

# **Il campionamento: quando, come, perché**

M. Mari



Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana  
Sezione di Siena

“Shelf life e campionamento ufficiale di alimenti per analisi microbiologiche”

Pisa 13 dicembre 2012

# Di cosa parleremo

- Il dato: come arrivare a trarre delle conclusioni dalla raccolta?
- Il campione: casuale o a caso?
- Cerchiamo di vedere come si fa in pratica un piano di campionamento



# **Il dato sperimentale: i due postulati**

- Un buon studio necessita di buoni dati
- Un buon dato necessita di un buon piano di campionamento



# **Cos'è il campionamento?**

Procedura attraverso la quale alcuni membri della popolazione sono selezionati come rappresentativi dell'intera popolazione



# Il campione: una piccola popolazione

- Deve ricalcare le caratteristiche dell'intera popolazione
- Avrà un errore di misura rispetto alla popolazione, ma esso è stimabile mediante metodi statistici
- E' sicuramente la scelta da fare per avere dati "a basso prezzo"



# come scegliere un metodo di campionamento

- Popolazione che deve essere studiata
  - dimensione/distribuzione geografica
  - eterogeneità della variabile
- Livello di precisione richiesta
- Perché sto campionando
  - Studio di popolazione (Survey o indagine)
  - Ricerca di Malattie infettive
- Risorse disponibili
- Importanza di avere una stima precisa dell'errore campionario



# Due modalità di campionamento

- Se devo escludere una prevalenza di malattia all'interno della popolazione
  - Uso la formula di Cannon e Roe
- Se faccio il campione per un'indagine userò questa formula

$$n = \frac{z^2 * p * (1-p)}{d^2}$$

**z:** livello di fiducia in z (95%  $\approx$  1.96)

**d:** precisione assoluta

**p:** prevalenza attesa



# Il campionamento in pillole (1)

- Prevalenza attesa, precisione, livello di confidenza sono i tre parametri per calcolare la numerosità campionaria
- Attenzione!!!!
  - Talvolta aumentando la prevalenza non si diminuisce la numerosità!
  - Il problema è dato da  $p * (1-p)$  che è al numeratore della formula
  - Il valore massimo sarà quando la prevalenza attesa è del 50% ( $p=0,5$ ) ed è con questa prevalenza che ho più numerosità campionaria



# Il campionamento in pillole (2)

- Per questo motivo quando non conosco la prevalenza metterò 50% !!
- Sarò così sicuro di avere un campione grande che è più sicuro



# Perché un campione grande è un campione migliore?

- Un campione più numeroso ha un minore errore rispetto al valore vero della popolazione a lui associata
- L'errore standard del campione è stimato infatti da:

- Dove:
$$es = \frac{s}{\sqrt{n}}$$

s = deviazione standard del campione  
n = numerosità del campione



# Il campione: una piccola popolazione

- *Deve ricalcare le caratteristiche dell'intera popolazione*
- Deve essere quindi "stratificato" su determinate caratteristiche che qualificano la popolazione:
  - Sesso
  - Età
  - Distribuzione territoriale di stabilimenti



# Distorsione o Bias

**Le stime che si ottengono da un campione possono essere affette anche da errori causati dalla struttura stessa del campione, diversa da quella della popolazione.**

**DISTORSIONI possono essere dovute a:**

- 1) sfortunati**
- 2) piano di campionamento fatto male**
- 3) molto rifiuti e quindi sostituzioni**



# **Un esempio: il piano di campionamento negli Stabilimenti CEE della Regione**



**“Shelf life e campionamento ufficiale di alimenti per analisi microbiologiche”**

**Pisa 13 dicembre 2012**

# Cosa devo sapere per campionare

- Elenco della popolazione da campionare ( $N$  = numero totale di stabilimenti CEE in Toscana). Non serve per il calcolo del campione ma per verificare che la popolazione non sia troppo piccola ( $n/N$  non dovrebbe essere  $> 5\%$ , altrimenti devo correggere la formula del campionamento)
- Distribuzione del numero per ASL, in modo da poter ponderare il numero nei vari strati.
- Un'idea della prevalenza di quello che vado a cercare
- Quante risorse ho a disposizione (influenzeranno la precisione del campione)
- Il livello di confidenza del campione detto anche errore di tipo alfa (in genere si prende il  $95\% = 0,95 = 1,96$  in unità  $z$ )



# Ottengo i dati

ASL	N. TOTALE STABILIMENTI
L101	45
L102	69
L103	124
L104	26
L105	81
L106	107
L107	128
L108	104
L109	100
L110	155
L111	81
L112	49
TOTALE	1069



“Shelf life e campionamento ufficiale di alimenti per analisi microbiologiche”

