



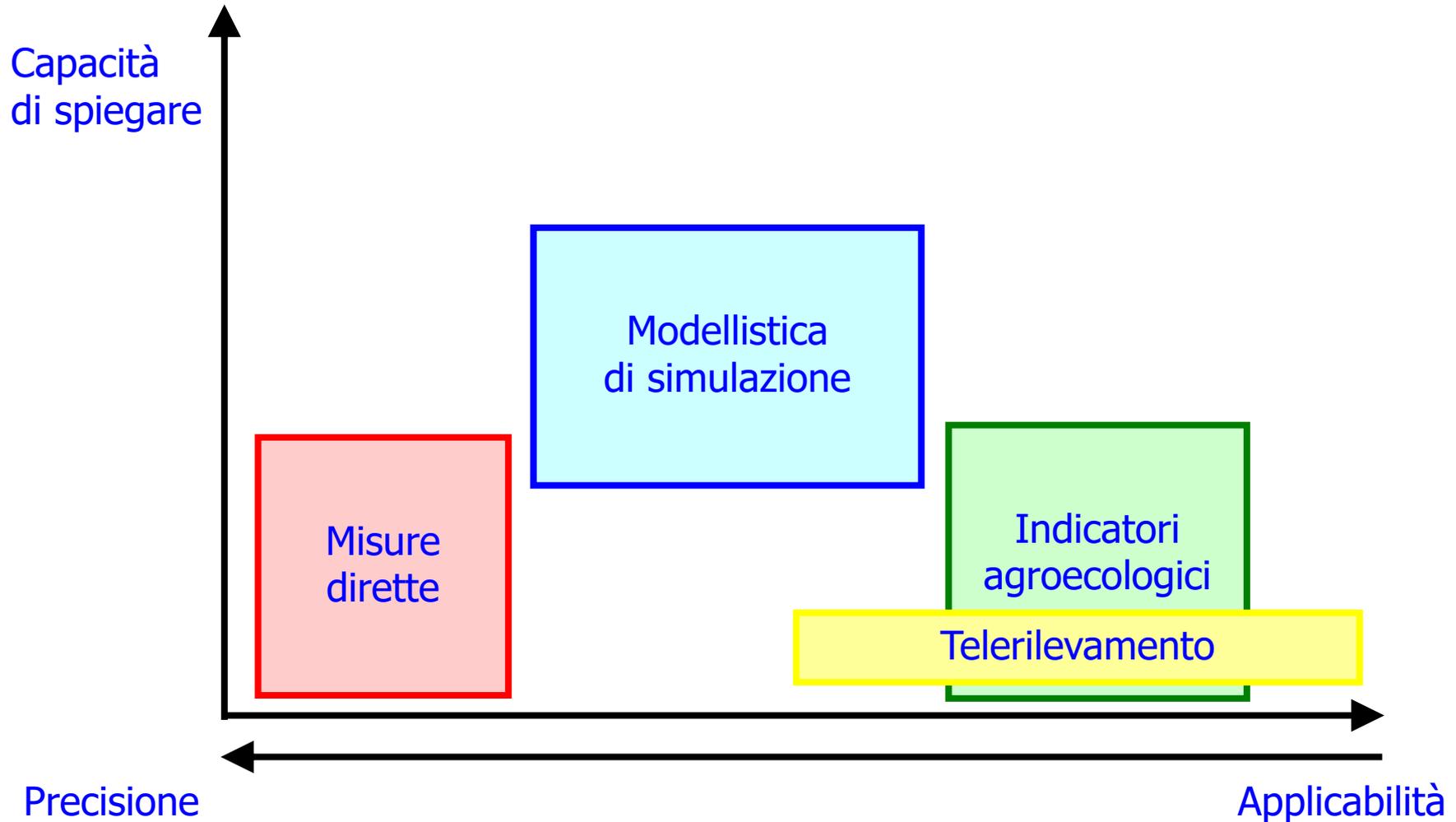
# **Analisi e gestione di sistemi culturali attraverso misure dirette (3)**

