107	041RM020	3	SI	LATTE DI MASSA CRUDO	0.015
O112	006FR329	4	SI	LATTE DI MASSA CRUDO	0.007
O112	006FR310	5	SI	LATTE DI MASSA CRUDO	0.015
O112	073FR004	8	SI	LATTE DI MASSA CRUDO	0.338
O112	045FR066	9	SI	LATTE DI MASSA CRUDO	0.019
O107	041RM031	10	SI	LATTE DI MASSA CRUDO	0.0046
O112	046FR007	11	SI	LATTE DI MASSA CRUDO	0.027
O112	006FR225	12	SI	LATTE DI MASSA CRUDO	0.01
O112	046FR012	14	SI	LATTE DI MASSA CRUDO	0.015
O107	041RM025	15	SI	LATTE DI MASSA CRUDO	0.05
O112	006FR261	17	SI	LATTE DI MASSA CRUDO	0.0285
O107	041RM002	18	SI	LATTE DI MASSA CRUDO	0.02
O112	073FR012	19	SI	LATTE DI MASSA CRUDO	0.022
O112	045FR086	21	SI	LATTE DI MASSA CRUDO	0.012
O112	006FR285	22	SI	LATTE DI MASSA CRUDO	0.006
O112	076FR070	23	SI	LATTE DI MASSA CRUDO	0.017
O112	006FR098	24	SI	LATTE DI MASSA CRUDO	0.027
O112	076FR017	26	SI	LATTE DI MASSA CRUDO	0.007
O112	006FR169	28	SI	LATTE DI MASSA CRUDO	0.018
O107	102RM022	29	SI	LATTE DI MASSA CRUDO	0.015
O112	033FR072	30	SI	LATTE DI MASSA CRUDO	0.0087
O107	041RM028	31	SI	LATTE DI MASSA CRUDO	0.041
O107	102RM021	32	SI	LATTE DI MASSA CRUDO	0.01
O107	041RM017	33	SI	LATTE DI MASSA CRUDO	5.91
O107	102RM061	38	SI	LATTE DI MASSA CRUDO	0.02
O112	006FR302	39	SI	LATTE DI MASSA CRUDO	0.016
O112	002FR002	42	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	002FR006	43	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	002FR009	44	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	002FR012	45	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR001	46	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR003	47	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR006	48	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR007	49	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR014	52	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR019	54	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR023	56	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR027	57	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR032	58	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR046	59	NO	LATTE DI MASSA CRUDO	0.006
O112	006FR047	60	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR048	61	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR052	62	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR054	63	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR055	64	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR061	65	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR062	66	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR070	67	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR073	68	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR076	69	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR077	70	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR084	72	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR085	73	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR091	74	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR097	75	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR099	76	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR103	77	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR110	78	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR117	79	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR127	80	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR141	81	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR142	82	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR147	83	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR149	84	SI	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR151	85	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR157	88	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR161	89	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR162	90	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR164	91	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR170	92	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR175	93	NO	LATTE DI MASSA CRUDO	< 0.0005
O112	006FR176	94	SI	LATTE DI MASSA CRUDO	0.0024

Es. tabella esiti quantitativi 1 dati continui

ESITO ppb	frequenza assoluta	frequenza %
< 0.0005	46	62,16
0.0006	1	1,35
0.0024	1	1,35
0.0046	1	1,35
0.006	1	1,35
0.007	2	2,70
0.0087	1	1,35
0.01	2	2,70
0.012	1	1,35
0.015	4	5,41
0.016	1	1,35
0.017	1	1,35
0.018	1	1,35
0.019	1	1,35
0.02	2	2,70
0.022	1	1,35
0.027	2	2,70
0.0285	1	1,35
0.041	1	1,35
0.05	1	1,35
0.338	1	1,35
5.91	1	1,35
Totale complessivo	74	

La tabella di frequenza degli esiti è grande quasi come il dataset  
 ...perché quasi tutti i valori di esito sono diversi tra loro  
 ...le frequenze assolute tendono ad assumere valore unitario per tutte le modalità

E' inutile fare la tabella di freq!

