

# Metodologia sperimentale per le scienze agrarie

Andrea Onofri e Dario Sacco

Update: v. 0.9 (18/01/2020), compil. 2020-02-04

# Indice

<b>Introduzione</b>	<b>8</b>
Organizzazione del testo . . . . .	8
Gli autori . . . . .	8
Pre-requisiti . . . . .	8
Packages aggiuntivi . . . . .	8
<b>1 Scienza e pseudo-scienza</b>	<b>9</b>
1.1 Scienza = dati . . . . .	11
1.2 Dati ‘buoni’ e ‘cattivi’ . . . . .	11
1.2.1 L’errore sperimentale . . . . .	11
1.2.2 La variabilità naturale . . . . .	11
1.2.3 Il campionamento . . . . .	11
1.3 Scienza = metodo . . . . .	11
1.4 Elementi fondamentali del disegno sperimentale . . . . .	11
1.4.1 Controllo degli errori . . . . .	11
1.4.2 Replicazione . . . . .	11
1.4.3 Randomizzazione . . . . .	11
1.4.4 Esperimenti invalidi . . . . .	11
1.5 Scienza e falsificabilità . . . . .	11
1.6 Chi valuta se un esperimento è attendibile? . . . . .	11
1.7 Conclusioni . . . . .	11
1.8 Per approfondire un po’... . . . .	11
<b>2 Progettare un esperimento</b>	<b>12</b>
2.1 Ipotesi scientifica → obiettivo dell’esperimento . . . . .	12
2.2 Identificazione dei fattori sperimentali . . . . .	13
2.2.1 Esperimenti (multi)fattoriali . . . . .	13
2.2.2 L’importanza del controllo . . . . .	14
2.2.3 Fattori sperimentali di trattamento e di blocco . . . . .	15
2.3 Identificazione delle unità sperimentali . . . . .	15

2.3.1	Cornice di campionamento . . . . .	15
2.3.2	Scelta delle unità sperimentali . . . . .	16
2.3.3	Unità sperimentali in campo: le parcelle . . . . .	16
2.4	Allocazione dei trattamenti e disegno sperimentale . . . . .	20
2.4.1	Disegni completamente randomizzati . . . . .	20
2.4.2	Disegni a blocchi randomizzati . . . . .	20
2.4.3	Disegni a quadrato latino . . . . .	23
2.5	Le variabili sperimentali . . . . .	24
2.5.1	Variabili nominali (categoriche) . . . . .	25
2.5.2	Variabili ordinali . . . . .	26
2.5.3	Variabili quantitative discrete . . . . .	26
2.5.4	Variabili quantitative continue . . . . .	26
2.5.5	Rilievi visivi e sensoriali . . . . .	27
2.5.6	Variabili di confondimento . . . . .	28
2.6	Impianto delle prove . . . . .	28
2.7	Scrivere un progetto/report di ricerca: semplici indicazioni . . . . .	29
2.8	Per approfondire un po'... . . . .	30
2.8.1	Esperimento 'cieco' e 'doppio cieco' . . . . .	30
<b>3</b>	<b>Modelli matematici ed analisi dei dati</b>	<b>31</b>
3.1	Verità 'vera' e modelli deterministici . . . . .	33
3.1.1	Qualche esempio di modello deterministico . . . . .	33
3.2	Genesi deterministica delle osservazioni sperimentali . . . . .	33
3.3	Errore sperimentale e modelli stocastici . . . . .	33
3.3.1	Funzioni di probabilità . . . . .	33
3.3.2	Funzioni di densità . . . . .	33
3.4	La distribuzione normale (curva di Gauss) . . . . .	33
3.5	Modelli 'a due facce' . . . . .	33
3.6	E allora? . . . . .	33
3.7	Le simulazioni Monte Carlo . . . . .	33
3.8	Analisi dei dati e 'model fitting' . . . . .	33
3.9	Per approfondire un po'... . . . .	33
3.9.1	Generazione dei dati sperimentali: un esempio più realistico . . . . .	33
3.9.2	Modelli stocastici non-normali . . . . .	33
3.9.3	Altre letture . . . . .	33
<b>4</b>	<b>Stime ed incertezza</b>	<b>34</b>
4.1	L'analisi dei dati: gli 'ingredienti' fondamentali . . . . .	36
4.2	Esempio: una soluzione erbicida . . . . .	36
4.2.1	Il modello dei dati . . . . .	36

4.2.2	Analisi dei dati: stima dei parametri . . . . .	36
4.2.3	La ‘sampling distribution’ . . . . .	36
4.2.4	L’errore standard . . . . .	36
4.3	Riepilogo 1: Caratterizzare l’incertezza di un esperimento . . .	36
4.4	L’intervallo di confidenza . . . . .	36
4.5	Qual è il senso dell’intervallo di confidenza? . . . . .	36
4.6	Come presentare i risultati degli esperimenti . . . . .	36
4.7	Alcune precisazioni . . . . .	36
4.7.1	Campioni numerosi e non . . . . .	36
4.7.2	Popolazioni gaussiane e non . . . . .	36
4.8	Analisi statistica dei dati: riassunto del percorso logico . . . .	36
4.9	Da ricordare . . . . .	36
4.10	Esercizi . . . . .	36
4.11	Per approfondire un po’... . . . .	36
4.12	<i>Coverage</i> degli intervalli di confidenza . . . . .	36
4.12.1	Intervalli di confidenza per fenomeni non-normali . . . .	36
4.12.2	Altre letture . . . . .	36
<b>5</b>	<b>Decisioni e incertezza</b>	<b>37</b>
5.1	Confronto tra una media osservata e una media teorica . . . .	38
5.1.1	Simulazione Monte Carlo . . . . .	38
5.1.2	Soluzione formale . . . . .	38
5.1.3	Interpretazione del P-level . . . . .	38
5.2	Confronto tra due medie: il test t di Student . . . . .	38
5.3	Confronto tra due proporzioni: il test $\chi^2$ . . . . .	38
5.4	Conclusioni . . . . .	38
5.5	Riepilogo . . . . .	38
5.6	Esercizi . . . . .	38
5.7	Per approfondire un po’... . . . .	38
5.7.1	Tipologie alternative di test t di Student . . . . .	38
5.7.2	Simulazione del P-value nel test d’ipotesi . . . . .	38
5.7.3	Altre letture . . . . .	38
<b>6</b>	<b>Modelli ANOVA ad una via</b>	<b>39</b>
6.1	Caso-studio: confronto tra erbicidi in vaso . . . . .	41
6.2	Descrizione del dataset . . . . .	41
6.3	Definizione di un modello lineare . . . . .	41
6.4	Parametrizzazione del modello . . . . .	41
6.4.1	Vincolo sul trattamento . . . . .	41
6.4.2	Vincolo sulla somma . . . . .	41
6.5	Assunzioni di base . . . . .	41

6.6	Stima dei parametri . . . . .	41
6.6.1	Coefficienti del modello . . . . .	41
6.6.2	Residui . . . . .	41
6.6.3	Stima di $\sigma$ . . . . .	41
6.6.4	SEM e SED . . . . .	41
6.7	Scomposizione della varianza . . . . .	41
6.8	Test d'ipotesi . . . . .	41
6.9	Operazioni con R . . . . .	41
6.10	Medie marginali attese . . . . .	41
6.11	Per concludere . . . . .	41
6.12	Per approfondire un po' . . . . .	41
6.12.1	Simulazione della <i>sampling distribution</i> per F . . . . .	41
6.12.2	Altre letture . . . . .	41
<b>7</b>	<b>La verifica delle assunzioni di base</b>	<b>42</b>
7.1	Procedure diagnostiche . . . . .	43
7.2	Analisi grafica dei residui . . . . .	43
7.2.1	Grafico dei residui contro i valori attesi . . . . .	43
7.2.2	QQ-plot . . . . .	43
7.3	Strumenti diagnostici formali . . . . .	43
7.4	Risultati contraddittori . . . . .	43
7.5	'Terapia' . . . . .	43
7.5.1	Correzione/Rimozione degli outliers . . . . .	43
7.5.2	Correzione del modello . . . . .	43
7.5.3	Trasformazione della variabile indipendente . . . . .	43
7.5.4	Impiego di metodiche statistiche avanzate . . . . .	43
7.5.5	Trasformazioni stabilizzanti . . . . .	43
7.6	Esempio . . . . .	43
7.7	Per approfondire un po' . . . . .	43
<b>8</b>	<b>Contrasti e confronti multipli</b>	<b>44</b>
8.1	Esempio . . . . .	46
8.2	I contrasti . . . . .	46
8.2.1	Varianza del contrasto e test d'ipotesi . . . . .	46
8.2.2	I contrasti con R . . . . .	46
8.3	I confronti multipli a coppie (pairwise comparisons) . . . . .	46
8.4	Display a lettere . . . . .	46
8.5	Problemi di molteplicità: tassi di errore per confronto e per esperimento . . . . .	46
8.5.1	Correzione per la molteplicità . . . . .	46
8.6	E le classiche procedure di confronto multiplo? . . . . .	46

