Metodologia sperimentale per le scienze agrarie

Andrea Onofri e Dario Sacco

Update: v. 1.0 (15/03/2021), compil. 2021-05-05

#### Indice

Pı	reme	ssa
	Obie	ettivi
		anizzazione
	_	ware statistico
		authors
		graziamenti
1	Scie	enza e pseudo-scienza 10
	1.1	$Scienza = dati \dots \dots$
	1.2	Dati 'buoni' e 'cattivi'
	1.3	Dati 'buoni' e metodi 'buoni'
	1.4	Il principio di falsificazione
	1.5	Falsificare un risultato
	1.6	Elementi fondamentali del disegno sperimentale
		1.6.1 Controllo degli errori
		1.6.2 Replicazione
		1.6.3 Randomizzazione
		1.6.4 Esperimenti invalidi
	1.7	Chi valuta se un esperimento è attendibile?
	1.8	Conclusioni
	1.9	Altre letture
<b>2</b>	Pro	gettare un esperimento 13
	2.1	Gli elementi della ricerca
	2.2	Ipotesi scientifica $\rightarrow$ obiettivo dell'esperimento 15
	2.3	Identificazione dei fattori sperimentali
		2.3.1 Esperimenti (multi-)fattoriali
		2.3.2 Controllo o testimone
	2.4	Le unità sperimentali
	2.5	Allocazione dei trattamenti

	2.6	Le variabili sperimentali
		2.6.1 Variabili nominali (categoriche)
		2.6.2 Variabili ordinali
		2.6.3 Variabili quantitative discrete
		2.6.4 Variabili quantitative continue
		2.6.5 Rilievi visivi e sensoriali
		2.6.6 Variabili di confondimento
	2.7	Esperimenti di campo
		2.7.1 Scegliere il campo
		2.7.2 Le unità sperimentali in campo
		2.7.3 Numero di repliche
		2.7.4 La mappa di campo
		2.7.5 Lay-out sperimentale
	2.8	Altre letture
0	D	
3		hiami di statistica descrittiva 16
	3.1	Dati quantitativi
		3.1.1 Indicatori di tendenza centrale
		1
	3.2	•
	3.2	Dati qualitativi
		3.2.2 Statistiche descrittive per le distribuzioni di frequenze . 17
		3.2.3 Distribuzioni di frequenza bivariate: le tabelle di con-
		tingenze
		3.2.4 Connessione
	3.3	Statistiche descrittive con R
	0.0	3.3.1 Descrizione dei sottogruppi
		3.3.2 Distribuzioni di frequenze e classamento
		3.3.3 Connessione
	3.4	Altre letture
	0.1	
4	Mo	delli statistici ed analisi dei dati 18
	4.1	Verità 'vera' e modelli deterministici
	4.2	Genesi deterministica delle osservazioni sperimentali 19
	4.3	Errore sperimentale e modelli stocastici
		4.3.1 Funzioni di probabilità
		4.3.2 Funzioni di densità
		4.3.3 La distribuzione normale (curva di Gauss) 19
	4.4	Modelli 'a due facce'

	4.5	E allora?	19							
	4.6	Le simulazioni Monte Carlo								
	4.7	Analisi dei dati e 'model fitting'								
	4.8	Modelli stocastici non-normali								
	4.9	Altre letture								
5	Stin	ne ed incertezza	<b>2</b> 0							
	5.1	Esempio: una soluzione erbicida	22							
		5.1.1 Analisi dei dati: stima dei parametri								
		5.1.2 La 'sampling distribution'	22							
		5.1.3 L'errore standard	22							
	5.2	Stima per intervallo	22							
	5.3	L'intervallo di confidenza	22							
	5.4	Qual è il senso dell'intervallo di confidenza?	22							
	5.5	Come presentare i risultati degli esperimenti	22							
	5.6	Alcune precisazioni	22							
		5.6.1 Campioni numerosi e non	22							
		5.6.2 Popolazioni gaussiane e non	22							
	5.7	Analisi statistica dei dati: riassunto del percorso logico								
	5.8	Da ricordare	22							
	5.9	Per approfondire un po'								
	5.10	Coverage degli intervalli di confidenza	22							
		5.10.1Intervalli di confidenza per fenomeni non-normali $$	22							
	5.11	Altre letture	22							
_	ъ.		23							
6	Decisioni ed incertezza									
	6.1	Confronto tra due medie: il test t di Student								
		6.1.1 L'ipotesi nulla e alternativa								
		6.1.2 La statistica T								
		6.1.3 Simulazione Monte Carlo								
		6.1.4 Soluzione formale								
		6.1.5 Interpretazione del P-level								
		6.1.6 Tipologie alternative di test t di Student								
	6.2	Confronto tra due proporzioni: il test $\chi^2$	24							
		6.2.1 Simulazione Monte Carlo	24							
		6.2.2 Soluzione formale	24							
	6.3	Conclusioni e riepilogo	24							
	6.4	Altre letture	24							
7	Mod	lelli ANOVA ad una via	25							
•		Caso-studio: confronto tra erbicidi in vaso	$\frac{20}{27}$							