## **Problematiche relative alla raccolta dei dati**



# Analisi di sistemi colturali attraverso Misure dirette

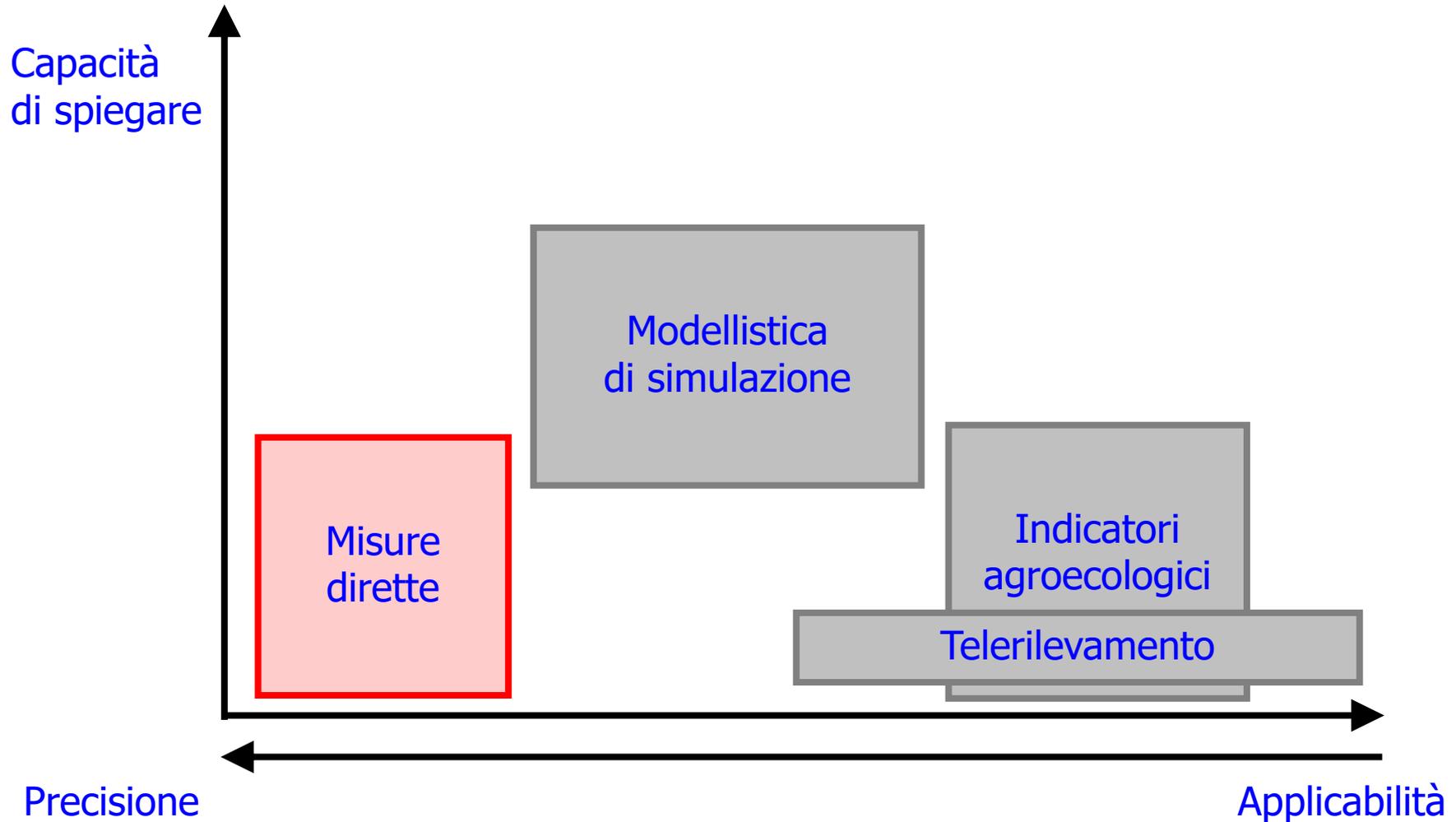
Sistemi Colturali





# Analisi di sistemi colturali attraverso Misure dirette

Sistemi Colturali





# Dimensione del campione (sample size)

Sistemi Culturali

- Avere un numero di *entità* (massa campione) *sufficienti*
  - abbastanza grande da permettere, analizzando un campione, di stimare le caratteristiche dell'intera popolazione
  - ... ma non di più

Sembra un problema di poco conto... ma non lo è!

Ignorarlo può significare sprecare risorse cruciali



# ... dimensionamento dell'unità sperimentale

Sistemi Culturali





# ... dimensionamento dell'unità sperimentale

## Sistemi Colturali

Coltura	Variabile misurata	Dimensioni del campione	CV (%)	Riferimento
orzo	AGB	8 piante per 3 ripetizioni	/	Yap et al., 1972
orzo	AGB	40 cm lineari per 4 ripetizioni	/	Brunetti et al., 1982
frumento	AGB	1 m <sup>2</sup> per 3 ripetizioni	/	Wittmer et al., 1982
mais	AGB	10-60 piante per 3 ripetizioni	/	Williams et al., 1965
soia	AGB	5-10 piante per 5 ripetizioni	/	Buttery et al., 1974
soia	AGB	4 piante per 3 ripetizioni	/	Gent, 1982
<i>Vigna unguiculata</i>	AGB	1 m lineare per 4 ripetizioni	/	Wien, 1982
girasole	AGB	2 piante per 5 ripetizioni	/	Abbate et al., 1982
orzo, pisello	AGB	40 cm lineari per 3 ripetizioni	22	Cervato e Piva, 1985
frumento, orzo	AGB, Nuptake	50 cm lineari per 3 ripetizioni	/	Delogu et al., 1998
frumento	AGB, LAI, SLA	0.25 m <sup>2</sup>	/	Olesen et al., 2002
frumento	AGB, Nuptake	0.17 m <sup>2</sup>	21	Rodriguez et al., 2000
riso	AGB, LAI, SLA	0.5 m <sup>2</sup>	/	Casanova et al., 1998
riso	AGB, LAI, Nuptake	1 m lineare (semina a file)	18	Dingkuhn et al., 1999
cereali		30 piante camminando a zig zag	/	Scottish Agricultural College
tutte		"buon senso" !?!?!?	?!?!?!?	Campbell et al., 1998



# ... dimensionamento dell'unità sperimentale

Sistemi Colturali

Prove nazionali varietali (Istituto Sperimentale per la Cerealicoltura)...

