

Metodologia sperimentale per le scienze agrarie

Andrea Onofri e Dario Sacco

Update: v. 1.0 (15/03/2021), compil. 2022-03-03

Indice

Premessa	9
Obiettivi	9
Organizzazione	9
Software statistico	9
The authors	9
Ringraziamenti	9
1 Scienza e pseudo-scienza	10
1.1 Scienza = dati	11
1.2 Dati ‘buoni’ e ‘cattivi’	11
1.3 Dati ‘buoni’ e metodi ‘buoni’	13
1.4 Il principio di falsificazione	15
1.5 Falsificare un risultato	16
1.6 Elementi fondamentali del disegno sperimentale	17
1.6.1 Controllo degli errori	18
1.6.2 Replicazione	19
1.6.3 Randomizzazione	20
1.6.4 Esperimenti invalidi	21
1.7 Chi valuta se un esperimento è attendibile?	26
1.8 Conclusioni	27
1.9 Altre letture	28
2 Progettare un esperimento	29
2.1 Gli elementi della ricerca	31
2.2 Ipotesi scientifica → obiettivo dell’esperimento	31
2.3 Identificazione dei fattori sperimentali	31
2.3.1 Esperimenti (multi-)fattoriali	31
2.3.2 Controllo o testimone	31
2.4 Le unità sperimentali	31
2.5 Allocazione dei trattamenti	31

2.6	Le variabili sperimentali	31
2.6.1	Variabili nominali (categoriche)	31
2.6.2	Variabili ordinali	31
2.6.3	Variabili quantitative discrete	31
2.6.4	Variabili quantitative continue	31
2.6.5	Rilievi visivi e sensoriali	31
2.6.6	Variabili di confondimento	31
2.7	Esperimenti di campo	31
2.7.1	Scegliere il campo	31
2.7.2	Le unità sperimentali in campo	31
2.7.3	Numero di repliche	31
2.7.4	La mappa di campo	31
2.7.5	Lay-out sperimentale	31
2.8	Altre letture	31
3	Richiami di statistica descrittiva	32
3.1	Dati quantitativi	33
3.1.1	Indicatori di tendenza centrale	33
3.1.2	Indicatori di dispersione	33
3.1.3	Incertezza delle misure derivate	33
3.1.4	Relazioni tra variabili quantitative: correlazione	33
3.2	Dati qualitativi	33
3.2.1	Distribuzioni di frequenze e classamento	33
3.2.2	Statistiche descrittive per le distribuzioni di frequenze	33
3.2.3	Distribuzioni di frequenza bivariate: le tabelle di con- tingenze	33
3.2.4	Connessione	33
3.3	Statistiche descrittive con R	33
3.3.1	Descrizione dei sottogruppi	33
3.3.2	Distribuzioni di frequenze e classamento	33
3.3.3	Connessione	33
3.4	Altre letture	33
4	Modelli statistici ed analisi dei dati	34
4.1	Verità ‘vera’ e modelli deterministici	35
4.2	Genesi deterministica delle osservazioni sperimentali	35
4.3	Errore sperimentale e modelli stocastici	35
4.3.1	Funzioni di probabilità	35
4.3.2	Funzioni di densità	35
4.3.3	La distribuzione normale (curva di Gauss)	35
4.4	Modelli ‘a due facce’	35

4.5	E allora?	35
4.6	Le simulazioni Monte Carlo	35
4.7	Analisi dei dati e ‘model fitting’	35
4.8	Modelli stocastici non-normali	35
4.9	Altre letture	35
5	Stime ed incertezza	36
5.1	Esempio: una soluzione erbicida	38
5.1.1	Analisi dei dati: stima dei parametri	38
5.1.2	La ‘sampling distribution’	38
5.1.3	L’errore standard	38
5.2	Stima per intervallo	38
5.3	L’intervallo di confidenza	38
5.4	Qual è il senso dell’intervallo di confidenza?	38
5.5	Come presentare i risultati degli esperimenti	38
5.6	Alcune precisazioni	38
5.6.1	Campioni numerosi e non	38
5.6.2	Popolazioni gaussiane e non	38
5.7	Analisi statistica dei dati: riassunto del percorso logico	38
5.8	Da ricordare	38
5.9	Per approfondire un po’...	38
5.10	Coverage degli intervalli di confidenza	38
5.10.1	Intervalli di confidenza per fenomeni non-normali	38
5.11	Altre letture	38
6	Decisioni ed incertezza	39
6.1	Confronto tra due medie: il test t di Student	40
6.1.1	L’ipotesi nulla e alternativa	40
6.1.2	La statistica T	40
6.1.3	Simulazione Monte Carlo	40
6.1.4	Soluzione formale	40
6.1.5	Interpretazione del P-level	40
6.1.6	Tipologie alternative di test t di Student	40
6.2	Confronto tra due proporzioni: il test χ^2	40
6.2.1	Simulazione Monte Carlo	40
6.2.2	Soluzione formale	40
6.3	Conclusioni e riepilogo	40
6.4	Altre letture	40
7	Modelli ANOVA ad una via	41
7.1	Caso-studio: confronto tra erbicidi in vaso	43

7.2	Descrizione del dataset	43
7.3	Definizione di un modello lineare	43
7.4	Parametrizzazione del modello	43
7.5	Assunzioni di base	43
7.6	Fitting del modello: metodo manuale	43
7.6.1	Stima dei parametri	43
7.6.2	Calcolo dei residui	43
7.6.3	Stima di σ	43
7.7	Scomposizione della varianza	43
7.8	Test d'ipotesi	43
7.9	Inferenza statistica	43
7.10	Fitting del modello con R	43
7.11	Medie marginali attese	43
7.12	Per concludere	43
7.13	Altre letture	43
8	La verifica delle assunzioni di base	44
8.1	Violazioni delle assunzioni di base	45
8.2	Procedure diagnostiche	45
8.3	Analisi grafica dei residui	45
8.3.1	Grafico dei residui contro i valori attesi	45
8.3.2	QQ-plot	45
8.4	Test d'ipotesi	45
8.5	Risultati contraddittori	45
8.6	'Terapia'	45
8.6.1	Correzione/Rimozione degli outliers	45
8.6.2	Correzione del modello	45
8.6.3	Trasformazione della variabile indipendente	45
8.6.4	Impiego di metodiche statistiche avanzate	45
8.6.5	Trasformazioni stabilizzanti	45
8.7	Esempio 1	45
8.8	Esempio 2	45
8.9	Altre letture	45
9	Contrasti e confronti multipli	46
9.1	Esempio	47
9.2	I contrasti	47
9.3	I contrasti con R	47
9.4	I confronti multipli a coppie (pairwise comparisons)	47
9.5	Display a lettere	47
9.6	Tassi di errore per confronto e per esperimento	47