COD ASL	COD AZ	n° unità fisica	FORAGGI A RISCHIO S/NO	MATRICE	ESITO ppb	esito 2
O107	041RM020	3	SI	LATTE DI MASSA CRUDO	0.015	POSITIVO
O112	006FR329	4	SI	LATTE DI MASSA CRUDO	0.007	POSITIVO
O112	006FR310	5	SI	LATTE DI MASSA CRUDO	0.015	POSITIVO
O112	073FR004	8	SI	LATTE DI MASSA CRUDO	0.338	POSITIVO
O112	045FR066	9	SI	LATTE DI MASSA CRUDO	0.019	POSITIVO
O107	041RM031	10	SI	LATTE DI MASSA CRUDO	0.0046	POSITIVO
O112	046FR007	11	SI	LATTE DI MASSA CRUDO	0.027	POSITIVO
O112	006FR225	12	SI	LATTE DI MASSA CRUDO	0.01	POSITIVO
O112	046FR012	14	SI	LATTE DI MASSA CRUDO	0.015	POSITIVO
O107	041RM025	15	SI	LATTE DI MASSA CRUDO	0.05	POSITIVO
O112	006FR261	17	SI	LATTE DI MASSA CRUDO	0.0285	POSITIVO
O107	041RM002	18	SI	LATTE DI MASSA CRUDO	0.02	POSITIVO
O112	073FR012	19	SI	LATTE DI MASSA CRUDO	0.022	POSITIVO
O112	045FR086	21	SI	LATTE DI MASSA CRUDO	0.012	POSITIVO
O112	006FR285	22	SI	LATTE DI MASSA CRUDO	0.006	POSITIVO
O112	076FR070	23	SI	LATTE DI MASSA CRUDO	0.017	POSITIVO
O112	006FR098	24	SI	LATTE DI MASSA CRUDO	0.027	POSITIVO
O112	076FR017	26	SI	LATTE DI MASSA CRUDO	0.007	POSITIVO
O112	006FR169	28	SI	LATTE DI MASSA CRUDO	0.018	POSITIVO
O107	102RM022	29	SI	LATTE DI MASSA CRUDO	0.015	POSITIVO
O112	033FR072	30	SI	LATTE DI MASSA CRUDO	0.0087	POSITIVO
O107	041RM028	31	SI	LATTE DI MASSA CRUDO	0.041	POSITIVO
O107	102RM021	32	SI	LATTE DI MASSA CRUDO	0.01	POSITIVO
O107	041RM017	33	SI	LATTE DI MASSA CRUDO	5.91	POSITIVO
O107	102RM061	38	SI	LATTE DI MASSA CRUDO	0.02	POSITIVO
O112	006FR302	39	SI	LATTE DI MASSA CRUDO	0.016	POSITIVO
O112	002FR002	42	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	002FR006	43	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	002FR009	44	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	002FR012	45	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR001	46	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR003	47	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR006	48	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR007	49	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR014	52	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR019	54	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR023	56	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR027	57	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR032	58	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR046	59	NO	LATTE DI MASSA CRUDO	0.0006	POSITIVO
O112	006FR047	60	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR048	61	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR052	62	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR054	63	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR055	64	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR061	65	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR062	66	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR070	67	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR073	68	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR076	69	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR077	70	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR084	72	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR085	73	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR091	74	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR097	75	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR099	76	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR103	77	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR110	78	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR117	79	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR127	80	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR141	81	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR142	82	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR147	83	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR149	84	SI	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR151	85	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR157	88	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR161	89	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR162	90	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR164	91	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR170	92	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR175	93	NO	LATTE DI MASSA CRUDO	< 0.0005	NEGATIVO
O112	006FR176	94	SI	LATTE DI MASSA CRUDO	0.0024	POSITIVO

Es. tabella esiti quantitativi 2 dati da continui a dicotomici

Ricodifica degli esiti in positivo/negativo (Dicotomica)

esito 2	frequenza assoluta	frequenza %
NEGATIVO	46	62,16
POSITIVO	28	37,84
Totale complessivo	74	

La tabella di frequenza degli esiti è molto piccola

TROPPO!

Non rappresenta la variabilità dell'esito analitico

Valori alti o bassi tra i positivi sono considerati allo stesso modo

Perdita di informazione (e descrizione)

## Es. tabella esiti quantitativi 3 continui a qualit. ordinale

classe di concentrazione (parti per bilione) di $\beta$ -HCH nel latte massale	Esito analitico qualitativo	Esito in base a conformità a limiti di legge
< 0,5	negativo	conforme
0,51-2,99	positivo	conforme
$\geq 3$	positivo	non-conforme

Riclassificazione degli esiti in negativo, positivo conforme, positivo non conforme (qualitativo ordinale)

Esito in base a conformità a limiti di legge	frequenza assoluta	frequenza %
neg-conforme	198	81,15
pos-conforme	16	6,56
pos-non-conforme	30	12,30
Totale complessivo	244	

La tabella di frequenza degli esiti è sintetica, informativa, rappresenta molto meglio la variabilità dell'esito analitico e discrimina il gradiente o la magnitudine dei positivi

Viene introdotto un NUOVO elemento informativo, ossia il riferimento ai **valori consentiti** dalla norma



# Tabelle di frequenza

## esiti

- Questi esempi oltre a rappresentare le modalità di esecuzione di semplici valutazioni descrittive degli **esiti** introducono...
- Una prima idea delle frequenze degli esiti e delle possibilità di loro aggregazione ai fini delle analisi

