

STUDI BIOMETRICI

Bruno Foggi
Dipartimento Biologia Vegetale
Università degli Studi di Firenze

Sommario

- 1.1 Introduzione
- 1.2 La statistica
- 1.3 Piano di lavoro e disegno sperimentale
 - 1.3.1 Le caratteristiche morfologiche
 - 1.3.2 Il campionamento
 - 1.3.3 Approvvigionamento del materiale

1.1 Introduzione

J. Ray (1628-1705) fu probabilmente uno dei primi botanici ad occuparsi della variazione delle piante in maniera scientifica e ad accorgersi che, anche all'interno di una specie, potevano esistere individui diversi da quelli normali detti "accidenti". Con il passare del tempo l'identificazione dei cosiddetti modelli normali e il loro range di variazione diviene uno strumento sempre più indispensabile per il tassonomo.

Le applicazioni matematiche allo studio della variabilità degli organismi si diffondono a partire dalla seconda metà del 1800 in seguito ai lavori di Darwin. Il primo studioso che applicò i metodi statistici allo studio della variazione in Natura fu, probabilmente, il belga M. A. Quetelet (1796-1874): si deve a lui la definizione di "distribuzione normale" e di "dispersione". Bisogna comunque arrivare all'inizio del 1900 per avere i primi lavori di studio della variabilità applicati al mondo vegetale, come quello di PEARSON (1900) su *Papaver rhoes*, e il primo libro di statistica applicata alla biologia di DAVEMPORT (1904).

1.2 La statistica

Gli individui sono variabili: se non ci fosse variazione non ci sarebbe bisogno della statistica. La statistica, definita come "insieme delle metodologie adottate nella raccolta e nell'elaborazione di dati di tipo quantitativo" (FOWLER & COWEN, 1993), risulta lo strumento principe, non solo per organizzare, sintetizzare e descrivere i dati di tipo quantitativo, ma anche per esplorarli, soprattutto dove questi siano molto numerosi, alla ricerca di indizi di discontinuità all'interno della variabilità.

Prima di affrontare qualsiasi studio che preveda l'uso dell'analisi statistica bisogna sempre avere in mente due cose fondamentali:

1) l'analisi statistica non permette di provare niente, ma solo di valutare il grado di affidabilità delle ipotesi e oggettivare i procedimenti che hanno permesso di valutarle;

2) la statistica non genera dati di nuovo tipo, essa è semplicemente uno strumento per rendere universali le osservazioni eseguite su una parte dell'insieme oggetto di studio, definita campione.

Nel caso della ricerca sistematica, uno dei principali usi delle metodologie statistiche è quello di provare se esistano o meno delle relazioni fra le informazioni derivate dall'analisi delle caratteristiche morfologiche di un gruppo di individui (oggetti o campioni), al tutto (popolazione statistica). La popolazione statistica è una astrazione, essa è costituita da unità di campionamento, dalle quali può venire estratto un campione. La popolazione statistica è ben diversa da quella bioecologica: per esempio alla prima partecipano tutti gli individui che possono entrare a far parte di un campione, mentre della seconda fanno parte anche quegli individui che non potranno essere campionati.

L'analisi statistica troverà il suo maggiore campo di applicazione nello studio della variazione a livello intraspecifico, o comunque all'interno di gruppi costituiti da specie molto simili: risulta

evidente che se due specie sono palesemente diverse, la statistica non serve: sarebbe come guardare un elefante con un microscopio.

Essa può comunque servire per:

- analizzare la variazione dei caratteri;
- ipotizzare se due o più gruppi siano diversi e, eventualmente, verificare quanto lo sono (verifica di ipotesi);
- verificare se i campioni si raggruppano fra loro e in quale modo: per esempio campioni provenienti da aree geografiche separate, oppure da habitat diversi (esplorazione dei dati);
- ricercare la presenza di correlazioni fra variabilità fenotipica e particolari situazioni ambientali e/o geografiche.

Le analisi statistiche possono essere divise in due grandi gruppi:

a) statistica descrittiva: ha lo scopo di organizzare, sintetizzare e descrivere le misure di un campione; essa non consente di formulare previsioni

b) statistica inferenziale: ha lo scopo di dedurre o prevedere i parametri di una popolazione statistica da un campione estratto da questa

1.3 Piano di lavoro o disegno sperimentale

Una ricerca deve essere vista come un processo interattivo fra attività speculative (teoria) e raccolta di dati (pratica), dove queste due fasi devono essere strettamente connesse (JAGER & LOOMAN, 1987). In funzione delle caratteristiche bio-ecologiche del gruppo e degli obiettivi che voglio raggiungere dovrà essere messo a punto un piano di lavoro o disegno sperimentale.