	7.2	Descrizione del dataset
	7.3	Definizione di un modello lineare
	7.4	Parametrizzazione del modello
	7.5	Assunzioni di base
	7.6	Fitting del modello: metodo manuale
		7.6.1 Stima dei parametri
		7.6.2 Calcolo dei residui
		7.6.3 Stima di $\sigma$
	7.7	Scomposizione della varianza
	7.8	Test d'ipotesi
	7.9	Inferenza statistica
	7.10	Fitting del modello con R
	7.11	Medie marginali attese
		Per concludere
	7.13	Altre letture
8	La v	verifica delle assunzioni di base 28
	8.1	Violazioni delle assunzioni di base
	8.2	Procedure diagnostiche
	8.3	Analisi grafica dei residui
		8.3.1 Grafico dei residui contro i valori attesi
		8.3.2 QQ-plot
	8.4	Test d'ipotesi
	8.5	Risultati contraddittori
	8.6	'Terapia'
		8.6.1 Correzione/Rimozione degli outliers
		8.6.2 Correzione del modello
		8.6.3 Trasformazione della variabile indipendente 29
		8.6.4 Impiego di metodiche statistiche avanzate 29
		8.6.5 Trasformazioni stabilizzanti
	8.7	Esempio 1
	8.8	Esempio 2
	8.9	Altre letture
9	Con	trasti e confronti multipli 30
	9.1	Esempio
	9.2	I contrasti
	9.3	I contrasti con R
	9.4	I confronti multipli a coppie (pairwise comparisons) 31
	9.5	Display a lettere
	9.6	Tassi di errore per confronto e per esperimento

	9.7	Aggiustamento per la molteplicità	31
	9.8	E le classiche procedure di confronto multiplo?	31
	9.9	Consigli pratici	31
	9.10	Altre letture	31
10	Mod	delli ANOVA con fattori di blocco	32
		Caso-studio: confronto tra erbicidi in campo	33
		Definizione di un modello lineare	33
	10.3	Stima dei parametri	33
		10.3.1 Coefficienti del modello	33
		10.3.2 Stima di $\sigma$	33
		Scomposizione della varianza	33
		Adattamento del modello con R	33
		Disegni a quadrato latino	33
		Caso studio: confronto tra metodi costruttivi	33
	10.8	Definizione di un modello lineare	33
11	La r	regressione lineare semplice	34
	11.1	Caso studio: effetto della concimazione azotata al frumento	35
	11.2	Analisi preliminari	35
	11.3	Definizione del modello lineare	35
	11.4	Stima dei parametri	35
	11.5	Valutazione della bontà del modello	35
		11.5.1 Valutazione grafica	35
		11.5.2 Errori standard dei parametri	35
		11.5.3 Test F per la mancanza d'adattamento	35
		11.5.4 Test F per la bontà di adattamento e coefficiente di	
		determinazione	35
	11.6	Previsioni	35
	11.7	Altre letture	35
12	Mod	delli ANOVA a due vie con interazione	36
	12.1	Il concetto di 'interazione'	37
	12.2	Effetti incrociati: interazione tra lavorazioni e diserbo chimico	37
	12.3	Definizione del modello lineare	37
	12.4	Calcoli manuali	37
		12.4.1 Scomposizione della varianza	37
	12.5	Calcoli con R	37
		12.5.1 Model fitting	37
		12.5.2 Verifica delle assunzioni di base	37
		12.5.3 Scomposizione della varianza	37