8.7	Consigli pratici . . . . .	46
8.8	Per approfondire un po'... . . . .	46
8.8.1	Intervallo di confidenza di un contrasto . . . . .	46
8.8.2	Correzione per la molteplicità: qualche dettaglio ulteriore . . . . .	46
8.8.3	Altre letture . . . . .	46
<b>9</b>	<b>Modelli ANOVA con fattori di blocco</b>	<b>47</b>
9.1	Caso-studio: confronto tra erbicidi in campo . . . . .	48
9.2	Definizione di un modello lineare . . . . .	48
9.3	Stima dei parametri . . . . .	48
9.3.1	Coefficienti del modello . . . . .	48
9.3.2	Residui e devianze . . . . .	48
9.4	Scomposizione della varianza . . . . .	48
9.5	Disegni a quadrato latino . . . . .	48
9.6	Caso studio: confronto tra metodi costruttivi . . . . .	48
9.7	Definizione di un modello lineare . . . . .	48
<b>10</b>	<b>Modelli ANOVA a due vie</b>	<b>49</b>
10.1	Il concetto di 'interazione' . . . . .	50
10.2	Tipi di interazione . . . . .	50
10.3	Caso-studio: interazione tra lavorazioni e diserbo chimico . . . . .	50
10.4	Definizione del modello lineare . . . . .	50
10.5	Stima dei parametri . . . . .	50
10.6	Verifica delle assunzioni di base . . . . .	50
10.7	Scomposizione delle varianze . . . . .	50
10.8	Medie marginali attese . . . . .	50
10.9	Calcolo degli errori standard (SEM e SED) . . . . .	50
10.10	Medie marginali attese e confronti multipli con R . . . . .	50
10.11	Per approfondire un po'.... . . . .	50
10.11.1	Anova a due vie: scomposizione 'manuale' della varianza . . . . .	50
<b>11</b>	<b>La regressione lineare semplice</b>	<b>51</b>
11.1	Caso studio: effetto della concimazione azotata al frumento . . . . .	52
11.2	Analisi preliminari . . . . .	52
11.3	Definizione del modello lineare . . . . .	52
11.4	Stima dei parametri . . . . .	52
11.5	Valutazione della bontà del modello . . . . .	52
11.5.1	Valutazione grafica . . . . .	52
11.5.2	Errori standard dei parametri . . . . .	52
11.5.3	Test F per la mancanza d'adattamento . . . . .	52

11.5.4 Test F per la bontà di adattamento e coefficiente di determinazione . . . . .	52
11.6 Previsioni . . . . .	52
11.7 Per approfondire un po'... . . . .	52
<b>12 La regressione non-lineare</b>	<b>53</b>
12.1 Caso studio: degradazione di un erbicida nel terreno . . . . .	55
12.2 Scelta della funzione . . . . .	55
12.3 Stima dei parametri . . . . .	55
12.3.1 Linearizzazione della funzione . . . . .	55
12.3.2 Approssimazione della vera funzione tramite una polinomiale in X . . . . .	55
12.3.3 Minimi quadrati non-lineari . . . . .	55
12.4 La regressione non-lineare con R . . . . .	55
12.5 Verifica della bontà del modello . . . . .	55
12.5.1 Analisi grafica dei residui . . . . .	55
12.5.2 Test F per la mancanza di adattamento (approssimato)	55
12.5.3 Errori standard dei parametri . . . . .	55
12.5.4 Coefficienti di determinazione . . . . .	55
12.6 Funzioni lineari e nonlineari dei parametri . . . . .	55
12.7 Previsioni . . . . .	55
12.8 Gestione delle situazioni 'patologiche' . . . . .	55
12.8.1 Trasformazione del modello . . . . .	55
12.8.2 Trasformazione dei dati . . . . .	55
12.9 Per approfondire un po'... . . . .	55
12.9.1 Riparametrizzazione delle funzioni non-lineari . . . . .	55
12.9.2 Altre letture . . . . .	55
<b>13 Appendice 1: breve introduzione ad R</b>	<b>56</b>
Cosa è R? . . . . .	57
Oggetti e assegnazioni . . . . .	57
Costanti e vettori . . . . .	57
Matrici . . . . .	57
Dataframe . . . . .	57
Quale oggetto sto utilizzando? . . . . .	57
Operazioni ed operatori . . . . .	57
Funzioni ed argomenti . . . . .	57
Consigli per l'immissione di dati sperimentali . . . . .	57
Immissione manuale di dati . . . . .	57
Immissione di numeri progressivi . . . . .	57
Immissione dei codici delle tesi e dei blocchi . . . . .	57

Leggere e salvare dati esterni . . . . .	57
Alcune operazioni comuni sul dataset . . . . .	57
Selezionare un subset di dati . . . . .	57
Ordinare un vettore o un dataframe . . . . .	57
Workspace . . . . .	57
Script o programmi . . . . .	57
Interrogazione di oggetti . . . . .	57
Altre funzioni matriciali . . . . .	57
Cenni sulle funzionalità grafiche in R . . . . .	57
13.1 Per approfondire un po'... . . . .	57
<b>14 Appendice 2: Richiami di statistica descrittiva</b>	<b>58</b>
Dati quantitativi: analisi chimiche e altre misurazioni fondamentali	59
Indicatori di tendenza centrale . . . . .	59
Indicatori di dispersione . . . . .	59
Arrotondamenti . . . . .	59
Descrizione dei sottogruppi . . . . .	59
14.0.1 Relazioni tra variabili quantitative: correlazione . . . . .	59
Dati qualitativi: conteggi e frequenze . . . . .	59
Distribuzioni di frequenze e classamento . . . . .	59
Statistiche descrittive per le distribuzioni di frequenze . . . . .	59
Distribuzioni di frequenza bivariate: le tabelle di contingenza . . . . .	59
Connessione . . . . .	59
Esercizi . . . . .	59
Esercizio 1 . . . . .	59
Esercizio 2 . . . . .	59
14.1 Per approfondire un po'... . . . .	59



# Introduzione

Placeholder

Organizzazione del testo

Gli autori

Pre-requisiti

Packages aggiuntivi

# Capitolo 1

## Scienza e pseudo-scienza

Placeholder



## 1.1 Scienza = dati

## 1.2 Dati ‘buoni’ e ‘cattivi’

### 1.2.1 L’errore sperimentale

### 1.2.2 La variabilità naturale

### 1.2.3 Il campionamento

## 1.3 Scienza = metodo

## 1.4 Elementi fondamentali del disegno sperimentale

### 1.4.1 Controllo degli errori

### 1.4.2 Replicazione

### 1.4.3 Randomizzazione

### 1.4.4 Esperimenti invalidi

Cattivo controllo degli errori

‘Confounding’ e correlazione spuria

Pseudo-repliche e randomizzazione poco attenta

## 1.5 Scienza e falsificabilità

## 1.6 Chi valuta se un esperimento è attendibile?

## 1.7 Conclusioni

## 1.8 Per approfondire un po’...

## Capitolo 2

# Progettare un esperimento

Qualunque sia l'ambito scientifico, nella progettazione di un esperimento possiamo individuare alcune fasi fondamentali, che proviamo ad elencare:

1. individuazione del background (ricerca bibliografica)
2. definizione dell'ipotesi scientifica e dell'obiettivo;
3. identificazione dei fattore/i sperimentale/i;
4. identificazione dei soggetti sperimentali e delle repliche;
5. identificazione delle variabili da rilevare;
6. allocazione dei trattamenti;
7. impianto dell'esperimento.