# Tabelle di frequenza

## Frequenza cumulativa

Sommando progressivamente le frequenze relative percentuali contigue corrispondenti ad ogni valore della variabile considerata otteniamo la frequenza cumulativa percentuale (o frequenza cumulativa)

# Tabelle di frequenza

Frequenza cumulativa

## Es. distribuzione aziende ovine positive per paratubercolosi Lazio e Toscana

classe prev %	Frequenze assolute	Frequenze relative	Frequenze relative percentuali	Frequenza cumulativa
a: 0	254	0,46	45,8	45,8
b: 0,1-5	85	0,15	15,3	61,2
c: 5,1-10	103	0,19	18,6	79,8
d: 10,1-20	76	0,14	13,7	93,5
e: 20,1-30	20	0,04	3,6	97,1
f: >30	16	0,03	2,9	100,0
Totale complessivo	554	1,00	100,0	

$$\begin{aligned} &= 45,8 + 15,3 \\ &= 61,2 + 18,6 \end{aligned}$$

È utile per avere una idea immediata della soglia di prevalenza che comprende la maggior parte delle aziende

In questo caso si può dire che circa 80% delle aziende ha una prevalenza di paratubercolosi  $\leq 10\%$

# Tabelle di frequenza

Frequenza cumulativa

Osservando la tabella della distribuzione di frequenza è possibile avere un colpo d'occhio immediato sulla situazione generale della popolazione (azienda) rispetto ad una determinata variabile e....

Valutare macroscopicamente aspetti positivi e negativi legati alla variabile

# Tabelle di frequenza

Frequenza cumulativa

classe prev %	Frequenze assolute	Frequenze relative	Frequenze relative percentuali	Frequenza cumulativa
a: 0	254	0,46	45,8	45,8
b: 0,1-5	85	0,15	15,3	61,2
c: 5,1-10	103	0,19	18,6	79,8
d: 10,1-20	76	0,14	13,7	93,5
e: 20,1-30	20	0,04	3,6	97,1
f: >30	16	0,03	2,9	100,0
Totale complessivo	554	1,00	100,0	

**Coda a  
destra o  
positiva**

Le frequenze si concentrano nei valori BASSI della variabile

Nell'esempio paratubercolosi è un aspetto positivo  
(prevalenza malattia bassa)

Se fosse un parametro produttivo (Latte/capo/die) sarebbe  
un aspetto negativo

# Tabelle di frequenza

Frequenza cumulativa

Giorni interparto	Frequenze assolute	Frequenze relative	Frequenze relative percentuali	Frequenza cumulativa
365-380	8	0,04	4,5	4,5
381-395	15	0,08	8,4	12,9
396-410	20	0,11	11,2	24,2
411-425	28	0,16	15,7	39,9
426-440	35	0,20	19,7	59,6
440-480	40	0,22	22,5	82,0
480+	32	0,18	18,0	100,0
Somma $\Sigma$	178	1	100	

**Coda a  
sinistra o  
negativa**

**75,8%**

Le frequenze si concentrano nei valori ALTI della variabile

Nell'esempio "intervallo interparto" è un aspetto negativo  
(minor produzione latte/capo/lattazione)

Se fosse un indice di fertilità sarebbe un aspetto positivo

## **Come rappresentare le distribuzioni di frequenza**

La distribuzione di frequenza può essere rappresentata graficamente.

E' utile avere un'idea di quale grafico usare in base al tipo di variabile considerata (qualitativa-quantitativa).

Istogramma

Torta

# Come rappresentare le distribuzioni di frequenza

## Istogramma

Utilizzabile per variabili quantitative e qualitative ordinali

Asse orizzontale (ascisse)

valori crescenti della variabile o classi crescenti

Asse verticale (ordinate)