*...è uno standard?*

- Per quanto riguarda la resa granellare:
  - ✓ tutta la parcella esclusi i bordi ( $1.2 \times 5$  m circa)
- Per quanto riguarda biomassa aerea, ecc:
  - ✓ 0.5 m lineari per due ripetizioni (CV = 25%)



# ... dimensionamento dell'unità sperimentale

## Sistemi Colturali

Gomez, 1972 "Techniques for field experiment with rice: layout, sampling, sources of error"...

*...è uno standard?*

In base a numerosi set di dati sperimentali trova delle dimensioni del campione (numero di piante) **diverse a seconda della variabile** che si vuole misurare.

Questo – dice – perché i metodi per misurare le diverse variabili hanno diversa accuratezza...

**Ad ogni modo, a seconda della variabile che misuriamo è necessaria una diversa dimensione del campione!!!**

Per la biomassa aerea suggerisce "20 piante per parcella".



# ... dimensionamento dell'unità sperimentale

Sistemi Colturali

Esempi da letteratura riguardanti metodi alternativi per la determinazione delle dimensioni del campione (1)

Wolkowski et al., 1988

Determinazione della biomassa aerea:

- raccoglie tutte le piante su una fila e ne pesa 5, 10, 15, 20, tutte (estrazioni random)
- calcola il coefficiente di variazione (CV)
- guarda quando il CV è più basso (CV più basso nel suo caso = 14%)

**...precampionamento?**



# ... dimensionamento dell'unità sperimentale

## Sistemi Colturali

Esempi da letteratura riguardanti metodi alternativi per la determinazione delle dimensioni del campione (2)

Yonezawa (1985)

Numero di piante da raccogliere per caratterizzazione genotipica:

- considera tutte le attività necessarie per giungere al dato finale e lo **sforzo** richiesto in ogni attività ed elabora matematicamente il tutto.
- conclude che:
  - ✓ campioni di sole 10 piante per ogni località o popolazione sono sufficienti
  - ✓ è più importante analizzare un rilevante numero di località o popolazioni

Queste conclusioni potrebbero essere estese a prove parcellari considerando le repliche in campo come località: meglio campione "piccolo" ma tante repliche?



# Tecniche classiche per dimensioni del campione

Sistemi Culturali

Tradizionalmente:

$$n = \left( t_{n-1, \alpha} \cdot \frac{s}{|\Delta|} \right)^2$$

Con:

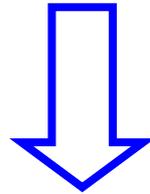
- $t_{n-1, \alpha}$  = valore di Student-t per n-1 gradi di libertà e livello di significatività  $\alpha$
- $\Delta$  = massimo errore accettabile (differenza tra media del campione e media della popolazione)
- $s$  = stima della deviazione standard della popolazione



# Tecniche classiche per dimensioni del campione

Sistemi Culturali

- Spesso le caratteristiche della popolazione sono sconosciute (variano da situazione a situazione)



E' **impossibile conoscere a priori l'errore** che può essere ritenuto **accettabile**: può essere più o meno grande a seconda del particolare campo sperimentale in esame e della sua storia.

In questi casi è **meglio analizzare la variazione relativa dell'errore** dovuto al campionamento **al crescere delle dimensioni del campione** in esame piuttosto che basarsi su criteri assoluti per l'accettazione o meno di un certo errore.



# Tecniche classiche per dimensioni del campione

Sistemi Culturali

- Con un **numero limitato di osservazioni** a disposizione (cosa che si verifica spesso nel caso di metodi lunghi o costosi) **può essere un azzardo valutare la normalità** (requisito per l'applicazione dei metodi classici per la determinazione delle dimensioni del campione).
- Può succedere che le **varianze non siano omogenee**.
- I metodi classici non prendono in considerazione **lo sforzo necessario** per ottenere il dato.



# Nuove metodologie per dimensionamento del campione

Sistemi Culturali

- Serve un metodo basato su **statistica non parametrica**: un metodo basato su **tecniche di ri-campionamento**.
- Il metodo **non** deve richiedere a priori informazioni
  - ✓ sulle caratteristiche della distribuzione e
  - ✓ sull'errore che siamo disposti ad accettare.



# Nuove metodologie per dimensionamento del campione

Sistemi Culturali

I metodi di ricampionamento si sono molto diffusi a partire dagli anni 60.

Concettualmente si basano sui metodi Monte Carlo ma si basano sull'**uso ripetuto dell'unico campione disponibile.**

I più importanti sono il bootstrap ed il jackknife.



# Nuove metodologie per dimensionamento del campione

Sistemi Culturali

Il visual jackknife!