		12.5.4 Medie marginali attese e confronti multipli con R	37
	12.6	Effetti innestati: valutazione di ibridi di mais	37
	12.7	Definizione del modello lineare	37
	12.8	Fitting del modello con R	37
	~		
13		t-plots, strip-plots e altri disegni sperimentali in campo	38
		Raggruppamenti tra parcelle	38
	13.2	Esperimenti a split-plot	38
		13.2.1 Definizione del modello lineare	38
		13.2.2 Model fitting con R	38
	13.3	Esperimenti a strip-plot	38
		13.3.1 Definizione del modello lineare	38
		13.3.2 Model fitting con R	38
	13.4	Altre letture	38
11	T o n	regressione non-lineare	39
14		Caso studio: degradazione di un erbicida nel terreno	41
		Scelta della funzione	41
		Stima dei parametri	41
	14.0	14.3.1 Linearizzazione della funzione	41
		14.3.2 Approssimazione della vera funzione tramite una poli-	41
		nomiale in X	41
			41
	111	14.3.3 Minimi quadrati non-lineari	41
		La regressione non-lineare con R	41
	14.3		
		14.5.1 Analisi grafica dei residui	41 41
		14.5.2 Test F per la mancanza di adattamento (approssimato)	
		14.5.3 Errori standard dei parametri	41
	110	14.5.4 Coefficienti di determinazione	41
		Funzioni lineari e nonlineari dei parametri	41
		Previsioni	
	14.8	Gestione delle situazioni 'patologiche'	41
		14.8.1 Trasformazione del modello	
	140	14.8.2 Trasformazione dei dati	41
	14.9	Per approfondire un po'	41
		14.9.1 Riparametrizzazione delle funzioni non-lineari	41
		14.9.2 Altre letture	41
15	Eser	rcizi	42
10		Capitoli 1 e 2	44
	10.1	15.1.1 Esercizio 1	

15.2 Ca	pitolo 3										44
15.	2.1 Esercizio	1									44
15.	2.2 Esercizio	2									44
15.	2.3 Esercizio	3		 							44
15.3 Ca	pitolo 4										44
15.	3.1 Esercizio	1		 							44
	3.2 Esercizio										
15.	3.3 Esercizio	3									44
15.	3.4 Esercizio	4									44
15.	3.5 Esercizio	5		 							44
15.	3.6 Esercizio	6		 							44
	3.7 Esercizio										
	3.8 Esercizio										
	pitolo 5										
	4.1 Esercizio										
15.	4.2 Esercizio	2		 							44
	4.3 Esercizio										
	4.4 Esercizio										
	4.5 Esercizio										
	pitolo 6										
	5.1 Esercizio										
	5.2 Esercizio										
	5.3 Esercizio										
	5.4 Esercizio										
	5.5 Esercizio										
	5.6 Esercizio										
	5.7 Esercizio										
	5.8 Esercizio										
	5.9 Esercizio										
	5.10 Esercizio										
	pitoli da 7 a 9										
	6.1 Esercizio										44
	6.2 Esercizio										44
	6.3 Esercizio										44
	6.4 Esercizio										44
	pitolo 10										44
	7.1 Esercizio										44
	7.2 Esercizio										44
	7.3 Esercizio										44
	pitolo 11										44
	8.1 Esercizio										44