Frequenze assolute o relative

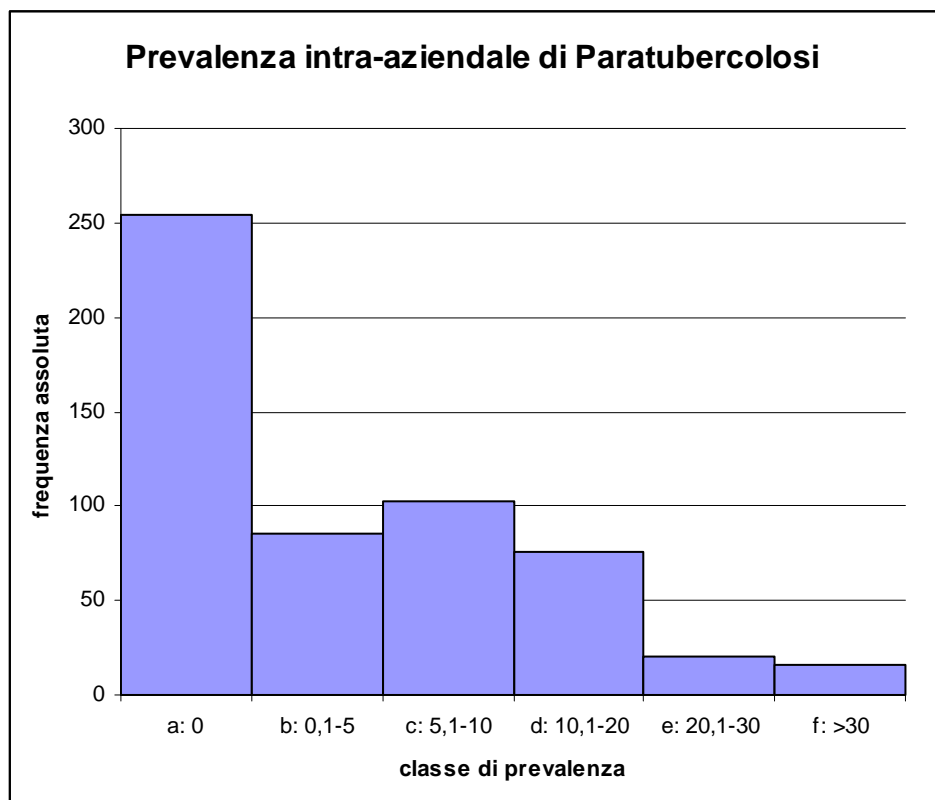
Sopra ciascuna classe riportata nelle ascisse si riporta una barra rettangolare la cui altezza corrisponde alla rispettiva frequenza nelle ordinate.

Le barre devono essere unite e nel caso delle classi bisogna evitare "salti" tra le classi.