# Nuove metodologie per dimensionamento del campione

Sistemi Culturali

- Il jackknife è basato sulla divisione del campione originale di  $N$  elementi in gruppi di  $k$  elementi.
- Possono essere generati  $N!/[(N-k)!k!]$  campioni virtuali (combinazioni senza ripetizioni) di  $N-k$  elementi eliminando ogni volta  $k$  valori differenti dal campione originale.
- Nel jackknife,  $k$  ha un valore unico ed in genere è molto piccolo rispetto ad  $N$  (es., se  $N$  è piccolo,  $k = 1$ ).
- Nel visual jackknife, invece, la generazione di campioni virtuali è ripetuta  $N-1$  volte, con  $k$  che assume valori da 1 a  $N-2$ , per un numero totale di campioni pari a:

$$\sum_{k=1}^{N-2} \frac{N!}{(N-k)!k!}$$



# Nuove metodologie per dimensionamento del campione

## Sistemi Culturali

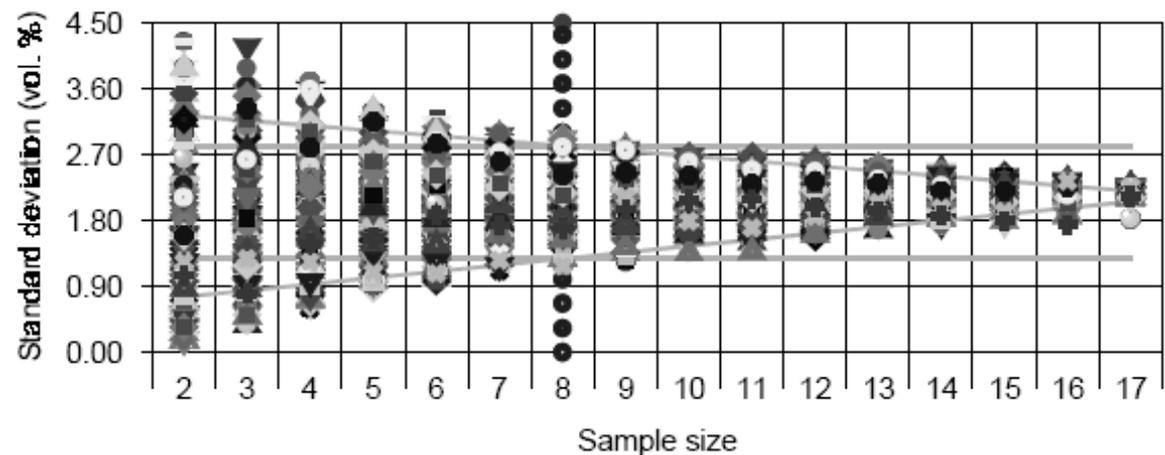
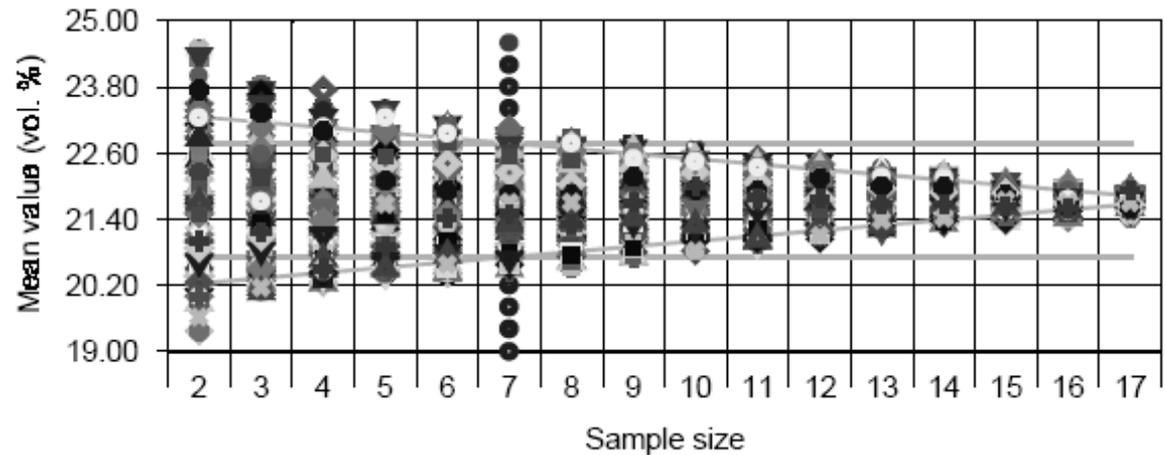
- Il numero di possibili campioni jackknife può diventare molto alto: gli autori del jackknife suggeriscono che ci si può fermare a 200-1000.
- Calcolare, per ogni campione generato di  $N-k$  unità campionarie, la media e la deviazione standard
- Disporre il tutto su due grafici che hanno:
  - ✓ in ascissa i valori di  $N-k$  (con  $k$  da  $N-2$  a  $1$ ; ovvero con  $(N-k)$  da  $2$  a  $(N-1)$ )
  - ✓ in ordinata le medie e le deviazioni standard