Secondo COCHRAN (1977) lo sviluppo di un piano di lavoro dovrebbe seguire i seguenti passi:

- 1) definire con chiarezza verso chi è diretta la ricerca: individui, popolazioni, altri gruppi
- 2) descrivere gli oggetti che verranno studiati
- 3) definire le variabili da osservare e studiare
- 4) definire il piano di campionamento
- 5) determinare il livello di accuratezza che voglio raggiungere
- 6) scegliere i metodi statistici più opportuni
- 7) valutare gli obiettivi raggiunti alla luce dei punti precedenti
- 8) intervenire con eventuali aggiustamenti e interazioni con la serie dei punti sopra descritti al passo opportuno

Nella pratica giornaliera un ricercatore segue questa serie di procedure più o meno intuitivamente ma la buona riuscita di un lavoro che utilizza tecniche statistiche, dipende dalla precisione con la quale vengono eseguiti i passi del piano di lavoro deciso.

Osservare e studiare le caratteristiche morfologiche sugli individui campionati sono importanti fasi di lavoro che vengono spesso eseguite, almeno in parte, utilizzando gli esiccata già esistenti nelle collezioni di un erbario, ed è quindi su queste che ci soffermeremo. Per la parte più propriamente statistica si rimanda ai vari testi e manuali sull'argomento, alcuni dei quali presenti in bibliografia..

1.3.1 Le caratteristiche morfologiche

Gli individui sono descritti attraverso le loro caratteristiche morfologiche che, per ragioni di semplicità, chiameremo d'ora in poi "caratteri". Questi possono essere per esempio: lunghezza e larghezza delle foglie basali, altezza degli individui maturi, larghezza dei petali o dei sepali, colore dei petali o dei sepali, pelosità dei vari organi ecc. In genere i caratteri da prendere in considerazione si possono ricercare fra quelli usati nelle descrizioni, nelle chiavi di riconoscimento, nelle diagnosi ecc. La scelta dei caratteri è un passo molto importante e deve prevedere una certa conoscenza del gruppo di piante oggetto dello studio, quindi il metodo seguito è induttivo, anche se una sorta di deduttività non può e non deve essere esclusa: risulta assurdo per esempio confrontare una quercia con una poa, le due specie sono talmente diverse che qualsiasi persona può essere in grado di valutarne i due diversi modelli di diversità. Sarà inutile prendere in considerazione parti di

una pianta che generalmente, in quel gruppo, non vengono prese in considerazione: come, per esempio, confrontare due graminacee attraverso l'uso del carattere "colore dei fiori" poiché tutte le graminacee hanno fiori incolore; in genere difficilmente nelle chiavi e nelle descrizioni vengono presi in considerazione caratteri derivati dall'apparato radicale, quindi scavare per raccogliere gli apparati radicali risulterà, in genere, un'azione che determina perdite di tempo; mentre nel caso del genere *Aristolochia* L., essendo molte chiavi basate sulla morfologia della caulirizza, sarà invece necessario considerare anche tale carattere. Da notare che in alcuni generi, o gruppi di specie, le misurazioni di alcuni caratteri devono seguire delle procedure standardizzate che servono per poterli paragonare fra loro (per esempio, nel genere *Festuca* L., la lunghezza del lemma deve essere riferita a quella del secondo lemma della spighetta). Comunque più caratteri vengono presi in considerazione maggiore sarà l'attendibilità dello studio. Alcuni autori sconsigliano l'uso di caratteri derivati da prodotti, divisioni o comunque da operazioni che prendono in considerazione altri caratteri, perché in questo caso questi verrebbero considerati due volte ed il loro peso nella valutazione totale sarebbe maggiore rispetto agli altri.

Ciascun carattere presenta modelli di variazione diversi e tale variazione può essere apprezzata in vari modi:

caratteri bistato o binari: essi sono riferiti alla valutazione della presenza o assenza di quel carattere o del suo stato, quando quel carattere può avere solo due stati (foglie glabre vs. foglie pelose; fiori bianchi vs. fiori colorati ecc);

caratteri multistato qualitativi: sono caratteri che presentano vari stati ma di valore discontinuo; essi sono distribuiti in classi di grandezza per cui ad una categoria corrisponde un valore numerico: glaucescenza nulla, media, molta; pelosità di un organo: nulla, media, molta;

caratteri multistato quantitativi: sono caratteri che variano in maniera continua, come le misure. Es. lunghezza dei sepali; larghezza delle foglie inferiori; ma anche numero dei peli su unità di superficie (il carattere multistato precedente viene quindi analizzato con maggiore dettaglio trasformandolo in un carattere quantitativo continuo).