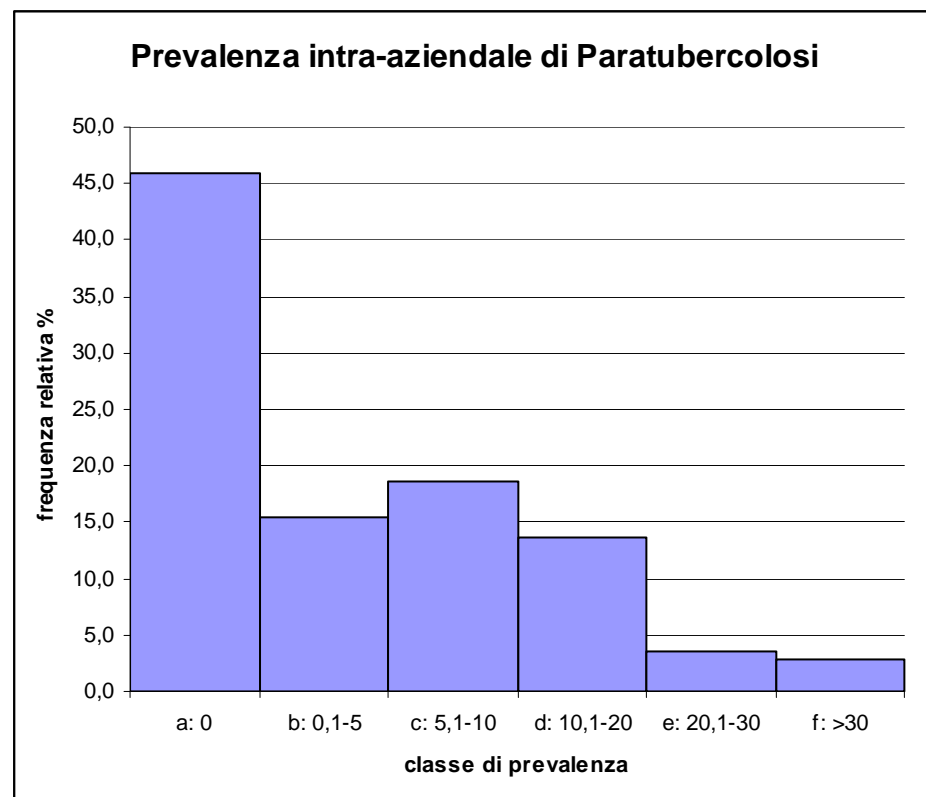


# Come rappresentare le distribuzioni di frequenza

## Istogramma



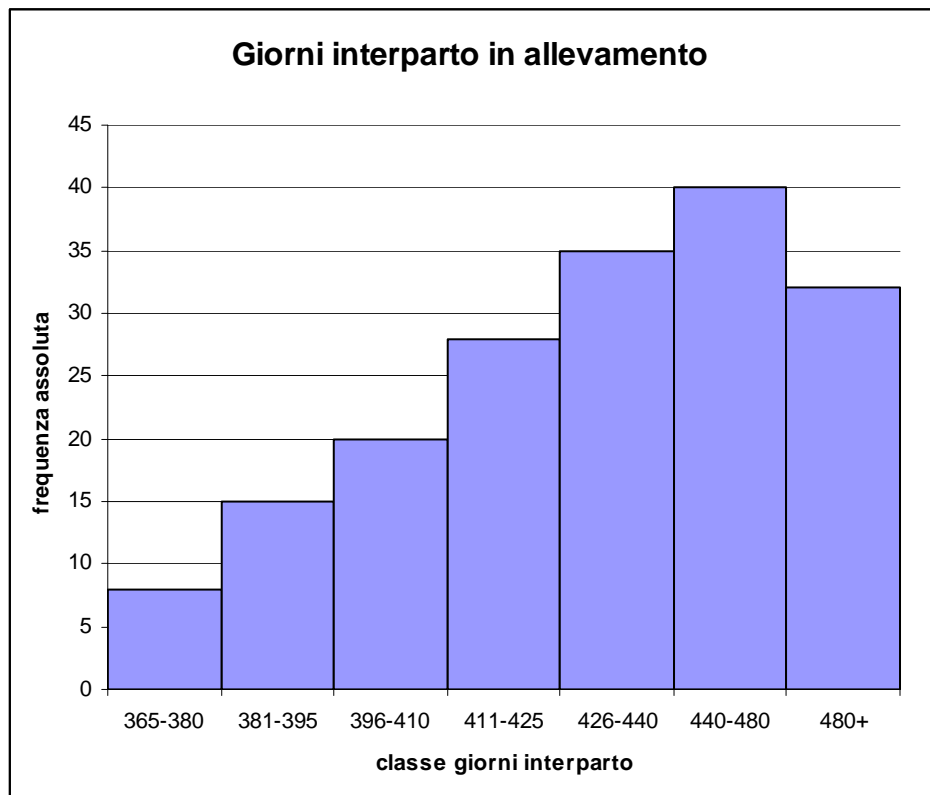
Frequenza assoluta



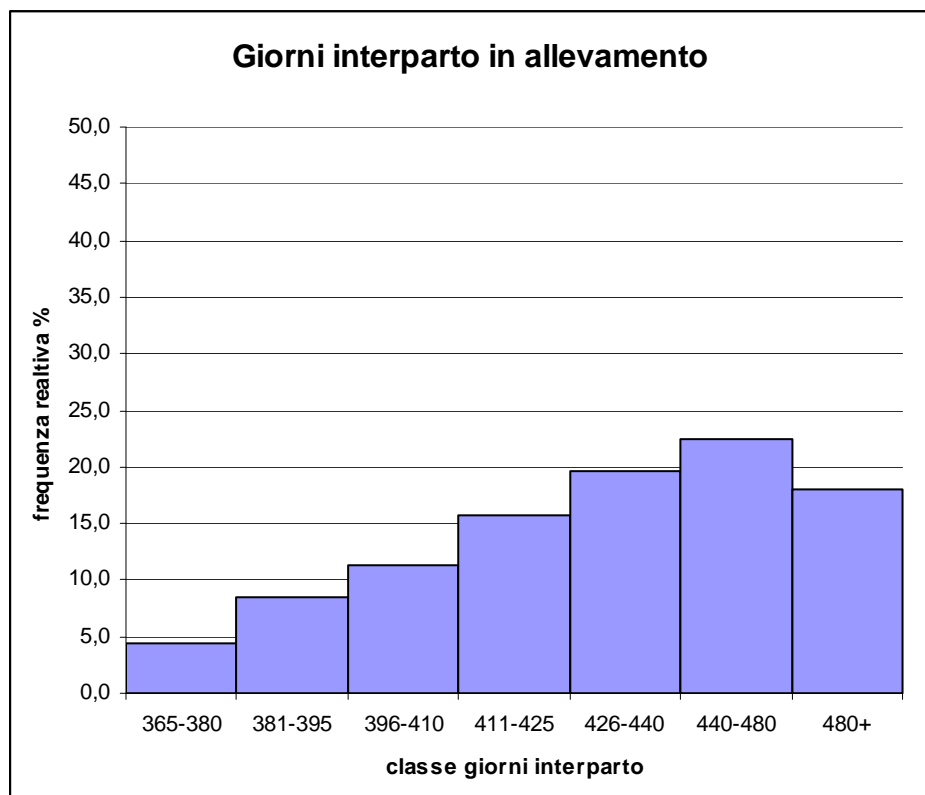
Frequenza relativa

# Come rappresentare le distribuzioni di frequenza

## Istogramma



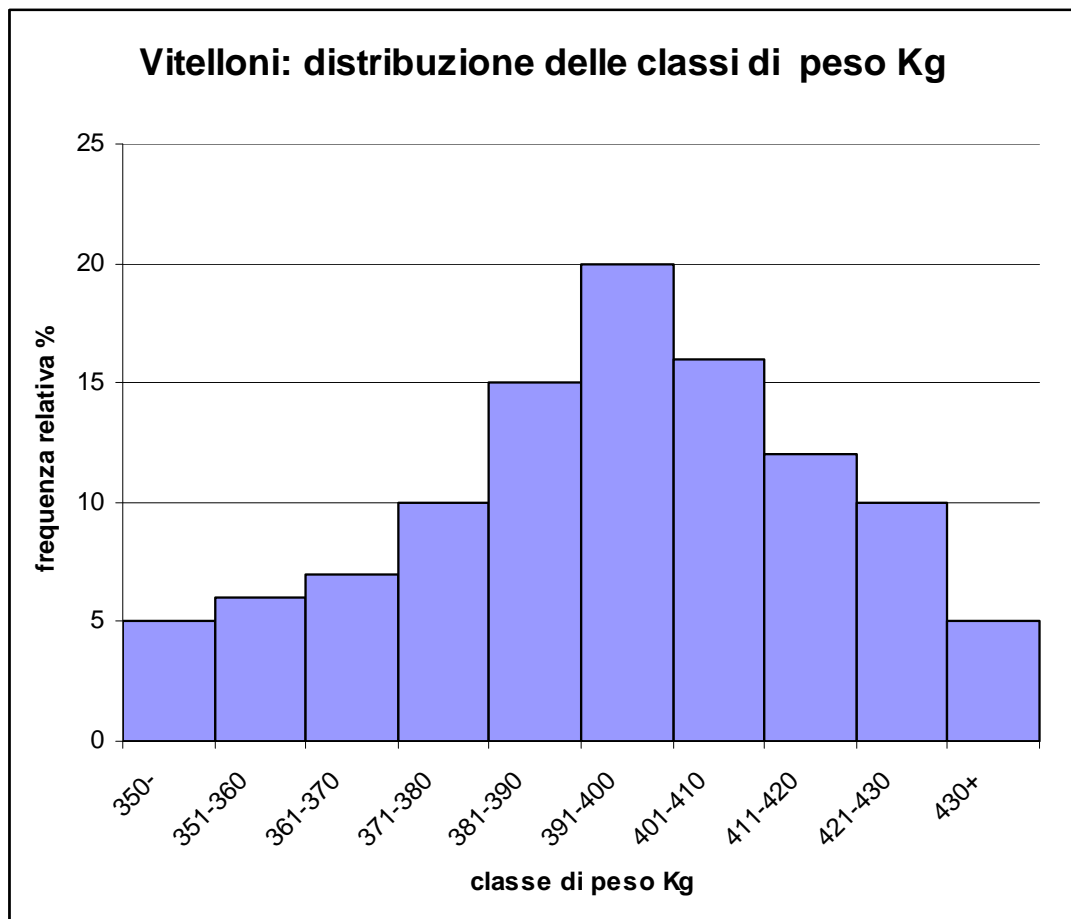
Frequenza assoluta



Frequenza relativa

# Come rappresentare le distribuzioni di frequenza

## Istogramma



**Distribuzione  
"normale"  