# Nuove metodologie per dimensionamento del campione

## Sistemi Culturali

- Si ottengono grafici di questo tipo.
- La dimensione ottimale del campione è quella oltre la quale la variabilità delle medie e delle deviazioni standard non decresce più sensibilmente al crescere di (N-k)
- Manualmente o con una procedura automatica





# Perché determinare il sample size? Esempio 1

Sistemi Colturali

Perché cambia con lo stadio di sviluppo, con la varietà e con le agrotecniche



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Field Crops Research 97 (2006) 135–141



[www.elsevier.com/locate/fcr](http://www.elsevier.com/locate/fcr)

Analysis of rice sample size variability due to development stage, nitrogen fertilization, sowing technique and variety using the visual jackknife

Roberto Confalonieri<sup>a,\*</sup>, Daniela Stroppiana<sup>b</sup>, Mirco Boschetti<sup>b,c</sup>,  
Davide Gusberti<sup>c</sup>, Stefano Bocchi<sup>c</sup>, Marco Acutis<sup>c</sup>

<sup>a</sup> *Institute for the Protection and Security of the Citizen, Joint Research Centre of the European Commission, AGRIFISH Unit, MARS-STAT Sector, TP 268 – 21020 Ispra (VA), Italy*

<sup>b</sup> *CNR-IREA, Institute for Electromagnetic Sensing of the Environment, Via Bassini 15, 20133 Milano, Italy*

<sup>c</sup> *Department of Crop Science, Section of Agronomy, University of Milano, Via Celoria 2, 20133 Milano, Italy*



# Perché determinare il sample size? Esempio 1

## Sistemi Colturali

Perché cambia con lo stadio di sviluppo, con la varietà e con le agrotecniche



\*: Development stage

§: kg<sub>N</sub> ha<sup>-1</sup>

(b)

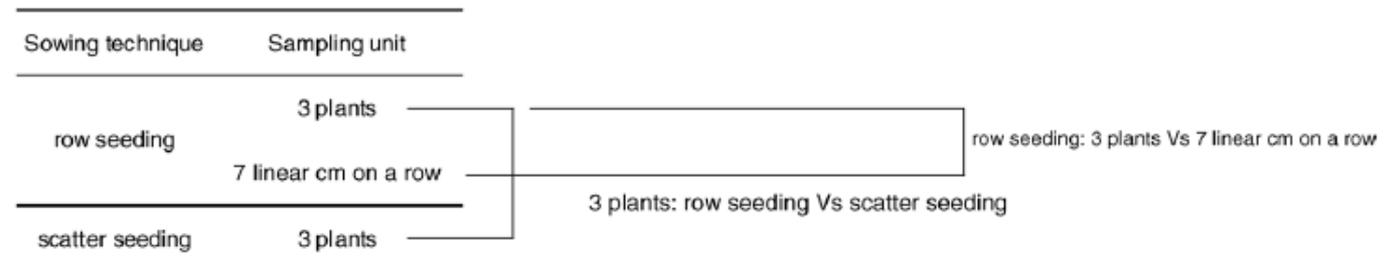


Fig. 2. Scheme of the comparisons among the sample sizes computed for the different situations. (a) Comparisons among development stages, varieties and levels of nitrogen fertilization. DVS1 and DVS2 represent tillering and stem elongation, respectively. (b) Comparisons among sowing techniques and sampling units.



# Perché determinare il sample size? Esempio 1

## Sistemi Colturali

Table 2

Comparisons among the sample sizes obtained according to the different sources of variability

		Variety			
		Volano		Gladio	
		DVS1 <sup>a</sup>	DVS2 <sup>a</sup>	DVS1 <sup>a</sup>	DVS2 <sup>a</sup>
N fertilization	N0	24	21	27	21
	N2	21	18	27	15
		Seeding technique <sup>b</sup>		Elementary observation <sup>c</sup>	
		Row	Scatter	Three plants	7 cm
Development stage	DVS1	27	21	27	33
	DVS2	18	18	18	30

<sup>a</sup> Development stage.

<sup>b</sup> Sampling unit = three plants.

<sup>c</sup> Only for row-seeded rice.

- aumenta il DVS => minore sample size
- aumenta l'N => minore sample size
- Gladio (DVS1) => maggiore sample size
- row => maggiore sample size
- cm-lineari => maggiore sample size



# Perché determinare il sample size? Esempio 2

Sistemi Colturali

Perché diverse variabili possono avere sample size molto diversi all'interno dello stesso sistema colturale

Field Crops Research 113 (2009) 125–130

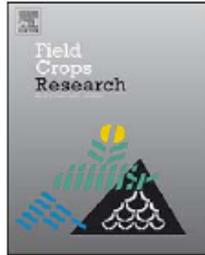


ELSEVIER

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: [www.elsevier.com/locate/fcr](http://www.elsevier.com/locate/fcr)



Analysis of sample size for variables related to plant, soil, and soil microbial respiration in a paddy rice field

Roberto Confalonieri <sup>a,\*</sup>, Alessia Perego <sup>b</sup>, Marcello Ermido Chiodini <sup>b</sup>, Barbara Scaglia <sup>b</sup>,  
Alexandra Stella Rosenmund <sup>b</sup>, Marco Acutis <sup>b</sup>

<sup>a</sup> European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Agriculture Unit, Agri4cast Action, via Fermi 2749, 21027 Ispra (VA), Italy

<sup>b</sup> Department of Crop Science, University of Milano, Via Celoria 2, 20133 Milano, Italy



# Perché determinare il sample size? Esempio 2

## Sistemi Colturali

Non si potevano usare tecniche di statistica parametrica!