9.7	Aggiustamento per la molteplicità	47
9.8	E le classiche procedure di confronto multiplo?	47
9.9	Consigli pratici	47
9.10	Altre letture	47
10	Modelli ANOVA con fattori di blocco	48
10.1	Caso-studio: confronto tra erbicidi in campo	49
10.2	Definizione di un modello lineare	49
10.3	Stima dei parametri	49
10.3.1	Coefficienti del modello	49
10.3.2	Stima di σ	49
10.4	Scomposizione della varianza	49
10.5	Adattamento del modello con R	49
10.6	Disegni a quadrato latino	49
10.7	Caso studio: confronto tra metodi costruttivi	49
10.8	Definizione di un modello lineare	49
11	La regressione lineare semplice	50
11.1	Caso studio: effetto della concimazione azotata al frumento . .	51
11.2	Analisi preliminari	51
11.3	Definizione del modello lineare	51
11.4	Stima dei parametri	51
11.5	Valutazione della bontà del modello	51
11.5.1	Valutazione grafica	51
11.5.2	Errori standard dei parametri	51
11.5.3	Test F per la mancanza d'adattamento	51
11.5.4	Test F per la bontà di adattamento e coefficiente di determinazione	51
11.6	Previsioni	51
11.7	Altre letture	51
12	Modelli ANOVA a due vie con interazione	52
12.1	Il concetto di 'interazione'	53
12.2	Effetti incrociati: interazione tra lavorazioni e diserbo chimico	53
12.3	Definizione del modello lineare	53
12.4	Calcoli manuali	53
12.4.1	Scomposizione della varianza	53
12.5	Calcoli con R	53
12.5.1	Model fitting	53
12.5.2	Verifica delle assunzioni di base	53
12.5.3	Scomposizione della varianza	53

12.5.4	Medie marginali attese e confronti multipli con R . . .	53
12.6	Effetti innestati: valutazione di ibridi di mais	53
12.7	Definizione del modello lineare	53
12.8	Fitting del modello con R	53
13	Breve introduzione ai modelli misti	54
13.1	Raggruppamenti tra parcelle	54
13.2	Esperimenti a split-plot	54
13.2.1	Definizione del modello lineare	54
13.2.2	Model fitting con R	54
13.3	Esperimenti a strip-plot	54
13.3.1	Definizione del modello lineare	54
13.3.2	Model fitting con R	54
13.4	Altre letture	54
14	La regressione non-lineare	55
14.1	Caso studio: degradazione di un erbicida nel terreno	57
14.2	Scelta della funzione	57
14.3	Stima dei parametri	57
14.3.1	Linearizzazione della funzione	57
14.3.2	Approssimazione della vera funzione tramite una poli- nomiale in X	57
14.3.3	Minimi quadrati non-lineari	57
14.4	La regressione non-lineare con R	57
14.5	Verifica della bontà del modello	57
14.5.1	Analisi grafica dei residui	57
14.5.2	Test F per la mancanza di adattamento (approssimato)	57
14.5.3	Errori standard dei parametri	57
14.5.4	Coefficienti di determinazione	57
14.6	Funzioni lineari e nonlineari dei parametri	57
14.7	Previsioni	57
14.8	Gestione delle situazioni ‘patologiche’	57
14.8.1	Trasformazione del modello	57
14.8.2	Trasformazione dei dati	57
14.9	Per approfondire un po’...	57
14.9.1	Riparametrizzazione delle funzioni non-lineari	57
14.9.2	Altre letture	57
15	Esercizi	58
15.1	Capitoli 1 e 2	60
15.1.1	Esercizio 1	60

15.1.2	Esercizio 2	60
15.2	Capitolo 3	60
15.2.1	Esercizio 1	60
15.2.2	Esercizio 2	60
15.2.3	Esercizio 3	60
15.3	Capitolo 4	60
15.3.1	Esercizio 1	60
15.3.2	Esercizio 2	60
15.3.3	Esercizio 3	60
15.3.4	Esercizio 4	60
15.3.5	Esercizio 5	60
15.3.6	Esercizio 6	60
15.3.7	Esercizio 7	60
15.3.8	Esercizio 8	60
15.4	Capitolo 5	60
15.4.1	Esercizio 1	60
15.4.2	Esercizio 2	60
15.4.3	Esercizio 3	60
15.4.4	Esercizio 4	60
15.4.5	Esercizio 5	60
15.5	Capitolo 6	60
15.5.1	Esercizio 1	60
15.5.2	Esercizio 2	60
15.5.3	Esercizio 3	60
15.5.4	Esercizio 4	60
15.5.5	Esercizio 5	60
15.5.6	Esercizio 6	60
15.5.7	Esercizio 7	60
15.5.8	Esercizio 8	60
15.5.9	Esercizio 9	60
15.5.10	Esercizio 10	60
15.6	Capitoli da 7 a 9	60
15.6.1	Esercizio 1	60
15.6.2	Esercizio 2	60
15.6.3	Esercizio 3	60
15.6.4	Esercizio 4	60
15.7	Capitolo 10	60
15.7.1	Esercizio 1	60
15.7.2	Esercizio 2	60
15.7.3	Esercizio 3	60
15.8	Capitolo 11	60

15.8.1	Esercizio 1	60
15.8.2	Esercizio 2	60
15.9	Capitoli 12 e 13	60
15.9.1	Esercizio 1	60
15.9.2	Esercizio 2	60
15.9.3	Esercizio 3	60
15.9.4	Esercizio 4	60
15.9.5	Esercizio 5	60
15.9.6	Esercizio 6	60
15.10	Capitolo 14	60
15.10.1	Esercizio 1	60
15.10.2	Esercizio 2	60
15.10.3	Esercizio 3	60
15.10.4	Esercizio 4	60
15.10.5	Esercizio 5	60
15.10.6	Esercizio 6	60
15.10.7	Esercizio 7	60
16	Appendice 1: breve introduzione ad R	61
	Cosa è R?	62
	Oggetti e assegnazioni	62
	Costanti e vettori	62
	Matrici	62
	Dataframe	62
	Quale oggetto sto utilizzando?	62
	Operazioni ed operatori	62
	Funzioni ed argomenti	62
	Consigli per l'immissione di dati sperimentali	62
	Immissione di numeri progressivi	62
	Immissione dei codici delle tesi e dei blocchi	62
	Immissione dei valori e creazione del dataframe	62
	Leggere e salvare dati esterni	62
	Alcune operazioni comuni sul dataset	62
	Selezionare un subset di dati	62
	Ordinare un vettore o un dataframe	62
	Workspace	62
	Script o programmi	62
	Interrogazione di oggetti	62
	Altre funzioni matriciali	62
	Cenni sulle funzionalità grafiche in R	62
	Altre letture	62

Premessa

Placeholder

Obiettivi

Organizzazione

Software statistico

The authors

Ringraziamenti

Capitolo 1

Scienza e pseudo-scienza

In una società caratterizzata dal sovraccarico cognitivo, è chiedersi (e chiedere) che cosa sia la scienza, cioè cosa distingue le informazioni scientifiche da tutto quello che invece non è altro che pura opinione, magari autorevole, ma senza il sigillo dell'oggettività.