Come abbiamo visto per la pelosità, in teoria tutte le caratteristiche possono essere quantificate e tradotte in misure, come per esempio il colore che, pur essendo un carattere qualitativo, può essere convertito in quantitativo continuo, misurando la quantità di pigmento per unità di volume.

I caratteri quantitativi sono per loro natura disposti in senso crescente (o decrescente) mentre quelli qualitativi multistato possono non esserlo, per cui vanno disposti in classi crescenti (o decrescenti) a seconda di come sono stati usati gli altri. Per esempio se abbiamo delle misure di lunghezza, larghezza o altro, i caratteri qualitativi, come la pelosità, dovranno essere codificati facendo coincidere l'ampiezza minore del carattere con il valore minimo dello stato, quindi pelosità nulla sarà, per esempio, uguale ad 1 (in genere è meglio non usare lo zero perché alcuni algoritmi matematici utilizzate da alcuni softwares prevedono l'uso dello 0 come dato mancante).

1.3.2 Il campionamento

Il campionamento non è un'attività isolata dalle altre, è il punto di partenza della ricerca quindi la strategia da adottare deve essere una diretta conseguenza dei risultati che si vogliono ottenere e del tipo di oggetti che devo studiare.

Affinchè un campione sia utilizzabile, il processo di campionamento deve sottostare ad alcune regole:

- nessuna considerazione deve essere fatta, a priori, sulla scelta degli individui da campionare: mai prendere il più grande o il più piccolo, il più bello o il più brutto ecc.;
- il campione deve essere raccolto attraverso una scelta casuale "...take the individuals methodically at random and without possible selection of individuals on the basis of magnitude of the character to be measured" (DAVEMPORT, 1904): nel caso di campioni d'erbario bisognerebbe numerarli e poi sceglierli attraverso un generatore di numeri casuali (esistono tabelle e programmi per PC, ma può essere sufficiente anche un sacchetto chiuso con dei numeri dentro);

- un individuo del campione deve essere completo, cioè deve contenere tutti gli elementi che servono per la sua descrizione e che verranno utilizzati per caratterizzarlo: essi devono essere misurabili per cui una prima analisi dev'essere fatta individuando i campioni idonei, sarà tra questi che verrà operata la scelta casuale;
- il campione deve essere numericamente significativo ovvero composto da un numero di individui tale da riassumere una parte significativa della variazione complessiva. Il numero ideale è quello necessario e sufficiente per descrivere il tutto. Per esempio considerando di dover lavorare su popolazioni, dall'analisi bibliografica disponibile, risulta che un buon campione di una popolazione è costituito da un numero di individui che va da 15 a 25; al di sopra l'informazione, in genere, non aumenta in modo tale da giustificare ulteriori spese di tempo e denaro;
- il campione deve essere omogeneo cioè non deve contenere delle sottopopolazioni, cosa che consente di diminuire il range di variazione complessivo e quindi di rendere più facile l'individuazione di discontinuità.

Nella ricerca biologica un campionamento completamente casuale può risultare difficile da effettuare ed in genere si ricorre ad una prima divisione della variabilità totale, definendo a priori dei gruppi in funzione di una o più caratteristiche della popolazione, e limitando la casualità solo all'interno di questi (campionamento stratificato).

Per approfondimenti e chiarimenti, soprattutto per quanto il campionamento in ecologia si consiglia di consultare ROSSI (1979).

1.3.3 Approvvigionamento del materiale

Abbiamo visto come la fase di campionamento sia di fondamentale importanza quando si voglia studiare la variazione degli organismi con tecniche biometriche. La disponibilità di materiale idoneo può quindi essere un fattore limitante a questi tipi di studio.

Durante la fase di esplorazione della variabilità di una specie alla ricerca della presenza di eventuali gruppi di diversità, si dovrà cercare di utilizzare campioni provenienti da tutta l'area di distribuzione presunta e da tutti i tipi di habitat dove essa vive: altitudini diverse, vari tipi di substrato, tipi di vegetazione, ecc. In questa fase il materiale presente negli erbari può risultare di fondamentale importanza, ma nella maggior parte dei principali musei le raccolte riflettono un modello di distribuzione casuale (purtroppo non in senso statistico !) e a volte bizzarro o capriccioso (DAVIS, 2001). Durante la fase di esplorazione della variabilità si può ovviare a questo problema ricorrendo al prestito di materiale da vari erbari e in questo i piccoli erbari, che contengono collezioni a carattere locale, possono giocare un ruolo molto importante. Quindi si dovrà utilizzare i grandi erbari per studiare *exsiccata* sparsi per l'intera area di distribuzione, mentre i piccoli erbari per ricoprire i "buchi" distributivi eventualmente presenti. Il materiale osservato dovrà essere molto abbondante perché dovrà essere completo dei caratteri scelti per essere misurati, ed essere distribuito più o meno omogeneamente in funzione di area geografica e tipi di habitat nei quali l'organismo oggetto di studio può vivere. La fase che utilizza i campioni provenienti dagli erbari può costituire un primo momento di esplorazione della situazione esistente. Durante questo primo approccio si possono evidenziare le aree dove dovrebbero essere eseguite ulteriori approfondimenti e indirizzare le raccolte sul campo.