**Table 1a**

Features of the pre-samplings carried out for the different plant-related variables. Shapiro–Wilk normality test was carried out for all the variables; Bartlett homoscedasticity test was carried out among the groups of coherent variables. Levene homoscedasticity test was carried out in case of deviation from normality.

Variable		Number of pre-sampling units <sup>*</sup>	Units	Mean	Standard deviation	Normality <sup>†</sup>	Homoscedasticity <sup>‡</sup>
Biomass	Total (aboveground)	54	kg ha <sup>-1</sup>	10250.36	2494.73		B–
	Leaves	54	kg ha <sup>-1</sup>	858.94	269.61		
	Stems	54	kg ha <sup>-1</sup>	6181.47	1509.92		
	Panicles	54	kg ha <sup>-1</sup>	3209.95	799.20		
Nitrogen concentration	Total (aboveground)	54	%	0.72	0.09		B–
	Leaves	54	%	0.66	0.11		
	Stems	54	%	0.49	0.09		
	Panicles	54	%	1.16	0.14		
Carbon concentration	Total (aboveground)	54	%	42.02	0.37		B–, L+
	Leaves	54	%	40.67	0.50		
	Stems	54	%	40.99	0.51		
	Panicles	54	%	44.36	0.58	S ( $P < 0.10$ )	
Spikelet sterility		54	%	14.78	7.01		–
Plant density		20	Plants m <sup>-2</sup>	478.00	114.00	S ( $P < 0.05$ )	–

<sup>\*</sup> Number of plants for variables describing plant features; number of determinations for plant density.

<sup>†</sup> S–indicates not normal according to the Shapiro–Wilk test; blanks indicate normality.

<sup>‡</sup> B and L–indicate not homoscedastic respectively according to the Bartlett and Levene test. The latter is used in case of deviation from normality; L+ indicates homoscedastic according to the Levene test; – indicates that homoscedasticity tests were not performed since the variable was not belonging to a group of coherent variables.



# Perché determinare il sample size? Esempio 2

## Sistemi Colturali

Non si potevano usare tecniche di statistica parametrica!

**Table 1b**

Features of the pre-samplings carried out for the variables related to soil and soil microbial activity. Shapiro–Wilk normality test was carried out for all the variables; Bartlett homoscedasticity test was carried out among the groups of coherent variables. Levene homoscedasticity test was carried out in case of deviation from normality.

Variable	Number of pre-sampling units*	Units	Mean	Standard deviation	Normality†	Homoscedasticity‡	
Soil microbial activity	9	mg CO <sub>2</sub> g DM <sup>-1</sup> 25 day <sup>-1</sup>	22.49	4.23		–	
N-NO <sub>3</sub> concentration	9	kg N-NO <sub>3</sub> <sup>-</sup> ha <sup>-1</sup>	2.86	1.36		B–	
N-NH <sub>4</sub> concentration	9	kg N-NH <sub>4</sub> <sup>+</sup> ha <sup>-1</sup>	3.53	0.94			
Total carbon concentration	9	%	1.42	0.12		B–, L–	
Total nitrogen concentration	9	%	0.13	0.01	S ( <i>P</i> < 0.10)		
Soil organic matter	9	%	2.45	0.20		–	
Water holding capacity	9	%	41.69	1.99		–	
pH (H <sub>2</sub> O)	9	–	5.69	0.12		B–	
pH (KCl)	9	–	4.62	0.14			
Texture	Sand	9	%	39.46	5.97		B–
	Clay	9	%	17.18	0.97		
	Silt	9	%	43.36	5.48		

\* Aggregated samples (four 125 cm<sup>3</sup> sub-samples).

† S–indicates not normal according to the Shapiro–Wilk test; blanks indicate normality.

‡ B and L–indicate not homoscedastic respectively according to the Bartlett and Levene test. The latter is used in case of deviation from normality; L+ indicates homoscedastic according to the Levene test; – indicates that homoscedasticity tests were not performed since the variable was not belonging to a group of coherent variables.