Per quanto affascinante possa sembrare l'idea del ricercatore che con un'improvviso colpo di genio elabora una stupefacente teoria, dovrebbe essere chiaro che l'intuizione, per quanto geniale ed innovativa, è solo un possibile punto di partenza, che non necessariamente prelude al progresso scientifico. In generale, almeno in ambito biologico, nessuna teoria acquisisce automaticamente valenza scientifica, ma rimane solo nell'ambito delle opinioni, indipendentemente dal fatto che nasca da un colpo di genio, oppure da un paziente e meticoloso lavoro di analisi intellettuale, che magari si concretizza in un modello matematico altamente elegante e complesso.

Che cosa è che permette ad una prova scientifica di uscire dall'ambito delle opinioni legate a divergenze di cultura, percezione e/o credenze individuali, per divenire, al contrario, oggettiva e universalmente valida? Che cosa è che distingue la verità scientifica da altre verità di natura metafisica, religiosa o pseudoscientifica?

A questo proposito, è utile leggere questi aforismi interessanti e significativi:

1. Analogy cannot serve as proof (Pasteur)
2. The interest I have in believing a thing is not a proof of the existence of that thing (Voltaire)
3. A witty saying proves nothing (Voltaire)

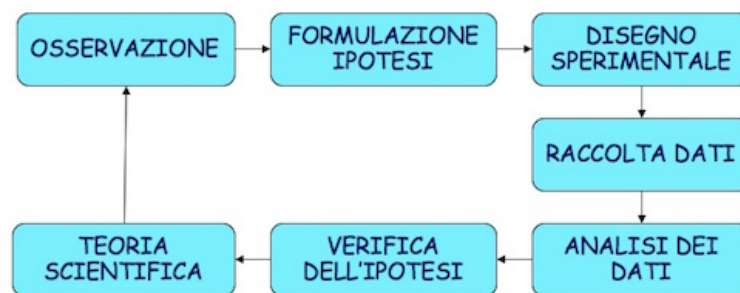


Figura 1.1: Il metodo scientifico Galileiano

1.1 Scienza = dati

La base di tutta la scienza risiede nel cosiddetto ‘metodo scientifico,’ che si fa comunemente risalire a Galileo Galilei (1564-1642) e che è riassunto nella figura seguente.

Senza andare troppo in profondità, è importante notare due aspetti:

1. il ruolo fondamentale dell’esperimento scientifico, che produce dati a supporto di ipotesi pre-esistenti;
2. lo sviluppo di teorie basate sui dati, che rimangono valide fino a che non si raccolgono altri dati che le confutano, facendo nascere nuove ipotesi che possono portare allo sviluppo di nuove teorie, più affidabili o più semplici.

Insomma, l’ingrediente fondamentale di una prova scientifica è quello di essere supportata dai dati sperimentali: di fatto, non esiste scienza senza dati! Resta famoso l’aforisma “In God we trust, all the others bring data,” attribuito all’ingegnere e statistico americano W. Edwards Deming (1900-1993), anche se pare che egli, in realtà, non l’abbia mai pronunciato.

1.2 Dati ‘buoni’ e ‘cattivi’

Detto che la scienza si basa sui dati, bisogna anche dire che non tutti i dati sono ugualmente ‘buoni.’ Nelle scienze biologiche, così come nelle altre scienze, è importante che i dati siano in grado di cogliere gli effetti che vogliamo studiare, senza introdurre distorsioni.

In agricoltura e nelle altre scienze quantitative abbiamo a che fare con fenomeni ‘misurabili’ e, di conseguenza, i nostri dati consistono di un set di misure di diverso tipo (ci torneremo nel secondo capitolo). L’aspetto più importante è che, per tutta una serie di motivi che dettaglieremo tra poco, le nostre misure non necessariamente riflettono il valore vero della caratteristica misurata nel nostro soggetto. Ciò è noto con il termine di **errore sperimentale**, che non significa che abbiamo necessariamente fatto qualcosa di sbagliato. Anzi, l’errore sperimentale è considerato una componente inevitabile di ogni esperimento, in grado di proiettare un alea d’incertezza su ogni risultato scientifico.

Ci sono tre fondamentali fonti di errore sperimentale:

1. Errore di misura
2. Variabilità dei soggetti sperimentali
3. Campionamento

Gli errori di misura sono legati allo strumento e dipendono dall’errata taratura, dall’impiego di un protocollo sbagliato, da inesattezze strumentali, da errori nella trascrizione dei risultati oppure dall’irregolarità dell’oggetto da misurare. Ad esempio, pensate alla misurazione dell’altezza di una pianta di mais: è facile riscontrare difficoltà legate, ad esempio, all’individuazione del punto esatto in cui inizia il culmo e del punto esatto dove termina l’infiorescenza apicale.

A parte gli errori di misura, ci sono anche altre sorgenti di errore meno evidenti e legate al fatto che, nel lavoro sperimentale, siamo di solito interessati non ad un singolo soggetto, ma ad un gruppo più o meno numeroso. Ad esempio, se dobbiamo misurare l’effetto di un erbicida, non possiamo farlo trattando una sola pianta, ma dobbiamo ripetere le misure su un gruppo di piante, il che ci porta ad avere un gruppo di misure, una diversa dall’altra. Quindi, qual è l’effetto dell’erbicida? Il fatto di avere tanti effetti diversi quante sono le piante studiate crea comunque un certo grado di variabilità che non dipende da alcun errore tecnico, ma è una caratteristica intrinseca del fenomeno biologico in studio.

Di per se’, la variabilità naturale dei soggetti sperimentali non sarebbe un grosso problema, in quanto potremmo calcolarci l’effetto medio e ritenerci soddisfatti in relazione alle finalità dell’esperimento. Tuttavia sorge un nuovo problema legato al fatto che spesso i soggetti sono così numerosi che non possiamo misurarli tutti e siamo costretti a misurare un campione composto da un ridotto numero di individui. Abbiamo un nuovo elemento di incertezza: come facciamo ad essere sicuri che la media, o qualunque altra statistica,

misurata nel nostro campione rifletta la media dell'intera popolazione? Anche se abbiamo fatto del tutto per scegliere un campione rappresentativo, è evidente che il campione perfetto non esiste: cosa potrebbe succedere se prendessimo un altro campione?

1.3 Dati 'buoni' e metodi 'buoni'

Quindi la ricerca scientifica non è esente da 'errori' in senso lato (componenti di incertezza). Tuttavia, gli errori non sono tutti uguali e si dividono in **sistematici** ed **accidentali (casuali)**. L'errore sistematico è provocato da difetti intrinseci dello strumento o incapacità peculiari dell'operatore e tende a ripetersi costantemente e con lo stesso segno in misure successive. Un esempio tipico è quello di una bilancia non tarata, che tende ad aggiungere 20 grammi ad ogni misura che effettuiamo. D'altra parte, l'errore accidentale, essendo di natura casuale, tende a ripresentarsi con valori e segni diversi. Di conseguenza, è ragionevole pensare che le repliche, nel lungo periodo, producano sovrastime e sottostime con uguale probabilità, in modo che la media tende a coincidere con il valore vero.