Nell'analizzare questi aspetti, faremo riferimento ad alcuni esempi pratici, che verranno presentati tra poco.

### 2.1 Ipotesi scientifica $\rightarrow$ obiettivo dell'esperimento

Trascurando la parte di ricerca bibliografica, che è pur fondamentale, nel metodo scientifico galileiano, il punto di partenza di un esperimento è l'**ipotesi scientifica**, che determina l'obiettivo dell'esperimento. Si tratta del passaggio fondamentale dal quale dipende in modo logico tutto il lavoro successivo. Gli obiettivi debbono essere:

1. rilevanti;
2. chiaramente definiti;
3. specifici;

4. misurabili;
5. raggiungibili/realistici;
6. temporalmente organizzati.

Il rischio che si corre con obiettivi mal posti è quello di eseguire una ricerca dispersiva, con raccolta di dati non necessari e/o mancanza di dati fondamentali, con costi più elevati del necessario e un uso poco efficiente delle risorse. In genere, prima si definisce un obiettivo generale, seguito da uno o più obiettivi specifici, in genere proiettati su un più breve spazio temporale e che possono essere visti anche come le fasi necessarie per raggiungere l'obiettivo generale.

## 2.2 Identificazione dei fattori sperimentali

A differenza degli esperimenti **osservazionali**, più tipici delle scienze naturali, dove ci si limita ad osservare quanto accade in natura, nelle scienze agrarie si tende lavorare con esperimenti manipolati, dove cioè le unità sperimentali sono sottoposte a stimoli differenti a seconda della risposta che si vuole valutare. Dopo aver definito l'obiettivo di un esperimento, pertanto, è necessario chiarire esattamente gli stimoli a cui saranno sottoposte le unità sperimentali. Uno 'stimolo' sperimentale prende il nome di **fattore sperimentale**, che può avere più **livelli**. I livelli del fattore sperimentale prendono il nome di **trattamenti (o tesi) sperimentali**.

### 2.2.1 Esperimenti (multi)fattoriali

In alcuni casi è necessario inserire in prova più di un fattore sperimentale. In questo caso si parla di esperimenti **fattoriali**, che possono essere **incrociati (crossed)** quando sono presenti in prova tutte le possibili combinazioni dei livelli di ogni fattore, oppure di esperimenti **innestati (nested)** quando i livelli di un fattore cambiano al cambiare dei livelli dell'altro. Il vantaggio di questi tipi di esperimento è che permettono di valutare non solo l'effetto di ogni singolo fattore, ma anche la risposta congiunta di due fattori, che a volte può essere più che additiva (i due trattamenti insieme permettono un risultato superiore al vantaggio derivante singolarmente dai due trattamenti) o meno che additiva (i due trattamenti in qualche modo si elidono).

Ad esempio:

1. Immaginiamo di voler studiare due fattori sperimentali: la varietà di girasole (tre livelli: A, B e C) e la concimazione (2 livelli: pollino e urea). Abbiamo quindi 6 possibili trattamenti (combinazioni): A-pollina, A-urea, B-pollina, B-urea, C-pollina e C-urea. Il disegno è completamente incrociato.
2. Immaginiamo di voler confrontare due specie in agricoltura biologica (orzo e tritcale), con tre varietà ciascuna (A, B e C per orzo, D, E e F per tritcale). Anche in questo caso abbiamo sei trattamenti: orzo-A, orzo-B, orzo-C, tritcale-D, tritcale-E e tritcale-F, ma il disegno è innestato, perché per il fattore sperimentale ‘varietà’ i livelli cambiano a seconda dei livelli del fattore ‘specie’.

### 2.2.2 L’importanza del controllo

In alcuni casi si pone il problema di inserire in prova un trattamento che funga da riferimento per tutti gli altri. In questi casi si parla comunemente di **controllo** o **testimone**, che può essere

1. non sottoposto a trattamento
2. trattato con placebo
3. trattato secondo le modalità usuali di riferimento

Ad esempio, è usuale includere in un confronto varietale, la varietà di riferimento, che ci consente di capire se le prestazioni delle nuove varietà sono effettivamente interessanti oppure no.

Per quello che riguarda invece gli studi tossicologici, è evidente l’importanza di includere un controllo non trattato. Il placebo è una preparazione che contiene tutti gli ingredienti della formulazione attiva, meno che il principio attivo. Il placebo si rende necessario in una serie di circostanze, ad esempio:

1. quando il soggetto è influenzabile e può reagire alla semplice idea di essere stato trattato (effetto placebo)
2. quando i co-formulanti o la soluzione impiegata per veicolare il principio attivo possono mostrare un effetto sul soggetto.

Poiché gli esperimenti ci permettono non solo di capire se il comportamento di un livello di un fattore è uguale o differente da un altro, ma anche di misurare l’incremento che possiamo avere rispetto ad un altro, è molto importante il ruolo del controllo, perché esso diventa il riferimento rispet-

to al quale noi possiamo misurare il miglioramento dell'innovazione testata ('trattato').

### 2.2.3 Fattori sperimentali di trattamento e di blocco

Finora abbiamo menzionato quelli che, in lingua inglese, vengono definiti *treatment factor* (trattamenti sperimentali). Tuttavia, possono esserci altri fattori sperimentali non allocati, ma 'innati' e legati alla collocazione spazio-temporale o alle caratteristiche dei soggetti. Questi fattori vengono definiti, sempre in inglese, *blocking factors*. Di questi fanno parte, ad esempio, il blocco, la località ed ogni altro elemento che permette di raggruppare i soggetti. Anche questi *blocking factors* devono essere chiaramente identificati ed elencati.

## 2.3 Identificazione delle unità sperimentali

### 2.3.1 Cornice di campionamento

Prima di iniziare un esperimento, deve essere chiaro qual è la popolazione di riferimento (cornice di campionamento). Ad esempio, se vogliamo determinare la produttività delle varietà di tabacco e la loro adattabilità alle condizioni umbre, la cornice di campionamento sarà la media e alta valle del Tevere. Se invece vogliamo individuare la risposta di bovini adulti ad un certo tipo di alimentazione, dovremo campionare soggetti di entrambi i sessi, diverse età (ma comunque adulti) e in buone condizioni di salute. Insomma, la cornice di campionamento deve essere composta da soggetti che presentano tutte le caratteristiche richieste dall'obiettivo dell'esperimento.

La scelta della cornice di campionamento è fondamentale, perché è ad essa che si riferiscono i risultati ottenuti; ad esempio, se la cornice di campionamento è la media valle del Tevere, i risultati ottenuti non potranno essere riferiti alla Pianura Padana oppure al Tavoliere delle Puglie.



### 2.3.2 Scelta delle unità sperimentali

All'interno della cornice di campionamento, adottando criteri rigorosi, andremo a scegliere i soggetti da includere in prova. I soggetti dovranno essere rappresentativi della cornice di campionamento, ma, per il resto, il più possibile omogenei, per minimizzare le fluttuazioni casuali di risposta.

Per quanto riguarda la sperimentazione di pieno campo, l'omogeneità dell'ambiente è fondamentale per aumentare la precisione dell'esperimento, cosa che si consegue, innanzitutto, con la scelta dell'appezzamento giusto. Questa scelta è particolarmente delicata ed è guidata soprattutto dall'esperienza, tenendo conto anche di aspetti come la facilità di accesso e la vicinanza di strutture (laboratori, capannoni...), che consentano un'accurata esecuzione degli eventuali prelievi. Oltre a scegliere correttamente l'appezzamento, è importante anche porre in atto alcune operazioni che consentano di incrementare ulteriormente l'omogeneità dell'appezzamento prescelto. Ad esempio, talvolta si usa far precedere la prova da una coltura di 'omogeneizzazione', ad esempio avena, che è molto avida di azoto e lascia nel terreno poca fertilità residua. Oppure un prato di erba medica, che, grazie agli sfalci periodici, lascia il terreno libero da piante infestanti.