Diversa è invece la situazione durante questa seconda fase in quanto essa può venire diretta verso lo studio della variabilità intrapopolazionale. In questo caso nessun erbario avrà a disposizione collezioni utili a questo scopo in quanto, in genere, queste si sono costituite nel tempo in seguito a studi di floristica, nei quali l'aspetto quantitativo non veniva preso mai in considerazione. Oggi è ampiamente accettato, come unico approccio pragmatico, che uno studio debba essere indirizzato verso un taxon, per cui occorre effettuare un'opera di raccolta selettiva ma intensa. Diviene quindi indispensabile per il sistematico, inteso come lo studioso della variabilità, andare a raccogliere i campioni direttamente sul campo, procedura che, però determina una forte lievitazione dei costi di questo tipo di ricerche. Di particolare importanza nella ricerca sistematica diviene quella che AUQUIER & KERGUÉLEN (1978) definiscono "*typification biosystematique*" che consiste nella

individuazione delle popolazioni dei “*loci classici*” e procedere al loro campionamento in modo da eliminare al massimo le insicurezze determinate da diagnosi incomplete.

La raccolta di grandi quantità di materiale potrebbe determinare un problema locale alla conservazione della biodiversità e quindi ci si può trovare di fronte ad un problema etico: documentare o conservare ? Oggi nessun ricercatore dotato di un minimo senso di responsabilità vorrebbe creare ulteriori problemi alla conservazione del patrimonio naturale, comunque è bene che chiunque si trovi a campionare in Natura dovrebbe per prima cosa pensare se questa operazione sia veramente utile e, quando questa lo sia effettivamente, chiedersi se essa provocherà perdite di diversità e se tali perdite potranno essere assorbite dalla popolazione locale.

Una buona norma dovrebbe essere che i campioni raccolti debbano entrare a far parte di collezioni di erbario in modo da poter essere consultati da altri ricercatori e fornire così un esauriente esempio della variabilità locale di quel taxon. Gli erbari divengono quindi, non solo una fonte di dati, ma anche un efficiente sistema di immagazzinamento: delle vere e proprie “biblioteche della diversità del mondo vivente” (TATTERSTALL, 1992).

Per avere ulteriori riscontri, alcuni campioni vivi, provenienti dalle popolazioni campionate, dovranno inoltre essere posti in coltura in condizioni standard in modo da verificare mantenimento dei caratteri. Ecco che quindi un altro tipo di museo, l’Orto Botanico, acquista nuova importanza: il mantenimento in vita di esempi della variabilità riscontrata in Natura.

Bibliografia

- AUQUIER P. & KERGUÉLEN M. 1978 Un groupe embrouillé de *Festuca* (Poaceae): les taxones désignés par l’épithète “*glauca*” en Europe occidentale et dans les régions voisines. *Lejeunia*, n.s., 89: 3-82 (1977).
- COCHRAN W. G. 1977 *Sampling techniques*. 3 ed. Wiley, New York.
- DAVEMPORT C. B. 1904 *Statistical methods with special references to biological variation*. II Ed., Chapman & Hall, London.
- DAVIS P. 2001 *Musei e ambiente naturale*. Clueb, Bologna.
- FOWLER J. & COWEN L. 1993 *Statistica per ornitologi e naturalisti*. Muzio ed., Padova.
- JAGER J. C. & LOOMAN C. W. N. 1987 *Data collection*. In JONGMAN R. H. G., TER BRAAK C. J. F. & VAN TONGEREN O. F. R. *Data analysis in community and landscape ecology*. Pudoc, Wageningen.
- PEARSON K. 1900 *The grammar of Science*. II Ed., Black, London.
- ROSSI O. 1979 *Metodi statistici di campionamento in ecologia*. Unicopli, Milano.
- TATTERSTALL I. 1992 *Systematics versus ecological diversity: the example of the Malgasy primates*. In ELDREDGE N. *Systematics, Ecology and the Biodiversity crisis*. Columbia University Press, New York.