È facile capire che le conseguenze degli errori sistematici e accidentali sono ben diverse. A questi propositi, dobbiamo considerare due aspetti molto importanti, cioè:

1. precisione
2. accuratezza

Con il termine **precisione** intendiamo due cose: la prima è relativa al numero di decimali che ci fornisce il nostro strumento di misura. E' evidente, ad esempio, come un calibro sia più preciso di un metro da sarto. Oltre a questo significato, abbastanza intuitivo, ce n'è un altro, più specificatamente legato agli esperimenti scientifici: la precisione di un dato ottenuto attraverso un processo di misurazione non è altro che la variabilità riscontrata quando la misurazione viene ripetuta più volte. L'errore casuale produce sempre un calo di precisione.

Il termine accuratezza ha invece un significato completamente diverso, riconducibile alla differenza tra la misura effettuata e il valore vero della caratteristica da misurare. Può sembrare una banalità, ma proviamo a pensare ad uno strumento non tarato, come, ad esempio, un gascromatografo, che restituisce sempre una concentrazione maggiorata del 20%. Se noi ripetessimo le analisi 100 volte, in assenza di altri errori, otterremmo sempre lo stesso

risultato, molto preciso, ma totalmente inaffidabile, nel senso che non riflette la concentrazione reale della soluzione in studio. L'errore sistematico, oltre a produrre un calo di precisione, produce anche inaccuratezza.

Comprendiamo bene che l'accuratezza è più importante della precisione: infatti una misura accurata, ma imprecisa, riflette bene la realtà, anche se in modo vago. Al contrario, una misura precisa, ma inaccurata, ci porta completamente fuori strada, perché non riflette la realtà. Con linguaggio tecnico, un dato non accurato si dice 'distorto' (*biased*) e, siccome la distorsione dipende dagli errori sistematici, questi ultimi vanno assolutamente evitati, ad esempio con la perfetta taratura degli strumenti e l'adozione di metodi di misura rigidamente standardizzati e accettati dalla comunità scientifica mondiale.

L'inaccuratezza preoccupa molto i laboratori di analisi, che spesso utilizzano standard di confronto, la cui misura è perfettamente nota e viene periodicamente confrontata con quella rilevabile dallo strumento stesso, per verificarne la taratura. Altro metodo utilizzato nelle procedure di accreditamento dei laboratori è il *ring test*, dove campioni reali della matrice da misurare sono inviati a più laboratori a livello nazionale, in modo da poter confrontare le misure ottenute e valutarne la variabilità. Con un *ring test*, un laboratorio può valutare la sua stessa affidabilità in confronto con laboratori simili, basandosi sull'eventuale differenza tra il risultato ottenuto e quelli ottenuti in tutti gli altri laboratori valutati.

Sfortunatamente la possibilità di raccogliere dati inaccurati è tutt'altro che remota. Gli scienziati americani Pons e Fleischmann, il 23 Marzo del 1989, diffusero pubblicamente la notizia di essere riusciti a riprodurre la fusione nucleare fredda, causando elevatissimo interesse nella comunità scientifica. Purtroppo le loro misure erano viziate da una serie di problemi e il loro risultato fu smentito da esperimenti successivi.

A parte questo clamoroso esempio, torniamo alla nostra domanda iniziale: come facciamo ad essere sicuri che i dati siano validi ed affidabili? La risposta è semplice: non possiamo mai essere sicuri, ma **dobbiamo fare del nostro meglio per applicare metodi rigorosi, così da minimizzare la possibilità di ottenere errori sistematici**. In altre parole, dati 'buoni' sono conseguenza di metodi 'buoni' e, pertanto **una prova scientifica è tale non perché siamo certi che corrisponda alla realtà, ma perché siamo ragionevolmente certi che sia stata ottenuta con metodi validi!**



Figura 1.2: Conseguenze di un esperimento sbagliato

1.4 Il principio di falsificazione

L'approccio che abbiamo indicato poco sopra ha un'importante conseguenza: anche se abbiamo utilizzato un metodo perfettamente valido non potremo mai avere la certezza di aver ottenuto un risultato corrispondente alla realtà e, quindi, ci dovremo sempre aspettare che ulteriori dati smentiscano la nostra conclusione. Questa è la base del principio di falsificazione, definito da Karl Popper (1902-1994): non potremo mai dimostrare che una nostra ipotesi è vera, ma potremo solo dimostrare che è falsa.

In pratica, tornando al metodo scientifico, partiamo da un'ipotesi e organizziamo un esperimento perfettamente valido che produce, di conseguenza, dati validi. Se i nostri dati sconfirmano l'ipotesi, abbiamo dimostrato che questa è falsa e dovremo quindi produrre una nuova ipotesi da sottoporre a verifica. Se invece i dati confermano la nostra ipotesi (o meglio, non la smentiscono) allora, non potremo concludere che l'ipotesi è vera, in quanto rimarrà sempre il dubbio che non abbiamo raccolto abbastanza dati. Tuttavia, in mancanza di altre informazioni, prenderemo per buona la nostra ipotesi, fino a che non sarà smentita. Una sorta di 'assoluzione' per insufficienza di prove, quindi. . .

Il principio di falsificazione è piuttosto importante nel mondo scientifico ed ha alcune importanti implicazioni:

1. Scienza non necessariamente significa ‘certezza’ o ‘verità.’ Tutto quello che possiamo fare con certezza è rigettare ipotesi (provare che sono false), ma non dimostrarne la validità.
2. Il nostro compito è quello di cercare di eliminare tutte le fonti di errore sistematico, per rendere il risultato il più accurato possibile.
3. Eliminato l’errore sistematico, l’eventuale errore casuale residuo deve essere sempre quantificato e visualizzato insieme ai risultati.
4. In considerazione dell’errore residuo, dobbiamo decidere se i dati raccolti consentono di rigettare la nostra ipotesi di partenza. Altrimenti, l’esperimento è inconclusivo e, pur non avendone la certezza, terremo per vera la nostra ipotesi di partenza fino a che non sarà smentita da future osservazioni.

Oltre al principio di falsificazione, la scienza fa largo uso del principio del ‘rasoio di Occam.’ Guglielmo di Occam (XIV secolo) era un frate francescano che, in un periodo in cui le dimostrazioni scientifiche iniziavano a divenire troppo complesse, voleva ribadire l’importanza della semplicità. Il suo principio è solitamente formulato come ‘*Entia non sunt multiplicanda praeter necessitatem*’ ed è noto come il ‘rasoio’ in quanto porta a respingere con nettezza (tagliare con il rasoio) le spiegazioni troppo complesse. Nella comunità scientifica, applichiamo questo principio preferendo sempre, tra due ipotesi alternative ugualmente buone, quella più semplice.

1.5 Falsificare un risultato

Se un esperimento è inconclusivo e porta ad eccettare un’ipotesi, è sempre possibile eseguire un ulteriore esperimento per rigettarla. Se anche questo secondo esperimento non riesce a rigettare l’ipotesi di partenza, allora la bontà di quest’ultima è certamente rafforzata. Parliamo quindi di esperimenti confermativi che costituiscono un elemento molto importante del metodo scientifico.