Gli sforzi operati al fine di rendere il più omogeneo possibile l'insieme delle unità sperimentali servono ad aumentare la capacità di rendere evidente l'effetto dei trattamenti praticati. Occorre tuttavia sottolineare che la variabilità è intrinseca nel sistema biologico e questa variabilità diventa il metro di giudizio per verificare se esiste veramente un effetto trattamento o no. Occorre pertanto assicurarsi che l'attività sperimentale non introduca (o introduca il meno possibile) ulteriore variabilità a quella normalmente presente in natura.

### 2.3.3 Unità sperimentali in campo: le parcelle

Solitamente, nella ricerca biologica, le unità sperimentali sono chiaramente definite: un animale, una persona, un insetto, un'aliquota di terreno. In genere, qualunque esse siano, debbono essere chiaramente identificate prima di procedere all'allocazione dei trattamenti.

Nella sperimentazione di pieno campo, le unità sperimentali sono costituite dalle cosiddette 'parcelle' (Figura 2.1), che, almeno inizialmente, non sono chiaramente identificate. La loro identificazione, dopo aver selezionato



Figura 2.1: Una prova sperimentale in campo (Foto D. Alberati)

l'apprezzamento giusto e averlo reso più uniforme possibile, viene usualmente eseguita su carta, redigendo la **mappa dell'esperimento**.

In primo luogo, si decide la **dimensione e la forma della parcella**. L'aspetto fondamentale è che ogni parcella deve contenere un numero di piante sufficientemente alto da essere rappresentativo. Per questo motivo le colture a bassa fittezza (es. mais) hanno sempre bisogno di parcelle più grandi che non quelle ad alta fittezza (es. frumento). La dimensione non deve tuttavia eccedere una certa soglia, in quanto con essa aumenta anche la variabilità del terreno e, di conseguenza, diminuisce l'omogeneità dell'esperimento. Per questo motivo, talvolta si preferisce diminuire la dimensione delle parcelle ed, avendo lo spazio sufficiente, aumentare il numero delle repliche. E' poi importante tenere conto di come saranno effettuare le pratiche colturali. Se queste saranno effettuate con normali attrezzature aziendali, occorre tenere conto della larghezza di lavoro degli organi lavoranti ed utilizzare multipli di questi.

Nello stabilire la dimensione delle parcelle, dovremo tener conto del fatto che la parte più delicata è il bordo, in quanto le piante che si trovano lungo il bordo esterno risentono di condizioni diverse dalle altre piante situate al centro della parcella (**effetto bordo**). Questo determina variabilità all'interno della parcella, che possiamo minimizzare raccogliendo solo la parte centrale. Si viene così a distinguere la superficie totale della parcella dalla superficie di raccolta (**superficie utile**), che può essere anche molto minore di quella

totale.

Tenendo conto degli aspetti detti in precedenza, riteniamo che le colture ad elevata fittezza (frumento, cereali, erba medica. . . ) dovrebbero avere parcelle di almeno 10-20 m<sup>2</sup>, mentre a bassa fittezza (mais, girasole. . . ) dovrebbero avere parcelle di almeno 20-40 m<sup>2</sup>. Queste dimensioni sono riferite alla superficie utile di raccolta, non alla dimensione totale: se si ritiene di dover raccogliere solo una parte della parcella per limitare l'effetto bordo, allora le dimensioni totali dovranno essere opportunamente aumentate, rispetto a quanto indicato sopra.

Per quanto riguarda la forma, le parcelle quadrate minimizzano l'effetto bordo, perché, a parità di superficie, hanno un perimetro più basso. Tuttavia esse sono di più difficile gestione, in quanto, considerando il fronte di lavoro di una seminatrice o una mietitrebbiatrice parcellare, possono richiedere la semina o la raccolta in più passate, il che finisce per essere una fonte di errore. Per questo motivo le parcelle sono usualmente rettangolari, con una larghezza pari a quella delle macchine operatrici.

Dopo aver stabilito la forma e la dimensione delle parcelle, si può procedere alla redazione della mappa, tenendo conto che il numero delle parcelle risulta dal prodotto tra il numero delle tesi sperimentali e il numero delle repliche.

In genere, si cerca di fare in modo che l'esperimento non sia troppo lungo (il che potrebbe aumentare la variabilità), ma neanche troppo largo, per evitare di avvicinarsi troppo alle scoline, dove possono manifestarsi ristagni idrici. Lungo il contorno della prova è possibile sistemare altre parcelle fuori esperimento con funzione di 'bordi'. In questo modo si evita che i bordi esterni delle parcelle esterne siano esposti a condizioni molto diverse dagli altri, cosa che potrebbe accentuare l'effetto 'bordo', di cui abbiamo parlato in precedenza. Queste parcelle di bordo verranno trattate in modo ordinario.

La figura 2.2 riporta la mappa di un esperimento sistemato su un appezzamento largo 30 metri e lungo 400 metri. In questo caso abbiamo disegnato otto file di parcelle in senso trasversale ( $8 \times 2.25 \text{ m} = 18 \text{ m}$  di larghezza), e quattro parcelle in senso longitudinale. Vediamo in figura che la mappa riporta tutte le informazioni relative al disegno sperimentale in modo da facilitare l'orientamento della mappa stessa. Un altro fondamentale aspetto è che le parcelle sono tutte chiaramente identificate con un numero.

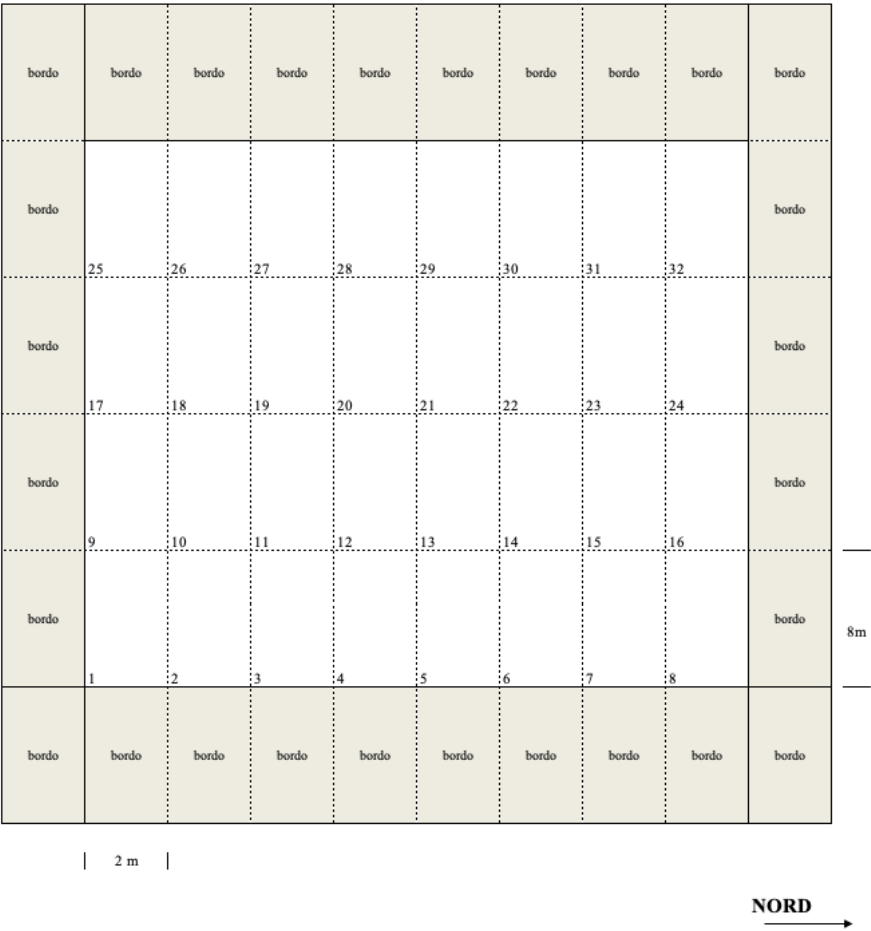


Figura 2.2: Mappa di campo di un esperimento con 32 parcelle

## 2.4 Allocazione dei trattamenti e disegno sperimentale

Il problema dell'allocazione dei trattamenti non si pone con gli esperimenti osservazionali, in quanto con questi si scelgono unità sperimentali già 'naturalmente' trattate.