simmetrica  
parametri  
morfometrici**

## **Come rappresentare le distribuzioni di frequenza Torta**

Particolarmente utile per rappresentare la distribuzione di frequenza di variabili qualitative (Nominali o Ordinali)

Solo per frequenza relative %

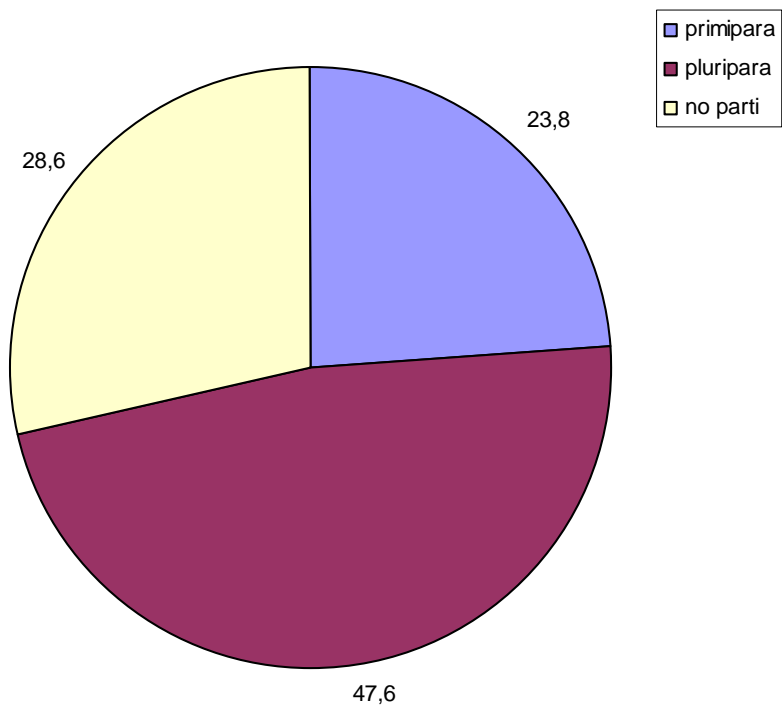
Cerchio diviso in spicchi nel quale ogni spicchio rappresenta la frequenza relativa della variabile per ogni classe.