A questo proposito, distinguiamo:

1. replicabilità
2. riproducibilità

Un esperimento è replicabile se, quando ripetuto in condizioni assolutamente analoghe (stessi soggetti, ambiente, strumenti...), restituisce risultati equivalenti. Per questo motivo, quando si pubblicano i risultati di un esperimento, è sempre necessario descrivere accuratamente i metodi impiegati, in modo da consentire a chiunque la verifica dei risultati.

In alcuni casi, tuttavia, questa verifica indipendente è pressoché impossibile; ad esempio, nelle scienze agronomiche, le caratteristiche genetiche e pedoclimatiche giocano un ruolo molto importante e non è facile replicare un esperimento di pieno campo esattamente nelle stesse condizioni. Per questo motivo, alcuni biostatistici distinguono la replicabilità dalla riproducibilità, definita come il grado di concordanza tra esperimenti ripetuti in condizioni diverse (diversi soggetti, diverso ambiente...). Se la replicabilità di un esperimento non può essere dimostrata, bisogna avere almeno un'idea della sua riproducibilità, ripetendo l'esperimento in condizioni diverse e discutendo attentamente le eventuali differenze riscontrate nei risultati.

1.6 Elementi fondamentali del disegno sperimentale

La metodica di organizzazione di un esperimento valido prende il nome di *disegno sperimentale* e le sue basi si fanno in genere risalire a Sir Ronald A. Fisher, vissuto in Inghilterra dal 7 Febbraio 1890 al 29 luglio 1962. Laureatosi nel 1912, lavora come statistico per il comune di Londra, fino a quando diviene socio della prestigiosa Eugenics Education Society di Cambridge, fondata nel 1909 da Francis Galton, cugino di Charles Darwin. Dopo la fine della guerra, Karl Pearson gli propone un lavoro presso il rinomato Galton Laboratory, ma egli non accetta a causa della profonda rivalità esistente tra lui e Pearson stesso. Nel 1919 viene assunto presso la Rothamsted Experimental Station, dove si occupa dell'elaborazione dei dati sperimentali e, nel corso dei successivi 7 anni, definisce le basi del disegno sperimentale ed elabora la sua teoria della "analysis of variance." Il suo libro più importante è "The design of experiment," del 1935. E' sua la definizione delle tre componenti fondamentali del disegno sperimentale:

1. controllo degli errori;
2. replicazione;
3. randomizzazione.

1.6.1 Controllo degli errori

Controllare gli errori, o, analogamente, eseguire un esperimento controllato significa fondamentalmente due cose:

1. adottare provvedimenti idonei ad evitare le fonti di errore, mantenendole al livello più basso possibile (alta precisione);
2. agire in modo da isolare l'effetto in studio (accuratezza), evitando che si confonda con effetti casuali e di altra natura. Ad esempio, se dobbiamo confrontare due fitofarmaci, dobbiamo fare in modo che i soggetti inclusi nell'esperimento differiscano tra di loro solo per il fitofarmaco impiegato e non per altro.

Mettere in pratica questi principi fondamentali richiede una vita di esperienza! Tuttavia, vogliamo solo sottolineare alcuni aspetti, come il rigore metodologico. È evidente che, ad esempio, se vogliamo sapere la cinetica di degradazione di un erbicida a 20 °C dovremo realizzare una prova esattamente a quella temperatura, con un erbicida uniformemente distribuito nel terreno, dentro una camera climatica capace di un controllo perfetto della temperatura. Gli strumenti dovranno essere ben tarati e sarà necessario attenersi scrupolosamente a metodi validati e largamente condivisi. Tuttavia, a proposito di rigore, non bisogna scordare quanto diceva C.F. Gauss a proposito della precisione nei calcoli, e che può essere anche riferito al rigore nella ricerca : *“Manca di mentalità matematica tanto chi non sa riconoscere rapidamente ciò che è evidente, quanto chi si attarda nei calcoli con una precisione superiore alla necessità”*

Oltre al rigore metodologico, è bene anche ricordare come un esperimento ben fatto passi sempre attraverso la giusta selezione dei soggetti sperimentali, che debbono essere omogenei, ma rappresentativi della popolazione alla quale intendiamo riferire i risultati ottenuti. Ad esempio, se si vuole ottenere un risultato riferito alla collina umbra, bisognerà scegliere parcelle di terreno omogenee, ma che rappresentano bene la variabilità pedo-climatica di quell'ambiente, né di più, né di meno.

Per concludere, vogliamo anche ricordare le cosiddette 'intrusioni' cioè quegli eventi che accadono in modo inaspettato e condizionano negativamente la riuscita di un esperimento in corso. E' evidente che, ad esempio, un'alluvione, l'attacco di insetti o patogeni, la carenza idrica hanno una pesante ricaduta sulla precisione di un esperimento e sulla sua riuscita. Per quanto possibile, controllare gli errori significa anche essere capaci di prevedere le eventuali intrusioni. In un suo famoso lavoro scientifico del 1984, lo scien-

ziato americano Stuart Hurlbert usa il termine ‘intrusione demoniaca’ per indicare quelle intrusioni che, pur casuali, avrebbero potuto essere previste con un disegno più accurato, sottolineando in questo caso la responsabilità dello sperimentatore.

Un esempio è questo: uno sperimentatore vuole studiare l’entità della predazione dovuta alle volpi e quindi usa campi senza staccionate (dove le volpi possono entrare) e campi protetti da staccionate (e quindi liberi da volpi). Se le staccionate, essendo utilizzate dai falchi come punto d’appoggio, finiscono per incrementare l’attività predatoria di questi ultimi, si viene a creare un’intrusione demoniaca, che rende l’esperimento distorto. Il demone, in questo caso, non è il falco, che danneggia l’esperimento, ma il ricercatore stesso, che non ha saputo prevedere una possibile intrusione.

1.6.2 Replicazione

In ogni esperimento, i trattamenti dovrebbe essere replicati su due o più unità sperimentali. Ciò permette di:

1. dimostrare che i risultati sono replicabili (ma non è detto che siano riproducibili!);
2. assicurare che eventuali circostanze aberranti casuali non abbiano provocato risultati distorti;
3. misurare la precisione dell’esperimento, come variabilità di risposta tra repliche trattate nello stesso modo;
4. incrementare la precisione dell’esperimento (più sono le repliche più l’esperimento è preciso, perché si migliora la stima della caratteristica misurata, diminuendo l’incertezza).

Per poter essere utili, le repliche debbono essere indipendenti, cioè debbono **aver subito tutte le manipolazioni necessarie per l’allocazione del trattamento in modo totalmente indipendente l’una dall’altra**. Le manipolazioni comprendono tutte le pratiche necessarie, come ad esempio la preparazione delle soluzioni, la diluizione dei prodotti, ecc..

La manipolazione indipendente è fondamentale, perché in ogni parte del processo di trattamento possono nascondersi errori più o meno grandi, che possono essere riconosciuti solo se colpiscono in modo casuale le unità sperimentali. Se la manipolazione è, anche solo in parte, comune, questi errori colpiscono tutte le repliche allo stesso modo, diventano sistematici e quindi non più riconoscibili. Di conseguenza, si inficia l’accuratezza dell’esperimento. Quando

le repliche non sono indipendenti, si parla di **pseudorepliche**, contrapposte alle **repliche vere**.