Per tutti gli esperimenti manipolati si pone invece il problema di scegliere quali soggetti trattare e come. A seconda di quali vincoli introduciamo nella selezione dei soggetti possiamo avere diversi schemi sperimentali. Quelli che descriviamo di seguito, sono quelli più comunemente usati nella sperimentazione di pieno campo, ma, con le opportune modifiche, possono trovare impiego anche in molte altre discipline scientifiche.

### 2.4.1 Disegni completamente randomizzati

Per queste prove, le più semplici, la scelta dei soggetti da trattare è totalmente casuale, senza vincoli di sorta. Il vantaggio principale è la semplicità; lo svantaggio sta nel fatto che tutte le eventuali differenze e disomogeneità tra unità sperimentali restano non riconosciute ed entrano nella definizione della variabilità residua. Per questo, i disegni completamente randomizzati sono utilizzati soprattutto per le situazioni di buona uniformità ambientale e tra i soggetti.

Come esempio mostriamo un disegno completamente randomizzato utilizzando le parcelle della figura 2.2, dove abbiamo allocato 8 trattamenti (identificati con le lettere da A ad H) con quattro repliche. Come si può notare, l'allocazione è completamente casuale (figura 2.3)

### 2.4.2 Disegni a blocchi randomizzati

Quando le unità sperimentali non sono totalmente omogenee, ma vi è una certa variabilità per una qualche caratteristica rilevante, potremo dividere i soggetti in base a questa caratteristica, in tanti gruppi quante sono le repliche.

Ad esempio, nel caso dello schema in figura 2.2, è lecito aspettarsi un gradiente trasversale, dato che il campo sarà certamente meno fertile vicino alle scoline. Per questo motivo, dato che abbiamo scelto di fare quattro repliche,

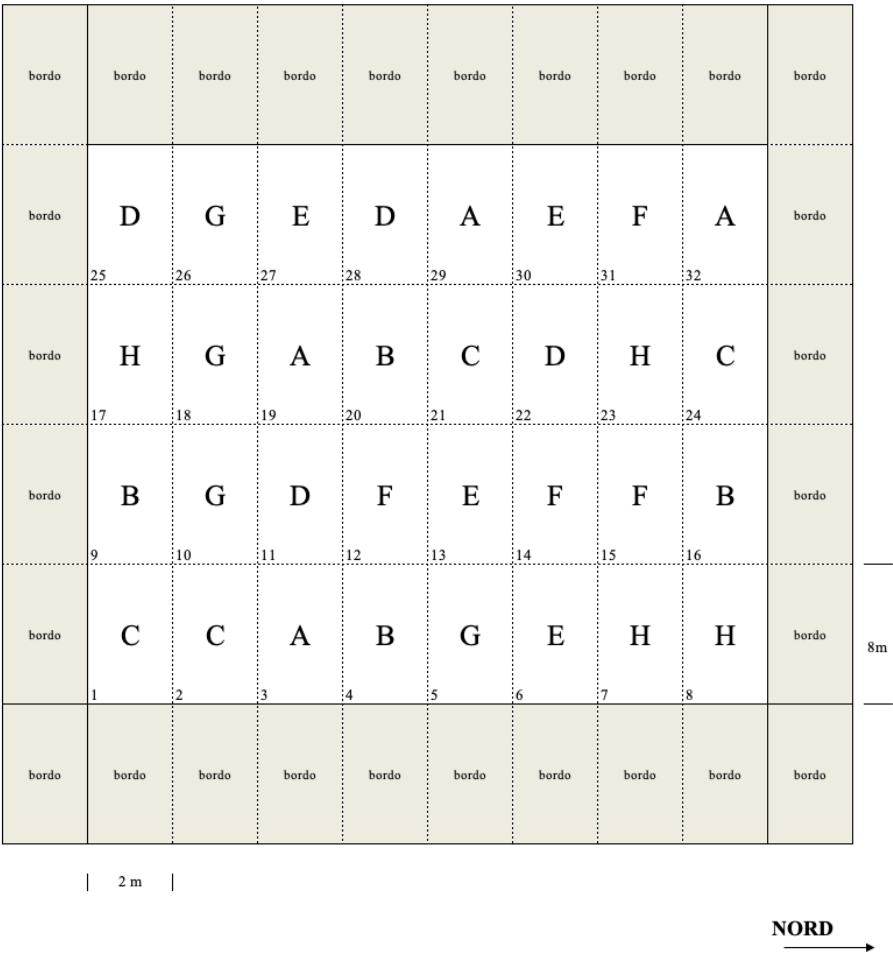


Figura 2.3: Esempio di uno schema sperimentale a randomizzazione completa

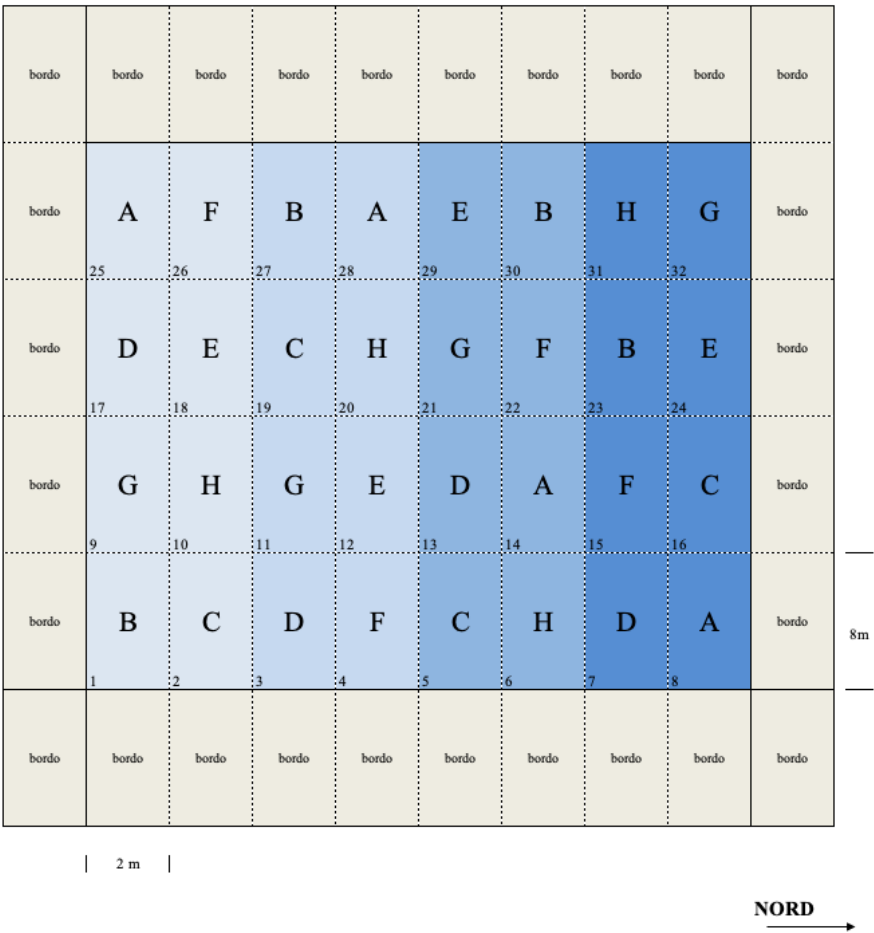


Figura 2.4: Esempio di uno schema sperimentale a blocchi randomizzati

divideremo l’appezzamento in quattro blocchi perpendicolari al gradiente di fertilità. Ad esempio il blocco 1 conterrà le parcelle 1, 9, 17, 25, 2, 10, 18 e 26, cioè le prime due colonne della mappa, con un numero di parcelle esattamente uguali al numero delle tesi. Il blocco 2 conterrà le colonne 3 e 4 e così via. Dato che il gradiente è trasversale, le parcelle di un stesso blocco saranno più omogenee che non parcelle su blocchi diversi. Dopo aver diviso la mappa in quattro blocchi di otto parcelle, possiamo allocare gli otto trattamenti a random all’interno di ogni blocco (2.4)

Un disegno a blocchi randomizzati non è solo tipico della sperimentazione di campo. Ad esempio, volendo determinare la contaminazione da micotossine nelle confezioni di datteri, a seconda della modalità di confezionamento (es.

carta, busta di plastica, scatola di plastica perforata), si può sospettare che il supermercato nel quale le confezioni vengono vendute potrebbe avere un certo effetto, legato alle modalità di conservazione. Per cui, invece che prelevare trenta confezioni (dieci per metodo) a caso nei supermercati di una città, scegliamo dieci supermercati e, in ognuno, prendiamo una confezione per tipo. In questo caso, il supermercato fa da blocco.