Pisa 13 dicembre 2012

# Calcolo la numerosità campionaria

$$n = \frac{z^2 * p * (1-p)}{d^2}$$

$$z: = 1.96$$

$$p: = 2\% (0,02)$$

$$d: = 5\% (0,05)$$

$$n = \frac{1,96^2 \cdot 0,02 \cdot (1 - 0,02)}{0,05^2} = 30$$

**Avrò bisogno di un campione di 30 unità da estrarre dalla popolazione di 1069 unità**



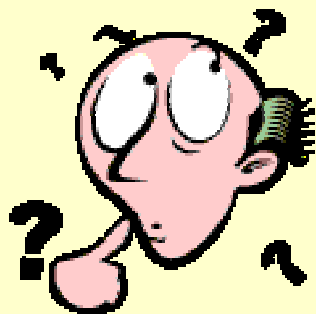
# Pondero il campione sul numero di stabilimenti per ASL

ASL	N. TOTALE STABILIMENTI	%	N
L101	45	4,21%	1
L102	69	6,45%	2
L103	124	11,60%	3
L104	26	2,43%	1
L105	81	7,58%	2
L106	107	10,01%	3
L107	128	11,97%	4
L108	104	9,73%	3
L109	100	9,35%	3
L110	155	14,50%	5
L111	81	7,58%	2
L112	49	4,58%	1
TOTALE	1069		30

# Posso valutare diversi scenari....

Prevalenza	Precisione	Confidenza	N
2%	2%	95%	161
2%	2%	99%	250
10%	2%	95%	478
10%	5%	95%	123
50%	2%	95%	740

**... E scegliere quello che più mi consentono le risorse, giocando sulla precisione ed il livello di confidenza.**



# **GRAZIE DELL'ATTENZIONE!!**

- Per chi volesse contattarmi....
- IZS Lazio e Toscana Sezione di Siena
- Tel 0577-41352
- Email [massimo.mari@izslt.it](mailto:massimo.mari@izslt.it)

# **Adesso lo fate voi!!!!**



**“Shelf life e campionamento ufficiale di alimenti per analisi microbiologiche”**

**Pisa 13 dicembre 2012**

# **Esempio 1**

**Un dipartimento di Prevenzione vuole stimare la prevalenza di casi di tubercolosi tra i bovini nell'area servita. Quanti bovini dovrebbero essere inclusi nel campione per ottenere una stima della prevalenza che varia di 5 punti percentuali attorno al vero valore, che si ipotizza non superi il 20%, con una confidenza del 95%?**

# ***Soluzione 1***

**20% proporzione attesa nella popolazione**

**95% livello di confidenza**

**5% precisione assoluta (15%-25%)**

**Applicando la formula, la dimensione campionaria richiesta è di 246 bovini.**

**Se è impossibile, in termini di tempo e soldi, studiare 246 bovini, allora il ricercatore dovrebbe abbassare il livello di confidenza, ad esempio al 90% (otterrebbe un campione di  $n = 174$ ).**

## **Esempio 2**

**Un ricercatore che lavora in un programma nazionale di immunizzazione cerca di stimare la proporzione di cani che hanno effettuato le vaccinazioni contro la Leptospirosi. Quanti cani dovrebbero essere inclusi nel campione per ottenere una stima della prevalenza che varia di 10 punti percentuali attorno al vero valore, con una confidenza del 95%**

**ATTENZIONE**

**(qui non è possibile fare nessuna assunzione sulla copertura vaccinale)?**

## ***Soluzione 2***

**50% proporzione attesa nella popolazione (non conosco la prevalenza!!)**

**95% livello di confidenza**

**5% precisione assoluta (15%-25%)**

**Applicando la formula, con  $P=0,50$  e  $d=0,10$  la dimensione campionaria richiesta è di 97 cani.**

## **Esempio 3**

**Quanto grande dovrebbe essere il campione per stimare la proporzione di pecore primipare che sono esposte al rischio di aborto da Chlamydia, con una precisione del 5% del vero valore e una confidenza del 95%? Si stima che la proporzione di pecore primipare che abortiscono per clamidiosi vari tra il 25 e il 40%.**

## ***Soluzione 3***

**25-40% proporzione attesa nella popolazione**

**95% livello di confidenza**

**5% precisione assoluta (di 25%-40%)**

P	Numerosità campionaria
0,25	4610
0,3	3585
0,35	2854
0,4	2305

**Pertanto, per ottenere gli obiettivi stabiliti potrebbe essere pianificato uno studio con circa 4610 pecore. Se necessario, potrebbe essere ridotta la numerosità, ma questo comporterebbe una perdita di precisione, di confidenza o di entrambi se il vero valore fosse vicino al 25%.**