Facilita i confronti tra 2 situazioni diverse

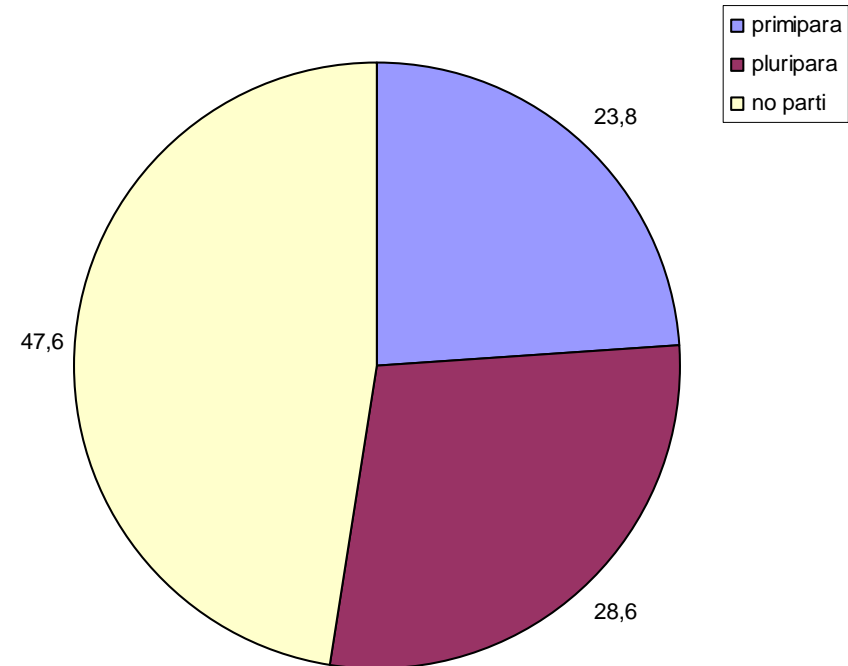
# Come rappresentare le distribuzioni di frequenza

## Torta

Azienda 1: Parità-Frequenze relative percentuali



Azienda 2: Parità-Frequenze relative percentuali



# Tabelle di contingenza a due (o più) entrate

- Vengono incrociate in una tabella 2x2 (o più) le frequenze relative dei valori di campi diversi del dataset
- Definiscono come si distribuisce una variabile in funzione dell'altra
- ..in particolare può descrivere come si distribuisce l'esito in funzione di un'altra variabile (esposizione)

# Tabelle di contingenza a due (o più) entrate

Se per ogni individuo (unità statistica) viene osservata più di una variabile è possibile procedere alla descrizione delle variabili congiuntamente considerate;

in particolare, permette di ipotizzare se le variabili si influenzano o se, al contrario, si manifestano una indipendentemente dall'altra;

Nel caso di 2 variabili, si tratta di raccogliere i dati in una tabella a doppia entrata (o **tabella di contingenza**) in grado di mostrare congiuntamente i valori delle 2 variabili.

# Tabelle di contingenza a due (o più) entrate

	indice incremento ponderale			
Sesso	Alto	Basso	Medio	Somma $\Sigma$ Sesso
Femmina	4	4	3	11
Maschio	3	4	2	9
Somma $\Sigma$ incremento	7	8	5	20

**distribuzione congiunta di Sesso e di Indice incremento:**

le frequenze congiunte (assolute), che si trovano al centro della tabella, stanno ad indicare quante unità statistiche hanno manifestato contemporaneamente un valore "sesso" e uno dei 3 valori "Indice incremento". Es 4 femmine con incremento Alto ecc..



# Tabelle di contingenza a due (o più) entrate

	indice incremento ponderale			
Sesso	Alto	Basso	Medio	Somma $\Sigma$ Sesso
Femmina	4	4	3	11
Maschio	3	4	2	9
Somma $\Sigma$ incremento	7	8	5	20

Frequenze marginali di **Sesso** e di **Indice incremento**: si  
"ritorna alla distribuzione di frequenza separata di  
ogni variabile eliminando l'effetto dell'altra.