Il numero di repliche dipende dal tipo di esperimento: più sono e meglio è, anche se è necessario trovare un equilibrio accettabile tra precisione e costo dell'esperimento. Nella sperimentazione di campo, due repliche sono poche, tre appena sufficienti, quattro costituiscono la situazione più comune, mentre un numero maggiore di repliche è abbastanza raro, non solo per la difficoltà di seguire l'esperimento, ma anche perché aumentano la dimensione della prova e, di conseguenza, la variabilità del terreno.

1.6.3 Randomizzazione

L'indipendenza di manipolazione non garantisce da sola un esperimento corretto. Infatti potrebbe accadere che le caratteristiche innate dei soggetti, o una qualche 'intrusione' influenzino in modo sistematico tutte le unità sperimentali trattate nello stesso modo, così da confondersi con l'effetto del trattamento. Un esempio banale è che potremmo somministrare un farmaco a quattro soggetti in modo totalmente indipendente, ma se i quattro soggetti fossero sistematicamente più alti di quelli non trattati finiremmo per confondere una caratteristica innata con l'effetto del farmaco. Oppure, se le piante di una certa varietà di sorgo si trovassero tutte più vicine alla scolina rispetto a quelle di un'altra varietà, potrebbero essere più danneggiate dal ristagno idrico, il cui effetto si confonderebbe con quello del trattamento stesso.

Questi problemi sono particolarmente insidiosi e si nascondono anche dietro ai particolari apparentemente più insignificanti. La randomizzazione è l'unico sistema per evitare, o almeno rendere molto improbabile, la confusione dell'effetto del trattamento con fattori casuali e/o comunque diversi dal trattamento stesso. La randomizzazione si declina in vari modi:

1. allocazione casuale del trattamento alle unità sperimentali. Gli esperimenti che prevedono l'allocazione del trattamento sono detti 'manipolativi' o 'disegnati.'
2. A volte l'allocazione del trattamento non è possibile o non è etica. Se volessimo studiare l'effetto delle cinture di sicurezza nell'evitare infortuni gravi, non potremmo certamente provocare incidenti deliberati. In questo caso la randomizzazione è legata alla scelta casuale di soggetti che sono 'naturalmente' trattati. Esperimenti di questo tipo, si dicono **osservazionali**. Un esempio è la valutazione dell'effetto dell'inquinamento con metalli pesanti nella salute degli animali: ovviamente non

è possibile, se non su piccola scala, realizzare il livello di inquinamento desiderato e, pertanto, dovremo scegliere soggetti che sono naturalmente sottoposti a questo genere di inquinamento, magari perché vivono vicino a zone industriali.

3. Se i soggetti sono immobili, la randomizzazione ha anche una connotazione legata alla disposizione spaziale e/o temporale casuale.

L'assegnazione casuale del trattamento, o la selezione casuale dei soggetti trattati, fanno sì che tutti i soggetti abbiano la stessa probabilità di ricevere qualunque trattamento oppure qualunque intrusione casuale. In questo modo, la probabilità che tutte le repliche di un trattamento abbiano qualche caratteristica innata o qualche intrusione comune che li penalizzi/avvantaggi viene minimizzata. Di conseguenza, confondere l'effetto del trattamento con variabilità casuale ('confounding'), anche se teoricamente possibile, diviene altamente improbabile.

1.6.4 Esperimenti invalidi

A questo punto, dovrebbe essere chiaro che un esperimento valido deve essere controllato, replicato e randomizzato: la mancanza anche di uno solo di questi elementi pone dubbi ragionevoli sull'affidabilità dei risultati. In particolare, gli esperimenti 'invalidi' sono caratterizzati da:

1. Cattivo controllo degli errori
2. Fondati sospetti di confounding
3. Mancanza di repliche vere
4. Confusione tra repliche vere e pseudo-repliche
5. Mancanza di randomizzazione
6. Presenza di vincoli alla randomizzazione, trascurati in fase di analisi.

Le conseguenze di queste problematiche sono abbastanza diverse.

Cattivo controllo degli errori

Bisogna verificare se il problema è relativo a questioni come la mancanza di scrupolosità, l'uso di soggetti poco omogenei o di un ambiente poco omogeneo, o altri aspetti che inficiano solo la precisione, ma non l'accuratezza dell'esperimento. In questo caso, l'esperimento è ancora valido (accurato), ma la bassa precisione probabilmente impedirà di trarre conclusioni forti. Quindi, un esperimento impreciso si 'elimina' da solo, perché sarà inconclusivo.

Di questi esperimenti bisogna comunque diffidare, soprattutto quando siano pianificati per mostrare l'assenza di differenze tra due trattamenti alternativi. Mostrare l'assenza di differenze è facile: basta fare male un esperimento, in modo che vi sia un alto livello di incertezza e quindi l'evidenza scientifica sia molto debole.

Diversa è la situazione in cui un cattivo controllo degli errori, ad esempio l'adozione di metodi sbagliati, porta a mancanza di accuratezza, cioè a risultati che non riflettono la realtà (campionamento sbagliato, ad esempio; oppure strumenti non tarati; impiego di metodi non validati e/o non accettabili). In questo caso venendo a mancare l'accuratezza, l'esperimento deve essere rigettato, in quanto non fornisce informazioni realistiche.

‘Confounding’ e correlazione spuria

Abbiamo appena menzionato il problema fondamentale della ricerca, cioè il **confounding**, vale a dire la confusione tra l'effetto del trattamento e un qualche altro effetto casuale, legato alle caratteristiche innate del soggetto o a qualche intrusione più o meno ‘demoniaca.’ Abbiamo detto che non possiamo mai avere la certezza dell'assenza di confounding, ma abbiamo anche detto che l'adozione di una pratica sperimentale corretta ne minimizza la probabilità.

Chiaramente, rimangono dei rischi che sono tipici di situazioni nelle quali il controllo adottato non è perfetto, come capita, ad esempio, negli esperimenti osservazionali. In questo ambito è piuttosto temuta la cosiddetta ‘correlazione spuria,’ una forma di confounding casuale per cui due variabili variano congiuntamente (sono direttamente o inversamente proporzionali), ma in modo del tutto casuale. Esistono, ad esempio, dati che mostrano una chiara correlazione tra le vendite di panna acida e le morti per incidenti in motocicletta (Figura 1.3). Chiaramente, non esistono spiegazioni scientifiche per questo effetto, che è, ovviamente, del tutto casuale. Il problema è che questa correlazione spuria non è sempre così semplice da rintracciare.