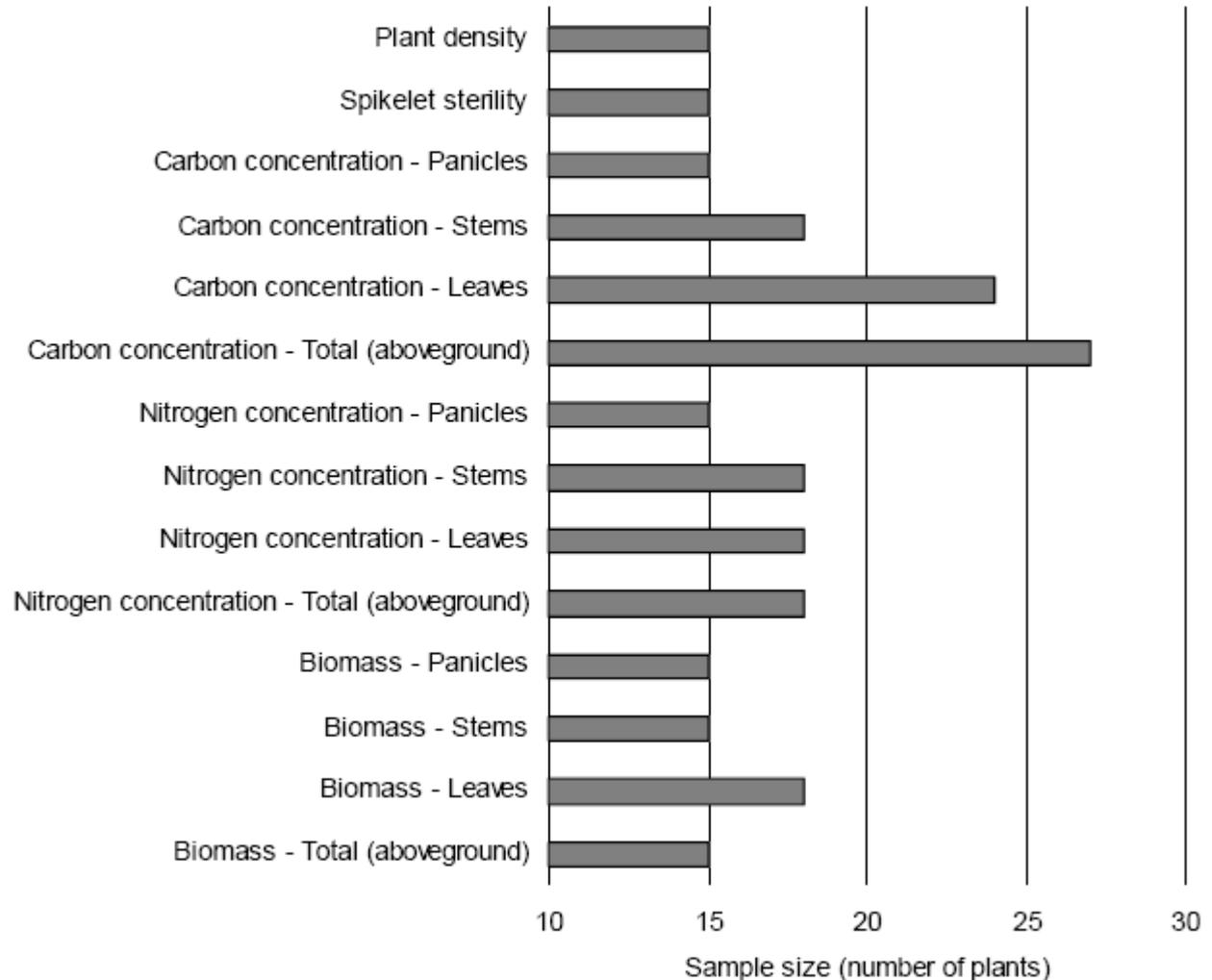
# Perché determinare il sample size? Esempio 2

Sistemi Colturali

Il sample size varia notevolmente a seconda della variabile indagata!

Sia per variabili che hanno a che fare con la pianta

Figure 1.a.