In generale, potremmo immaginare un esperimento con trattamenti che vengono allocati a caso agli animali di una stalla e ripetuti su stalle diverse, o a soggetti in gruppi omogenei di età e così via.

Il vantaggio del disegno a blocchi randomizzati sta nel fatto che ci permette di utilizzare soggetti sperimentali con una più bassa omogeneità iniziale, aspetto importante quando il numero di unità sperimentali richieste comincia ad essere elevato. Infatti, le differenze tra soggetti sperimentali, almeno in parte, possono essere spiegate attraverso l'appartenenza ad un determinato gruppo e possono quindi essere scorporate dal calcolo della variabilità residua.

### 2.4.3 Disegni a quadrato latino

In questo caso, le unità sperimentali presentano due 'gradienti', cioè vi sono differenze legate a due elementi importanti, oltre al trattamento sperimentale. Vediamo un esempio.

Una certo oggetto richiede un solo operatore per essere costruito, ma l'operazione può essere eseguita in quattro modi diversi. Vogliamo capire qual è il modo più veloce e, pertanto, pianifichiamo un esperimento. L'unità sperimentale è il lavoratore. I metodi sono quattro e, volendo lavorare con quattro repliche, avremmo bisogno di sedici operatori per disegnare un esperimento completamente randomizzato. Possiamo tuttavia considerare che un operatore, in quattro turni successivi, può operare con tutti e quattro i metodi. Quindi possiamo disegnare un esperimento in cui il turno fa da unità sperimentale e l'operatore fa da blocco (esperimento a blocchi randomizzati).

Tuttavia, in ogni blocco (operatore) vi è un gradiente, nel senso che i turni successivi al primo sono via via meno efficienti, perché l'operatore accumula stanchezza. Per tener conto di questo potremmo lasciare all'operatore un congruo periodo di tempo tra un turno e l'altro. Oppure, potremmo introdurre un vincolo ulteriore, per ogni operatore, randomizzando i quattro metodi tra i turni, in modo che ogni metodo, in operatori diversi, capiti in tutti i turni. In sostanza, l'operatore fa da blocco, perché in esso sono contenuti tutti i metodi. Ma anche il turno (per tutti gli operatori) fa da



	Operatore 1	Operatore 2	Operatore 3	Operatore 4
Turno 1	A	C	D	B
Turno 2	D	A	B	C
Turno 3	B	D	C	A
Turno 4	C	B	A	D

Figura 2.5: Allocazione di quattro metodi di lavoro (A, B, C e D), tra quattro operatori, in quattro turni, seguendo uno schema a quadrato latino

blocco, in quanto in esso sono ancora contenuti tutti i metodi. Proviamo a schematizzare, nella figura seguente (2.5 ).

Questo schema prende il nome di quadrato latino, in quanto il numero delle repliche è uguale al numero dei trattamenti e, schematizzando lo schema su una mappa, otteniamo una griglia quadrata, nella quale ogni trattamento occupa tutte le righe e tutte le colonne. Chi lo conosce, riconosce in questo schema i principi di fondo del Sudoku.

Questo disegno è utile, perché possiamo dar conto sia delle differenze tra righe (turni), che delle differenze tra colonne (operatori), in modo da ridurre al minimo possibile la variabilità inspiegata. Lo svantaggio sta nel fatto che, dovendo avere tante repliche quanti sono i trattamenti, è utilizzabile solo per esperimenti abbastanza piccoli.

[NOTA: Secondo me occorre inserire il disegno a split plot. Per ora non lo faccio per andare avanti e per decidere come vogliamo trattarlo. Metteresti i mixed in un corso magistrale?. Facendo una teoria leggera, secondo me si può fare, anche perchè introduce il concetto delle misure ripetute. Cosa facciamo?]

## 2.5 Le variabili sperimentali

Per ogni singolo carattere, l'insieme delle modalità/valori che ognuno dei soggetti presenta prende il nome di **variabile** (proprio perché varia, cioè assume diversi valori, a seconda del soggetto).

Fondamentalmente, esistono due grandi gruppi di variabili: quelle che definiscono lo ‘stimolo’ sperimentale e quelle che misurano la risposta dei soggetti trattati. Delle prime abbiamo già parlato, in quanto sono definite all’inizio dell’esperimento e rappresentano i livelli del/dei trattamento/i sperimentali. Ad esempio, fanno parte del primo gruppo le variabili come le modalità di lavorazione del suolo (aratura, minima lavorazione, semina su sodo) o il valore di azoto apportato (0, 50, 100, 150 kg N ha<sup>-1</sup>). Le variabili che definiscono uno ‘stimolo’ sperimentale sono anche dette variabili ‘indipendenti’ in quanto non dipendono da nessun altro interno all’esperimento, a parte la volontà dello sperimentatore.

Durante e al termine dell’esperimento vengono invece rilevate le variabili ‘risposta’ che esprimono l’effetto dei trattamenti sperimentali, come, ad esempio, la produzione della coltura, il peso delle piante trattate con un diserbante, la quantità di azoto lisciviato dopo la fertilizzazione. Queste variabili vengono normalmente definite ‘dipendenti’, in quanto i valori assunti da ogni unità sperimentale dipendono, normalmente’ dallo stimolo che hanno ricevuto.

Sia le variabili indipendenti sia quelle dipendenti possono essere di diversi tipi, che è molto importante saper riconoscere, perché condizionano il tipo di analisi statistica da eseguire. In particolare, distinguiamo:

1. variabili nominali (categoriche);
2. variabili ordinali;
3. variabili quantitative discrete;
4. variabili quantitative continue.

### 2.5.1 Variabili nominali (categoriche)

Le variabili nominali esprimono, per ciascun soggetto, l’appartenenza ad una determinata categoria o raggruppamento ed il valore che assumono lo definiamo modalità. L’unica caratteristica delle categorie è l’esclusività, cioè un soggetto che appartiene ad una di esse non può appartenere a nessuna delle altre. Variabili nominali sono, ad esempio, il sesso, la varietà, il tipo di diserbante impiegato, la lavorazione e così via. Le variabili categoriche permettono di raggruppare i soggetti, ma non possono essere utilizzate per fare calcoli, se non per definire le frequenze dei soggetti in ciascun gruppo.

### 2.5.2 Variabili ordinali

Anche le variabili ordinali esprimono, per ciascun soggetto, l'appartenenza ad una determinata categoria o raggruppamento. Tuttavia, le diverse categorie sono caratterizzate, oltre che dall'esclusività, anche da una relazione di ordine, nel senso che è possibile stabilire una naturale graduatoria tra esse. Ad esempio, la risposta degli agricoltori a domande relative alla loro percezione sull'utilità di una pratica agronomica può essere espressa utilizzando una scala con sei categorie (0, 1, 2, 3, 4 e 5), in ordine crescente da 0 a 5 (scala Likert). Di conseguenza possiamo confrontare categorie diverse ed esprimere un giudizio di ordine (2 è maggiore di 1, 3 è minore di 5), ma non possiamo eseguire operazioni matematiche, tipo sottrarre dalla categoria 3 la categoria 2 e così via, dato che la distanza tra le categorie non è necessariamente la stessa.

### 2.5.3 Variabili quantitative discrete

Le variabili discrete sono caratterizzate dal fatto che possiedono, oltre alle proprietà dell'esclusività e dell'ordine, anche quella dell'equidistanza tra gli attributi (es., in una scala a 5 punti, la distanza – o la differenza – fra 1 e 3 è uguale a quella fra 2 e 4 e doppia di quella tra 1 e 2). Una tipica variabile discreta è il conteggio di piante infestanti all'interno di una parcella di terreno.