# Tabelle di contingenza a due (o più) entrate

	paratubercolosi		
Sesso	negativo	positivo	Somma $\Sigma$ Sesso
Femmina	8	3	11
Maschio	5	4	9
Somma $\Sigma$ paraTBC	13	7	20

In altre parole: “confronto veloce” della frequenza di una variabile in funzione dell'altra

In questo esempio, confronto positivi paraTBC tra maschi e femmine

**Concetto importante quando si fanno confronti mediante analisi statistica**

# Tabelle di contingenza a due (o più) entrate

Normalmente le tabelle di contingenza incrociano 1 variabile di esposizione o fattore di rischio (ed es. sesso; trattamento si/no) con la variabile di esito (outcome)

Questo consente di verificare come si distribuiscono i positivi ed i negativi tra i valori della variabile di esposizione considerata

..in altre parole ci consentono di fare una prima ipotesi sulla eventuale associazione tra fattore di rischio e positività

Le analisi statistiche ci diranno poi se è così o no

# Un esempio ipotesi

introduzione/circolazione stagionale del virus in assenza di sintomatologia clinica nei cavalli successivamente al focolaio 1998

## obiettivi dello studio

- 1) valutare la presenza di cavalli sieropositivi e stimarne la prevalenza per coorte di nascita per verificare la circolazione virale successiva al 1998
- 2) monitorare nel 2006 eventuali sieroconversioni nel corso della stagione a rischio

# Un esempio metodologia

## area di studio:

buffer 9 Km dal centro geografico del P. del Fucecchio  
censimento della popolazione di equidi: 2.200 capi

## campionamento:

**Studio Trasversale:** casuale semplice di 350 capi residenti  
stabilmente (p.a. 20%; E.S. 5%; L.C. 99%)

-singolo controllo sierologico a maggio 2006 (ELISA competitiva)

# Un esempio risultati

- coorte di nascita = proxy dell'esposizione (presenza nell'area al momento del focolaio 1998)

coorte di nascita (esposizione)	positivo	negativo	Totale
1980-1998 (exp+)	14	88	102
1999-2006 (exp-)	3	230	233
Totale	17	318	335

- O.R. 12,2 (I.C. 95%: 3,4 - 43,5)
- i cavalli sieropositivi hanno mostrato una probabilità circa 12 volte maggiore di appartenere alle coorti di nascita precedenti o contemporanea al focolaio 1998

## conclusioni

- 1) la sieropositività dei cavalli è associata alla presenza nell'area al momento del focolaio epidemico del 1998 - massima esposizione;
- 2) il virus West Nile potrebbe aver circolato anche successivamente senza dar luogo a epidemie, in assenza di forme cliniche nei cavalli e di sieroconversioni nei polli sentinella;
- 3) possibile circolazione virale tra il 1999 ed il 2005, sporadica e riferibile ad introduzioni stagionali ed occasionali del virus;
- 4) sussiste il rischio reintroduzione del virus e di innesco del ciclo locale di trasmissione

# Altro esempio

## Es. tabella esiti quantitativi continui 2

classe di concentrazione (parti per bilione) di $\beta$ -HCH nel latte massale	Esito analitico qualitativo	Esito in base a conformità a limiti di legge
< 0,5	negativo	conforme
0,51-2,99	positivo	conforme
$\geq 3$	positivo	non-conforme

1. Riclassificazione degli esiti in negativo, positivo conforme, positivo non conforme (qualitativo ordinale)

Esito in base a conformità a limiti di legge	frequenza assoluta	frequenza %
neg-conforme	198	81,15
pos-conforme	16	6,56
pos-non-conforme	30	12,30
Totale complessivo	244	

2. Riclassificazione degli esiti in casi, non casi

classe di concentrazione (parti per bilione) di $\beta$ -HCH nel latte massale	Esito analitico qualitativo	Esito in base a conformità a limiti di legge	caso non caso
< 0,5	negativo	conforme	non-casi
0,51-2,99	positivo	conforme	non-casi
$\geq 3$	positivo	non-conforme	casi

	Numero di aziende bovine			Tasso d'attacco
	CASI	NON-CASI	Totale	
esposte	33	23	56	59%
NON esposte	1	187	188	0,50%
Totale	34	210	244	



	Numero di aziende bovine			Tasso d'attacco
	CASI	NON-CASI	Totale	
esposte	33	23	56	59%
NON esposte	1	187	188	0,50%
Totale	34	210	244	

- 1) Da una tabella di contingenza è facilmente calcolabile il tasso d'attacco
- 2) Il tasso d'attacco è la % di Esiti positivi (casi; successi; ecc..) rilevata per ogni categoria (classe) dell'altra variabile considerata (esposizione)
- 3) Osservando la tabella di contingenza e calcolando il tasso d'attacco possiamo avere una prima evidenza della possibile esistenza di una associazione tra evento e esposizione
- 4) La statistica ci consentirà verificare se la differenza osservata è dovuta al caso o no (test di significatività)

## ..concludendo...

Senza arrivare all'analisi statistica o meglio..prima di arrivare all'analisi statistica, le tabelle descrittive ci hanno detto molto su

1. Le frequenze delle variabili e degli esiti nel gruppo di record
2. Le modalità di aggregazione delle variabili più rappresentative ed informative
3. Il rapporto tra variabili di esposizione ed esiti

..preparando il terreno all'analisi statistica...