	15.8.2 Esercizio 2	44
	15.9 Capitoli 12 e 13	44
	15.9.1 Esercizio 1	44
		44
		44
	15.9.4 Esercizio 4	44
	15.9.5 Esercizio 5	44
	15.9.6 Esercizio 6	44
	15.10Capitolo 14	44
	15.10.1 Esercizio 1	44
	15.10.2 Esercizio 2	44
		44
	15.10.4 Esercizio 4	44
	15.10.5 Esercizio 5	44
	15.10.6 Esercizio 6	44
	15.10.7 Esercizio 7	44
16	11	45
		46
	88	46
		46
		46
		46
	• 00	46
	1	46
	0	46
	0 1	46
	1 0	46
		46
		46
	00	46
	Alcune operazioni comuni sul dataset	46
	Selezionare un subset di dati	46
	Ordinare un vettore o un dataframe	46
	Workspace	46
	Script o programmi	46
	Interrogazione di oggetti	46
	Altre funzioni matriciali	46
	Cenni sulle funzionalità grafiche in R	46
	Altre letture	46

#### Premessa

Placeholder

Obiettivi

Organizzazione

Software statistico

The authors

Ringraziamenti

### Capitolo 1 Scienza e pseudo-scienza

- 1.1 Scienza = dati
- 1.2 Dati 'buoni' e 'cattivi'
- 1.3 Dati 'buoni' e metodi 'buoni'
- 1.4 Il principio di falsificazione
- 1.5 Falsificare un risultato
- 1.6 Elementi fondamentali del disegno sperimentale
- 1.6.1 Controllo degli errori
- 1.6.2 Replicazione
- 1.6.3 Randomizzazione
- 1.6.4 Esperimenti invalidi

Cattivo controllo degli errori

'Confounding' e correlazione spuria

Pseudo-repliche e randomizzazione poco attenta

- 1.7 Chi valuta se un esperimento è attendibile?
- 1.8 Conclusioni
- 1.9 Altre letture

### Progettare un esperimento

- 2.1 Gli elementi della ricerca
- 2.2 Ipotesi scientifica  $\rightarrow$  obiettivo dell'esperimento
- 2.3 Identificazione dei fattori sperimentali
- 2.3.1 Esperimenti (multi-)fattoriali
- 2.3.2 Controllo o testimone
- 2.4 Le unità sperimentali
- 2.5 Allocazione dei trattamenti
- 2.6 Le variabili sperimentali
- 2.6.1 Variabili nominali (categoriche)
- 2.6.2 Variabili ordinali
- 2.6.3 Variabili quantitative discrete
- 2.6.4 Variabili quantitative continue
- 2.6.5 Rilievi visivi e sensoriali
- 2.6.6 Variabili di confondimento
- 2.7 Esperimenti di campo
- 2.7.1 Scegliere il campo
- 2.7.2 Le unità sperimentali in campo
- 2.7.3 Numero di repliche
- 2.7.4 La mappa di campo
- 2.7.5 Lay-out sperimentale

#### Richiami di statistica descrittiva

#### 3.1 Dati quantitativi

- 3.1.1 Indicatori di tendenza centrale
- 3.1.2 Indicatori di dispersione
- 3.1.3 Incertezza delle misure derivate
- 3.1.4 Relazioni tra variabili quantitative: correlazione
- 3.2 Dati qualitativi
- 3.2.1 Distribuzioni di frequenze e classamento
- 3.2.2 Statistiche descrittive per le distribuzioni di frequenze
- 3.2.3 Distribuzioni di frequenza bivariate: le tabelle di contingenze
- 3.2.4 Connessione
- 3.3 Statistiche descrittive con R.
- 3.3.1 Descrizione dei sottogruppi
- 3.3.2 Distribuzioni di frequenze e classamento
- 3.3.3 Connessione
- 3.4 Altre letture

#### Modelli statistici ed analisi dei dati

- 4.1 Verità 'vera' e modelli deterministici
- 4.2 Genesi deterministica delle osservazioni sperimentali
- 4.3 Errore sperimentale e modelli stocastici
- 4.3.1 Funzioni di probabilità
- 4.3.2 Funzioni di densità
- 4.3.3 La distribuzione normale (curva di Gauss)
- 4.4 Modelli 'a due facce'
- 4.5 E allora?
- 4.6 Le simulazioni Monte Carlo
- 4.7 Analisi dei dati e 'model fitting'
- 4.8 Modelli stocastici non-normali
- 4.9 Altre letture