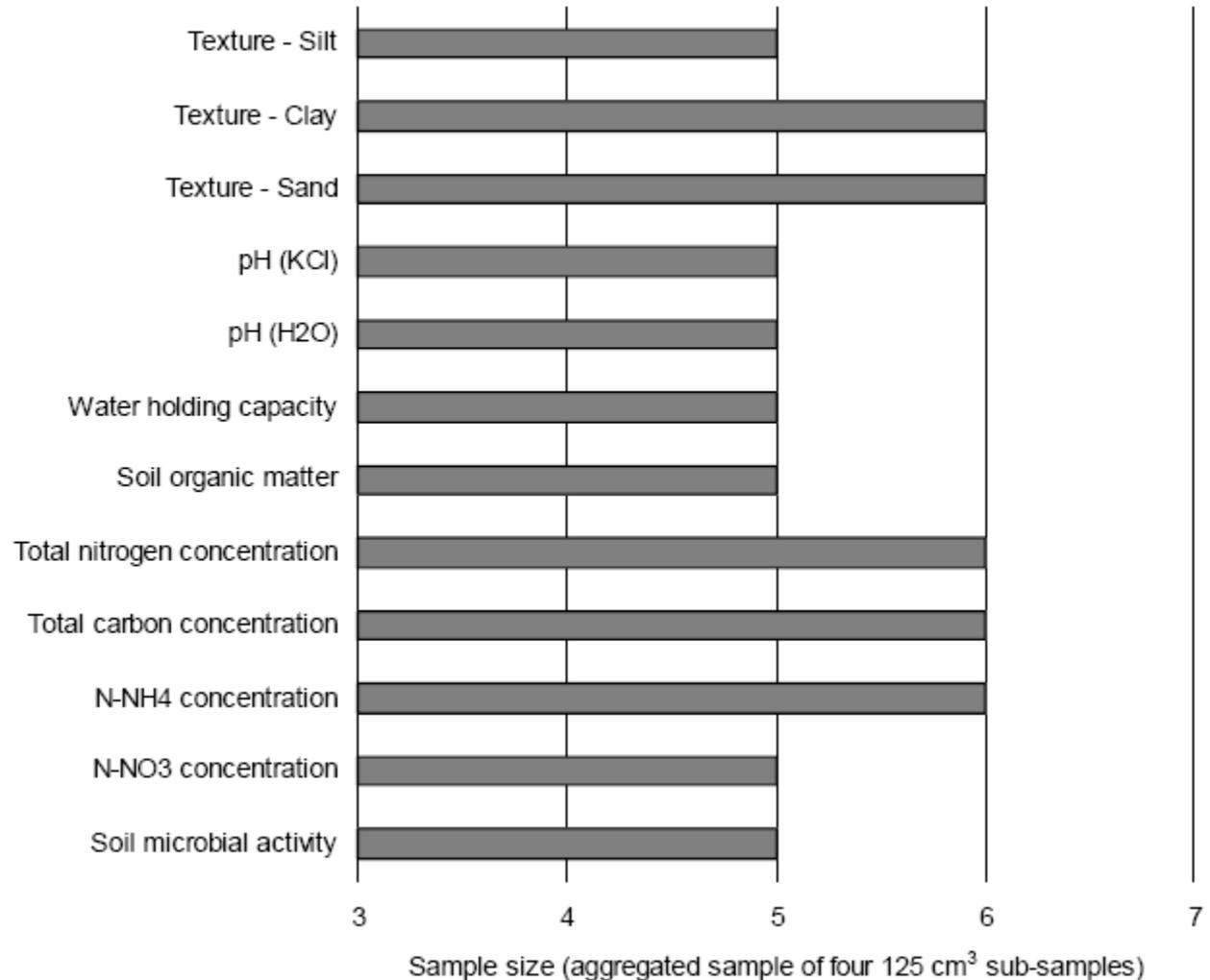


# Perché determinare il sample size? Esempio 2

Sistemi Culturali

...che per variabili che hanno a che fare con il suolo

Figure 1.b.





# Perché determinare il sample size? Esempio 2

Sistemi Culturali

- < sample size per  $[N]$  nelle cariossidi



# Perché determinare il sample size? Esempio 2

Sistemi Culturali

- < sample size per  $[N]$  nelle cariossidi (1 sink, molti source)



# Perché determinare il sample size? Esempio 2

Sistemi Colturali

- $<$  sample size per  $[N]$  nelle cariossidi (1 sink, molti source)
- $[C]$  nei vari organi della pianta presenta le maggiori differenze di sample size



# Perché determinare il sample size? Esempio 2

Sistemi Colturali

- $<$  sample size per  $[N]$  nelle cariossidi (1 sink, molti source)
- $[C]$  nei vari organi della pianta presenta le maggiori differenze di sample size (respirazione di mantenimento ( $\text{kg CH}_2\text{O kg}^{-1} \text{ day}^{-1}$ ) (temperatura dipendente) nelle foglie è circa 30 volte maggiore che negli steli e circa 7 volte maggiore che nelle cariossidi. Piccole variazioni di  $T$  (morfologia della canopy, "pozze") influenzano i tassi di respirazione e quindi la variabilità finale di  $[C]$ )
- Tra le biomasse, le foglie hanno anche  $>$  sample size rispetto agli altri organi (  )



Ok... interessante (spero)... ma chi li fa tutti quei conti (centinaia di combinazioni virtuali da generare, medie e deviazioni standard su di loro, analisi, ecc...)



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Environmental Modelling & Software 22 (2007) 1796–1800

---

---

Environmental  
Modelling & Software

---

---

[www.elsevier.com/locate/envsoft](http://www.elsevier.com/locate/envsoft)

Short communication

## Resampling-based software for estimating optimal sample size

R. Confalonieri <sup>a,\*</sup>, M. Acutis <sup>b</sup>, G. Bellocchi <sup>c</sup>, G. Genovese <sup>a</sup>

<sup>a</sup> *European Commission Directorate General Joint Research Centre, Institute for the Protection and Security of the Citizen, Agriculture and Fisheries Unit, via E. Fermi 1-TP 268, I-21020 Ispra (VA), Italy*

<sup>b</sup> *University of Milan, Department of Crop Science, via Celoria 2, I-20133 Milan, Italy*

<sup>c</sup> *European Commission Directorate General Joint Research Centre, Institute for Health and Consumer Protection, Biotechnology and GMOs Unit, via E. Fermi 1-TP 331, I-21020 Ispra (VA), Italy*

[http://www.robortoconfalonieri.it/software\\_download.htm](http://www.robortoconfalonieri.it/software_download.htm)