A volte il confounding non è casuale, ma è legato ad una variabile esterna che si agisce all'insaputa dello sperimentatore. Ad esempio, è stato osservato che il tasso di crimini è più alto nelle città che hanno più chiese. La spiegazione di questo paradosso sta nel fatto che esiste un ‘confounder,’ cioè l'ampiezza della popolazione. Nelle grandi città si riscontrano sia una maggiore incidenza criminale, sia un grande numero di chiese. In sostanza, la popolazione determina sia l'elevato numero di chiese che l'elevato numero di crimini, ma queste

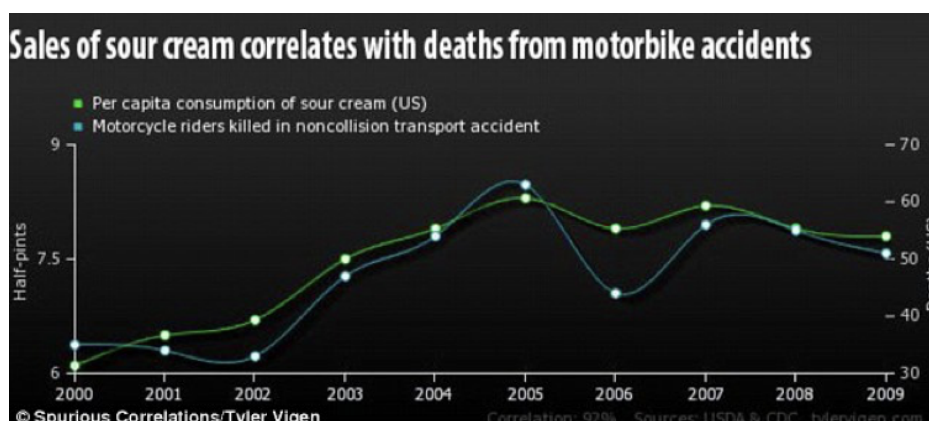


Figura 1.3: Esempio di correlazione spuria

ultime due variabili non sono legate tra loro da una relazione causa-effetto (A implica B e A implica C, ma B non implica C).

Il confounding non casuale è spesso difficile da evidenziare, soprattutto se le correlazioni misurate sono spiegabili. Inoltre, non è eliminabile con un'accurata randomizzazione, ma solo con l'esecuzione di un esperimento totalmente controllato, nel quale ci si preoccupa di rilevare tutte le variabili necessarie per spiegare gli effetti riscontrati. Di questo è importante tener conto soprattutto negli esperimenti osservazionali, dove il controllo è sempre più difficile e meno completo.

Pseudo-repliche e randomizzazione poco attenta

Per evidenziare questi problemi e comprendere meglio la differenza tra un esperimento corretto e uno non corretto, è utilissima la classificazione fatta da Hurlbert (1984), che riportiamo di seguito.

Vengono mostrati 8 soggetti, sottoposti a due trattamenti (bianco e nero), con 8 disegni sperimentali diversi.

Il disegno A1 è corretto, in quanto si tratta di un esperimento completamente randomizzato. Ugualmente, è valido il disegno A2, nel quale le unità sperimentali sono state divise in quattro gruppi omogenei e sono state trattate in modo randomizzato all'interno di ogni gruppo.

Il disegno A3 è quantomeno 'sospetto': vi sono repliche vere, ma l'allocazione dei trattamenti non è randomizzata ed avviene con un processo sistematico per il quale 'nero' e 'bianco' si alternano. Cosa succederebbe se vi fosse

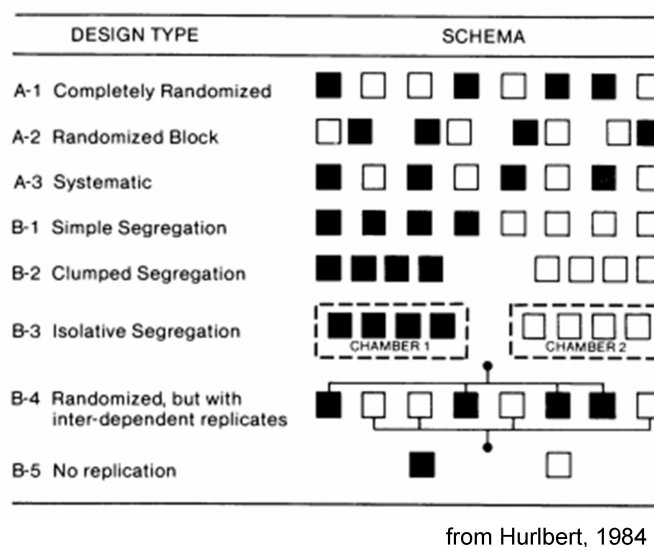


Figura 1.4: Indicazioni per una corretta randomizzazione (Hurlbert, 1984)

un gradiente di fertilità decrescente da destra verso sinistra? Le unità nere sarebbero avvantaggiate rispetto alle bianche! Insomma, rimangono sospetti di confounding, a meno che non si sia assolutamente certi dell'assenza di gradienti, come capita ad esempio se all'interno dei blocchi, dobbiamo creare una sequenza spazio-temporale. Vediamo tre esempi:

1. ho quattro piante e, per ogni pianta, voglio confrontare un ramo basso con uno alto: è evidente che i due trattamenti sono sempre ordinati in modo sistematico (basso prima di alto).
2. Dobbiamo valutare l'effetto di fitofarmaci somministrati in due epoche diverse (accestimento e inizio-levata); anche qui non possiamo randomizzare, giacché un'epoca precede sempre l'altra.
3. Dobbiamo confrontare la presenza di residui di un fitofarmaco a due profondità e non possiamo randomizzare, perché una profondità precede sempre l'altra nello spazio.

In queste situazioni l'esperimento rimane valido, anche se la randomizzazione segue un processo sistematico e non casuale.

Il disegno B1 è usualmente invalido: non vi è randomizzazione e ciò massimizza i problemi del disegno A3: la separazione delle unità sperimentali

‘bianche’ e ‘nere’ non consente una valutazione adeguata dell’effetto del trattamento, che è confuso con ogni potenziale differenza tra la parte destra e la sinistra dell’ambiente in cui la sperimentazione viene eseguita. Ovviamente, la separazione può essere non solo spaziale, ma anche temporale. Anche in questo caso diamo alcuni esempi in cui una situazione come quella descritta in B1 è valida:

1. Vogliamo confrontare la produzione in pianura e in collina. Ovviamente dobbiamo scegliere campioni in due situazioni fisicamente separate
2. Vogliamo confrontare la pescosità di due laghetti
3. Vogliamo confrontare la produttività di due campi contigui.

Queste situazioni sono valide, anche se con una restrizione: non siamo in grado di stabilire a chi debba essere attribuito l’effetto. Ad esempio, per la prima situazione, pianura e collina possono dare produzioni diverse per il suolo diverso, il clima diverso, la precessione colturale diversa o un qualunque altro elemento che differenzi le due località.

Il disegno B2 è analogo al disegno B1, ma il problema è più grave, perché la separazione fisica è più evidente. Questo disegno è totalmente sbagliato, a meno che non siamo specificatamente interessati all’effetto località (vedi sopra).

Il disegno B3 è analogo al disegno B2, ma costituisce una situazione molto frequente nella pratica scientifica. Immaginiamo infatti di voler confrontare la germinazione dei semi a due temperature diverse, utilizzando due camere climatiche e mettendo, in ognuna di esse, quattro capsule Petri identiche. In questa situazione, l’effetto temperatura è totalmente confuso con l’effetto ‘camera climatica (località)’ e risente di ogni malfunzionamento relativo ad una sola delle due camere. Inoltre, le unità sperimentali con lo stesso trattamento di temperature non sono manipolate in modo indipendente, dato che condividono la stessa camera climatica. Di conseguenza, non si può parlare di repliche vere, bensì di **pseudorepliche**.