Le variabili discrete consentono la gran parte delle operazioni matematiche e permettono di calcolare molte importanti statistiche come la media, la mediana, la varianza e la deviazione standard.

### 2.5.4 Variabili quantitative continue

Le variabili quantitative continue possiedono tutte le proprietà precedentemente esposte (esclusività delle categorie, ordine, distanza) oltre alla continuità, almeno in un certo intervallo. Tipiche variabili continue sono l'altezza, la produzione, il tempo, la fittezza. . .

Dato che gli strumenti di misura nella realtà sono caratterizzati da una certa risoluzione, si potrebbe arguire che misure su scala continua effettivamente non esistono. Tuttavia questo argomento è più teorico che pratico e, nella ricerca biologica, consideriamo continue tutte le misure nelle quali la risoluzione dello strumento è sufficientemente piccola rispetto alla grandezza da

misurare, purchè la scelta dello strumento sia effettuata in modo oculato. Viceversa, le variabili continue sono piuttosto rare nelle scienze economiche e sociali in genere.

La quantità di informazione fornita dagli strumenti di valutazione cresce passando dalle scale nominali, di più basso livello, a quelle quantitative continue, di livello più elevato. Variabili esprimibili con scale quantitative continue o discrete possono essere espresse anche con scale qualitative, adottando un'opportuna operazione di classamento. Il contrario, cioè trasformare in quantitativa una variabile qualitativa, non è invece possibile. Va tuttavia ricordato che le variabili discrete tendono ad assumere un comportamento simile alle continue quando il numero di valori che può essere assunto dalla variabile risulta molto ampio.

### 2.5.5 Rilievi visivi e sensoriali

Nella pratica sperimentale è molto frequente l'adozione di metodi di rilievo basati sull'osservazione di un fenomeno attraverso uno dei sensi (più spesso, la vista, ma anche gusto e olfatto) e l'assegnazione di una valutazione su scala categorica, ordinale o, con un po' di prudenza, quantitativa discreta o continua. Ad esempio, il ricoprimento delle piante infestanti, la percentuale di controllo di un erbicida e la sua fitotossicità vengono spesso rilevati visivamente, su scale da 0 a 100 o simili.

I vantaggi di questa tecnica sono molteplici:

1. Basso costo ed elevata velocità
2. Possibilità di tener conto di alcuni fattori perturbativi esterni, che sono esclusi dalla valutazione, contrariamente a quello che succede con metodi oggettivi di misura
3. non richiede strumentazione costosa

A questi vantaggi fanno da contraltare alcuni svantaggi, cioè:

1. Minor precisione (in generale)
2. Soggettività
3. L'osservatore può essere prevenuto
4. Difficoltà di mantenere uniformità di giudizio
5. Richiede esperienza specifica e allenamento

I rilievi sensoriali sono ben accettati nella pratica scientifica in alcuni ambiti ben definiti, anche se richiedono attenzione nell'analisi dei dati non potendo essere assimilati *tout court* con le misure oggettive su scala continua.

### 2.5.6 Variabili di confondimento

Quando si pianificano i rilievi da eseguire, oppure anche nel corso dell'esecuzione di un esperimento, bisogna tener presente non soltanto la variabile che esprime l'effetto del trattamento, ma anche tutte le variabili che misurano possibili fattori di confondimento.

Ad esempio, immaginiamo di voler valutare la produttività di una specie arborea in funzione della varietà. Immaginiamo anche di sapere che, per questa specie, la produttività dipende anche dall'età. Se facciamo un esperimento possiamo utilizzare alberi della stessa età per minimizzare la variabilità dei soggetti. Tuttavia, se questo non fosse possibile, per ogni albero dobbiamo rilevare non solo la produttività, ma anche l'età, in modo da poter valutare anche l'effetto di questo fattore aggiuntivo e separarlo dall'effetto della varietà. In questo modo l'esperimento diviene molto più preciso.

## 2.6 Impianto delle prove

Da questo punto in poi, subentrano le competenze agronomiche e fitopatologiche necessarie per condurre gli esperimenti. Occorre, tuttavia, ricordare alcune pratiche usuali nella sperimentazione di pieno campo, destinate a migliorare l'efficienza della prova.

1. Seminare a densità più alte e poi diradare, per assicurare una migliore uniformità di impianto
2. Prelevare da ogni parcella più campioni ed, eventualmente, omogeneizzarli o mediare i risultati ottenuti (vedi il caso dei 1000 semi)
3. Considerare le caratteristiche naturalmente meno variabili (es. la produzione areica e non la produzione per pianta)

Si vuole inoltre ricordare che gli esperimenti parcellari configurano una situazione nella quale, per l'elevata cura che si pone nelle tecniche agronomiche, la produttività è almeno del 20% superiore rispetto a quanto avviene nella normale pratica agricola.

## 2.7 Scrivere un progetto/report di ricerca: semplici indicazioni

Quanto abbiamo finora esposto costituisce uno schema generale che può essere adottato per redigere un progetto di ricerca o un report sui risultati ottenuti (tesi, pubblicazione). Bisogna provare che la ricerca che si è eseguita è precisa, accurata e replicabile/riproducibile e, di conseguenza, i risultati sono validi.

Nella redazione di un progetto di ricerca o di un report, è fondamentale tratteggiare bene i seguenti elementi:

1. Titolo della ricerca
2. Descrizione del problema e background scientifico
3. Ipotesi scientifica, motivazioni e obiettivi
4. Tipo di esperimento e durata
5. Disegno sperimentale: trattamenti sperimentali (tesi) a confronto con dettagli relativi all'applicazione
6. Unità sperimentali e criteri per la loro selezione. Dettagli su repliche e randomizzazione
7. Dettagli su eventuali tecniche di 'blocking'
8. Variabili da rilevare/rilevate
9. Dettagli su come le variabili saranno/sono state rilevate
10. Esposizione dei risultati (solo nel report)
11. Discussione (solo nel report)
12. Conclusioni (solo nel report)

Alcuni aspetti che divengono elemento di valutazione del progetto e/o del report sono i seguenti:

1. La selezione dei metodi deve essere coerente con gli obiettivi
  2. Descrizione dettagliata dei materiali e metodi (bisogna che chiunque sia in grado di replicare l'esperimento)
  3. Esposizione dei risultati chiara e convincente
  4. Discussione approfondita e con molti riferimenti alla letteratura per confermare quanto ottenuto in passato o giustificare risultati differenti.
-

## 2.8 Per approfondire un po'...

### 2.8.1 Esperimento 'cieco' e 'doppio cieco'

Per concludere questa parte, è opportuno menzionare il fatto che, in un esperimento scientifico, il fatto che lo sperimentatore e il soggetto siano coscienti del trattamento somministrato può portare a risultati distorti. Per esempio, nell'eseguire un rilievo, lo sperimentatore può essere influenzato dal sapere con quale diserbante è stata trattata una parcella, cercando inconsciamente conferme alle sue conoscenze pregresse. D'altro canto, nei soggetti sperimentali dotati di coscienza (uomo) sapere di essere stati trattati può influenzare l'esito del trattamento (effetto placebo).

Per evitare questi problemi, soprattutto in ambito medico, un esperimento può essere pianificato come:

1. cieco: l'unità sperimentale o lo sperimentatore non sono coscienti dei dettagli del trattamento;
2. doppio cieco: né l'unità sperimentale né lo sperimentatore sono a conoscenza dei dettagli del trattamento

Un esperimento cieco e/o doppio cieco possono non essere eticamente corretti oppure inutili, nel qual caso si torna ad un esperimento tradizionale 'aperto' (*open experiment*: Tutti sanno tutto')

1. Cochran, W.G., Cox, G.M., 1950. *Experimental design*. John Wiley & Sons, Inc., Books.
2. LeClerg, E.L., Leonard, W.H., Clark, A.G., 1962. *Field Plot Technique*. Burgess Publishing Company, Books.