# Capitolo 5 Stime ed incertezza

- 5.1 Esempio: una soluzione erbicida
- 5.1.1 Analisi dei dati: stima dei parametri
- 5.1.2 La 'sampling distribution'
- 5.1.3 L'errore standard
- 5.2 Stima per intervallo
- 5.3 L'intervallo di confidenza
- 5.4 Qual è il senso dell'intervallo di confidenza?
- 5.5 Come presentare i risultati degli esperimenti
- 5.6 Alcune precisazioni
- 5.6.1 Campioni numerosi e non
- 5.6.2 Popolazioni gaussiane e non
- 5.7 Analisi statistica dei dati: riassunto del percorso logico
- 5.8 Da ricordare
- 5.9 Per approfondire un po'...
- 5.10 Coverage degli intervalli di confidenza
- 5.10.1 Intervalli di confidenza per fenomeni non-normali
- 5.11 Altre letture

#### Decisioni ed incertezza

6.1	Confronto tra due medie: il te	est t	$\operatorname{di}$	Stu-
	dent			

- 6.1.1 L'ipotesi nulla e alternativa
- 6.1.2 La statistica T
- 6.1.3 Simulazione Monte Carlo
- 6.1.4 Soluzione formale
- 6.1.5 Interpretazione del P-level
- 6.1.6 Tipologie alternative di test t di Student
- 6.2 Confronto tra due proporzioni: il test  $\chi^2$
- 6.2.1 Simulazione Monte Carlo
- 6.2.2 Soluzione formale
- 6.3 Conclusioni e riepilogo
- 6.4 Altre letture

#### Modelli ANOVA ad una via

- 7.1 Caso-studio: confronto tra erbicidi in vaso
- 7.2 Descrizione del dataset
- 7.3 Definizione di un modello lineare
- 7.4 Parametrizzazione del modello
- 7.5 Assunzioni di base
- 7.6 Fitting del modello: metodo manuale
- 7.6.1 Stima dei parametri
- 7.6.2 Calcolo dei residui
- 7.6.3 Stima di  $\sigma$
- 7.7 Scomposizione della varianza
- 7.8 Test d'ipotesi
- 7.9 Inferenza statistica
- 7.10 Fitting del modello con R
- 7.11 Medie marginali attese
- 7.12 Per concludere ...
- 7.13 Altre letture

#### La verifica delle assunzioni di base

- 8.1 Violazioni delle assunzioni di base
- 8.2 Procedure diagnostiche
- 8.3 Analisi grafica dei residui
- 8.3.1 Grafico dei residui contro i valori attesi
- 8.3.2 QQ-plot
- 8.4 Test d'ipotesi
- 8.5 Risultati contraddittori
- 8.6 'Terapia'
- 8.6.1 Correzione/Rimozione degli outliers
- 8.6.2 Correzione del modello
- 8.6.3 Trasformazione della variabile indipendente
- 8.6.4 Impiego di metodiche statistiche avanzate
- 8.6.5 Trasformazioni stabilizzanti
- 8.7 Esempio 1
- 8.8 Esempio 2
- 8.9 Altre letture

#### Contrasti e confronti multipli

- 9.1 Esempio
- 9.2 I contrasti
- 9.3 I contrasti con R
- 9.4 I confronti multipli a coppie (pairwise comparisons)
- 9.5 Display a lettere
- 9.6 Tassi di errore per confronto e per esperimento
- 9.7 Aggiustamento per la molteplicità
- 9.8 E le classiche procedure di confronto multiplo?
- 9.9 Consigli pratici
- 9.10 Altre letture

## Modelli ANOVA con fattori di blocco

- 10.1 Caso-studio: confronto tra erbicidi in campo
- 10.2 Definizione di un modello lineare
- 10.3 Stima dei parametri
- 10.3.1 Coefficienti del modello
- 10.3.2 Stima di  $\sigma$
- 10.4 Scomposizione della varianza
- 10.5 Adattamento del modello con R
- 10.6 Disegni a quadrato latino
- 10.7 Caso studio: confronto tra metodi costruttivi
- 10.8 Definizione di un modello lineare