Altri esempi di **pseudorepliche** sono schematizzati con il codice B4. Ad esempio:

1. trattare piante in vaso ed analizzare in modo indipendente i singoli individui invece che tutto il vaso;
2. trattare una parcella di terreno e prelevare da essa più campioni, analizzandoli separatamente;
3. trattare una capsula Petri ed analizzare separatamente i semi germinati al suo interno.

Questi disegni, in assenza di repliche vere aggiuntive non sono da considerarsi validi. Ad esempio, se io ho due vasetti trattati in modo totalmente indipendente e da ciascuno di essi prelevo due piante e le analizzo separatamente, il disegno è caratterizzato da due repliche vere e due pseudorepliche per ogni replica ed è, pertanto, valido.

Il disegno B5 è invece evidentemente invalido, per totale mancanza di repliche.

1.7 Chi valuta se un esperimento è attendibile?

Quanto detto finora vorrebbe chiarire come il punto centrale della scienza non è la certezza delle teorie, bensì il metodo che viene utilizzato per definirle. Ognuno di noi è quindi responsabile di verificare che le informazioni in suo possesso siano ‘scientificamente’ attendibili, cioè ottenute con un metodo sperimentale adeguato. Il fatto è che non sempre siamo in grado di compiere questa verifica, perché non abbiamo strumenti ‘culturali’ adeguati, se non nel ristretto ambito delle nostre competenze professionali. Come fare allora?

L’unica risposta accettabile è quella di controllare l’attendibilità delle fonti di informazione. In ambito biologico, le riviste autorevoli sono caratterizzate dal procedimento di ‘*peer review*,’ nel quale i manoscritti scientifici, prima della pubblicazione, sono sottoposti ad un comitato editoriale ed assegnati ad un ‘editor,’ il quale legge il lavoro e contemporaneamente lo invia a due o tre scienziati anonimi e particolarmente competenti in quello specifico settore scientifico (*reviewers* o revisori).

I revisori, insieme all’*editor*, compiono un attento lavoro di esame e stabiliscono se l’evidenza scientifica presentata è sufficientemente ‘forte.’ Le eventuali critiche vengono presentate all’autore, che è tenuto a rispondere in modo convincente, anche ripetendo gli esperimenti se necessario. Il processo richiede spesso interi mesi ed è abbastanza impegnativo per uno scenziato. E’ piuttosto significativa l’immagine presentata in scienceBlog.com, che allego qui.

In sostanza il meccanismo di *peer review* porta a rigettare un lavoro scientifico in presenza di qualunque ragionevole dubbio metodologico. Desideriamo sottolineare che abbiamo parlato di dubbio metodologico, dato che il dubbio

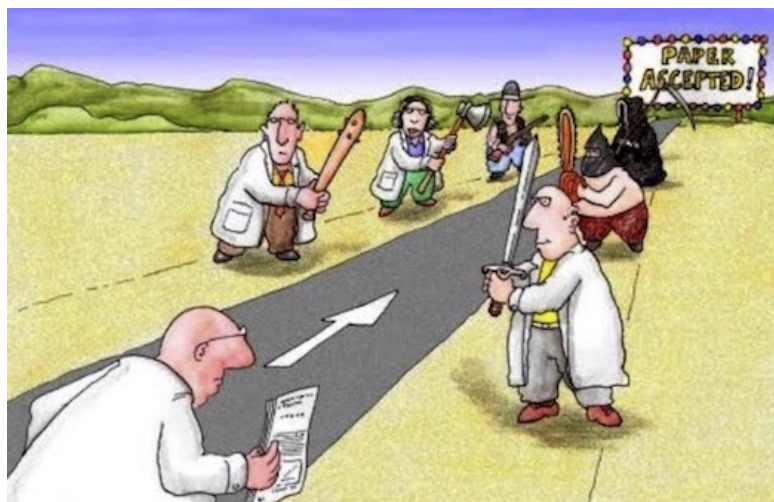


Figura 1.5: Il processo di peer review

sul risultato non può essere allontanato completamente e i reviewer controlleranno solo che il rischio di errore sia al disotto della soglia massima arbitrariamente stabilita (di solito pari al 5%). Questo procedimento, se effettuato con competenza, dovrebbe aiutare a separare la scienza dalla pseudo-scienza e, comunque, ad eliminare la gran parte degli errori metodologici dai lavori scientifici.

1.8 Conclusioni

In conclusione, possiamo ripartire dalla domanda iniziale: “Che cosa è la scienza?” per rispondere che è scienza tutto ciò che è supportato da dati che abbiano passato il vaglio della *peer review*, dimostrando di essere stati ottenuti con un procedimento sperimentale privo di vizi metodologici e di essere sufficientemente affidabili in confronto alle fonti di incertezza cui sono associati.

Qual è il *take-home message* di questo capitolo? Fidatevi solo delle riviste scientifiche attendibili, cioè quelle che adottano un serio processo di *peer review* prima della pubblicazione.

1.9 Altre letture

1. Fisher, Ronald A. (1971) [1935]. The Design of Experiments (9th ed.). Macmillan. ISBN 0-02-844690-9.
2. Hurlbert, S., 1984. Pseudoreplication and the design of ecological experiments. Ecological Monographs, 54, 187-211
3. Kuehl, R. O., 2000. Design of experiments: statistical principles of research design and analysis. Duxbury Press (CHAPTER 1)

Capitolo 2

Progettare un esperimento

Placeholder

2.1 Gli elementi della ricerca

2.2 Ipotesi scientifica → 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

Capitolo 3

Richiami di statistica descrittiva

Placeholder

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

Capitolo 4

Modelli statistici ed analisi dei dati

Placeholder

- 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

Placeholder

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

Capitolo 6

Decisioni ed incertezza

Placeholder

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 χ^2

6.2.1 Simulazione Monte Carlo

6.2.2 Soluzione formale

6.3 Conclusioni e riepilogo

6.4 Altre letture

Capitolo 7

Modelli ANOVA ad una via

Placeholder

- 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 σ
- 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

Capitolo 8

La verifica delle assunzioni di base

Placeholder

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

Capitolo 9

Contrasti e confronti multipli

Placeholder

- 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

Capitolo 10

Modelli ANOVA con fattori di blocco

Placeholder

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 σ

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

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 Altre letture

Capitolo 12

Modelli ANOVA a due vie con interazione

Placeholder

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

Capitolo 13

Breve introduzione ai modelli misti

Placeholder

13.1 Raggruppamenti tra parcelle

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

Capitolo 14

La regressione non-lineare

Placeholder

- 14.1 Caso studio: degradazione di un erbicida nel terreno
- 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’...

Capitolo 15

Esercizi

Placeholder

15.1 Capitoli 1 e 2

15.1.1 Esercizio 1

15.1.2 Esercizio 2

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

Capitolo 16

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 di numeri progressivi

Immissione dei codici delle tesi e dei blocchi

Immissione dei valori e creazione del dataframe

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