## Capitolo 3

# Modelli matematici ed analisi dei dati

Placeholder





## 3.1 Verità ‘vera’ e modelli deterministici

### 3.1.1 Qualche esempio di modello deterministico

## 3.2 Genesi deterministica delle osservazioni sperimentali

## 3.3 Errore sperimentale e modelli stocastici

### 3.3.1 Funzioni di probabilità

### 3.3.2 Funzioni di densità

## 3.4 La distribuzione normale (curva di Gauss)

## 3.5 Modelli ‘a due facce’

## 3.6 E allora?

## 3.7 Le simulazioni Monte Carlo

## 3.8 Analisi dei dati e ‘model fitting’

## 3.9 Per approfondire un po’...

### 3.9.1 Generazione dei dati sperimentali: un esempio più realistico

### 3.9.2 Modelli stocastici non-normali

t di Student

F di Fisher

La distribuzione binomiale

### 3.9.3 Altre letture

# Capitolo 4

## Stime ed incertezza

Placeholder



- 4.1 L'analisi dei dati: gli 'ingredienti' fondamentali
- 4.2 Esempio: una soluzione erbicida
  - 4.2.1 Il modello dei dati
  - 4.2.2 Analisi dei dati: stima dei parametri
  - 4.2.3 La 'sampling distribution'
  - 4.2.4 L'errore standard
- 4.3 Riepilogo 1: Caratterizzare l'incertezza di un esperimento
- 4.4 L'intervallo di confidenza
- 4.5 Qual è il senso dell'intervallo di confidenza?
- 4.6 Come presentare i risultati degli esperimenti
- 4.7 Alcune precisazioni
  - 4.7.1 Campioni numerosi e non
  - 4.7.2 Popolazioni gaussiane e non
- 4.8 Analisi statistica dei dati: riassunto del percorso logico
- 4.9 Da ricordare
- 4.10 Esercizi

# Capitolo 5

## Decisioni e incertezza

Placeholder

## 5.1 Confronto tra una media osservata e una media teorica

### 5.1.1 Simulazione Monte Carlo

### 5.1.2 Soluzione formale

### 5.1.3 Interpretazione del P-level

## 5.2 Confronto tra due medie: il test t di Student

## 5.3 Confronto tra due proporzioni: il test $\chi^2$

## 5.4 Conclusioni

## 5.5 Riepilogo

## 5.6 Esercizi

## 5.7 Per approfondire un po'...

### 5.7.1 Tipologie alternative di test t di Student

### 5.7.2 Simulazione del P-value nel test d'ipotesi

### 5.7.3 Altre letture

# Capitolo 6

## Modelli ANOVA ad una via

Placeholder





- 6.1 Caso-studio: confronto tra erbicidi in vaso
- 6.2 Descrizione del dataset
- 6.3 Definizione di un modello lineare
- 6.4 Parametrizzazione del modello
  - 6.4.1 Vincolo sul trattamento
  - 6.4.2 Vincolo sulla somma
- 6.5 Assunzioni di base
- 6.6 Stima dei parametri
  - 6.6.1 Coefficienti del modello
  - 6.6.2 Residui
  - 6.6.3 Stima di  $\sigma$
  - 6.6.4 SEM e SED
- 6.7 Scomposizione della varianza
- 6.8 Test d'ipotesi
- 6.9 Operazioni con R
- 6.10 Medie marginali attese
- 6.11 Per concludere ...
- 6.12 Per approfondire un po'...

- 6.12.1 Simulazione della *sampling distribution* per F

## Capitolo 7

# La verifica delle assunzioni di base

Placeholder

## 7.1 Procedure diagnostiche

## 7.2 Analisi grafica dei residui

### 7.2.1 Grafico dei residui contro i valori attesi

### 7.2.2 QQ-plot

## 7.3 Strumenti diagnostici formali

## 7.4 Risultati contraddittori

## 7.5 ‘Terapia’

### 7.5.1 Correzione/Rimozione degli outliers

### 7.5.2 Correzione del modello

### 7.5.3 Trasformazione della variabile indipendente

### 7.5.4 Impiego di metodiche statistiche avanzate

### 7.5.5 Trasformazioni stabilizzanti

## 7.6 Esempio

## 7.7 Per approfondire un po’ ...

# Capitolo 8

## Contrasti e confronti multipli

Placeholder



## 8.1 Esempio

## 8.2 I contrasti

### 8.2.1 Varianza del contrasto e test d'ipotesi

### 8.2.2 I contrasti con R

## 8.3 I confronti multipli a coppie (pairwise comparisons)

## 8.4 Display a lettere

## 8.5 Problemi di molteplicità: tassi di errore per confronto e per esperimento

### 8.5.1 Correzione per la molteplicità

## 8.6 E le classiche procedure di confronto multiplo?

## 8.7 Consigli pratici

## 8.8 Per approfondire un po'...

### 8.8.1 Intervallo di confidenza di un contrasto

### 8.8.2 Correzione per la molteplicità: qualche dettaglio ulteriore

### 8.8.3 Altre letture

## Capitolo 9

# Modelli ANOVA con fattori di blocco

Placeholder



## **9.1 Caso-studio: confronto tra erbicidi in campo**

## **9.2 Definizione di un modello lineare**

## **9.3 Stima dei parametri**

### **9.3.1 Coefficienti del modello**

### **9.3.2 Residui e devianze**

## **9.4 Scomposizione della varianza**

## **9.5 Disegni a quadrato latino**

## **9.6 Caso studio: confronto tra metodi costruttivi**

## **9.7 Definizione di un modello lineare**

# Capitolo 10

## Modelli ANOVA a due vie

Placeholder

- 10.1 Il concetto di 'interazione'
- 10.2 Tipi di interazione
- 10.3 Caso-studio: interazione tra lavorazioni e diserbo chimico
- 10.4 Definizione del modello lineare
- 10.5 Stima dei parametri
- 10.6 Verifica delle assunzioni di base
- 10.7 Scomposizione delle varianze
- 10.8 Medie marginali attese
- 10.9 Calcolo degli errori standard (SEM e SED)
- 10.10 Medie marginali attese e confronti multipli con R
- 10.11 Per approfondire un po'....
  - 10.11.1 Anova a due vie: scomposizione 'manuale' della varianza

# Capitolo 11

## La regressione lineare semplice

Placeholder

## 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 Per approfondire un po'...

# Capitolo 12

## La regressione non-lineare

Placeholder



## 12.1 Caso studio: degradazione di un erbicida nel terreno

## 12.2 Scelta della funzione

## 12.3 Stima dei parametri

### 12.3.1 Linearizzazione della funzione

### 12.3.2 Approssimazione della vera funzione tramite una polinomiale in $X$

### 12.3.3 Minimi quadrati non-lineari

## 12.4 La regressione non-lineare con R

## 12.5 Verifica della bontà del modello

### 12.5.1 Analisi grafica dei residui

### 12.5.2 Test F per la mancanza di adattamento (approssimato)

### 12.5.3 Errori standard dei parametri

### 12.5.4 Coefficienti di determinazione

## 12.6 Funzioni lineari e nonlineari dei parametri

## 12.7 Previsioni

## 12.8 Gestione delle situazioni ‘patologiche’

### 12.8.1 Trasformazione del modello

### 12.8.2 Trasformazione dei dati

## 12.9 Per approfondire un po’...



# Capitolo 13

## Appendice 1: breve introduzione ad R

Placeholder

## 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 manuale di dati

Immissione di numeri progressivi

Immissione dei codici delle tesi e dei blocchi

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

## Capitolo 14

### Appendice 2: Richiami di statistica descrittiva

Placeholder

## Dati quantitativi: analisi chimiche e altre misurazioni fondamentali

Indicatori di tendenza centrale

Indicatori di dispersione

Arrotondamenti

Descrizione dei sottogruppi

14.0.1 Relazioni tra variabili quantitative: correlazione

## Dati qualitativi: conteggi e frequenze

Distribuzioni di frequenze e classamento

Statistiche descrittive per le distribuzioni di frequenze

Distribuzioni di frequenza bivariate: le tabelle di contingenza

Connessione

## Esercizi

Esercizio 1

Esercizio 2

14.1 Per approfondire un po'...