# La regressione lineare semplice

- 11.1 Caso studio: effetto della concimazione azotata al frumento
- 11.2 Analisi preliminari
- 11.3 Definizione del modello lineare
- 11.4 Stima dei parametri
- 11.5 Valutazione della bontà del modello
- 11.5.1 Valutazione grafica
- 11.5.2 Errori standard dei parametri
- 11.5.3 Test F per la mancanza d'adattamento
- 11.5.4 Test F per la bontà di adattamento e coefficiente di determinazione
- 11.6 Previsioni
- 11.7 Altre letture

### Modelli ANOVA a due vie con interazione

- 12.1 Il concetto di 'interazione'
- 12.2 Effetti incrociati: interazione tra lavorazioni e diserbo chimico
- 12.3 Definizione del modello lineare
- 12.4 Calcoli manuali
- 12.4.1 Scomposizione della varianza
- 12.5 Calcoli con R
- 12.5.1 Model fitting
- 12.5.2 Verifica delle assunzioni di base
- 12.5.3 Scomposizione della varianza
- 12.5.4 Medie marginali attese e confronti multipli con R
- 12.6 Effetti innestati: valutazione di ibridi di mais
- 12.7 Definizione del modello lineare
- 12.8 Fitting del modello con R

#### Split-plots, strip-plots e altri disegni sperimentali in campo

13.1	Raggruppamenti	tra	parcel	lle
		OI CL	Pott CC:	

- 13.2 Esperimenti a split-plot
- 13.2.1 Definizione del modello lineare
- 13.2.2 Model fitting con R
- 13.3 Esperimenti a strip-plot
- 13.3.1 Definizione del modello lineare
- 13.3.2 Model fitting con R
- 13.4 Altre letture

#### La regressione non-lineare

14.1	Caso studio:	degradazione	$\operatorname{di}$	un	erbici-
	da nel terren	.0			

- 14.2 Scelta della funzione
- 14.3 Stima dei parametri
- 14.3.1 Linearizzazione della funzione
- 14.3.2 Approssimazione della vera funzione tramite una polinomiale in X
- 14.3.3 Minimi quadrati non-lineari
- 14.4 La regressione non-lineare con R
- 14.5 Verifica della bontà del modello
- 14.5.1 Analisi grafica dei residui
- 14.5.2 Test F per la mancanza di adattamento (approssimato)
- 14.5.3 Errori standard dei parametri
- 14.5.4 Coefficienti di determinazione
- 14.6 Funzioni lineari e nonlineari dei parametri
- 14.7 Previsioni
- 14.8 Gestione delle situazioni 'patologiche'
- 14.8.1 Trasformazione del modello
- 14.8.2 Trasformazione dei dati
- 14.9 Per approfondire un po'...

#### Esercizi

#### 15.1 Capitoli 1 e 2

- 15.1.1 Esercizio 1
- 15.2 Capitolo 3
- 15.2.1 Esercizio 1
- 15.2.2 Esercizio 2
- 15.2.3 Esercizio 3
- 15.3 Capitolo 4
- 15.3.1 Esercizio 1
- 15.3.2 Esercizio 2
- 15.3.3 Esercizio 3
- 15.3.4 Esercizio 4
- 15.3.5 Esercizio 5
- 15.3.6 Esercizio 6
- 15.3.7 Esercizio 7
- 15.3.8 Esercizio 8
- 15.4 Capitolo 5
- 15.4.1 Esercizio 1
- 15.4.2 Esercizio 2
- 15.4.3 Esercizio 3
- 15.4.4 Esercizio 4
- 15.4.5 Esercizio 5

### Appendice 1: breve introduzione ad R

Cosa è R?

Oggetti e assegnazioni

Costanti e vettori

Matrici

**Dataframe** 

Quale oggetto sto utilizzando?

Operazioni ed operatori

Funzioni ed argomenti

Consigli per l'immissione di dati sperimentali

Immissione di numeri progressivi

Immissione dei codici delle tesi e dei blocchi

Immissione dei valori e creazione del datframe

Leggere e salvare dati esterni

Alcune operazioni comuni sul dataset

Selezionare un subset di dati

Ordinare un vettore o un dataframe

Workspace

Script o programmi

Interrogazione di oggetti