

DIMENSIONAMENTO AMOSTRAL EM ESTUDOS DESCRITIVOS DE COMUNIDADES DE ORGANISMOS BÊNTICOS SÉSSEIS E SEMI-SÉSSEIS.

ROSSO, S.

Resumo:

Dada a grande importância do correto dimensionamento das amostras face à responsabilidade de se revelar da maneira mais fiel possível os fenômenos e estruturas que caracterizam os ecossistemas, bem como o anseio de nossos pesquisadores por uma síntese sobre esse tema, vimos prestar aqui nossa contribuição. Inicialmente enumeramos as principais condições que fazem uma boa amostra, ressaltando a apreciação preliminar da heterogeneidade do domínio sobre o qual devem recair as conclusões e melhores forma, tamanho e número de elementos amostrais. Após analisar a questão do tamanho de cada elemento, abordamos com detalhes dois modos de determinar o tamanho total da amostra, a saber, aplicando modelos estatísticos teóricos e suas propriedades, ou então estratégias espectrais. No primeiro caso, apresentamos diversas técnicas e algoritmos aplicáveis quando as variáveis se distribuem normalmente, ou seguem os modelos de Poisson ou binomial negativo, considerando também os pressupostos mais importantes e como satisfazê-los. No último caso, analisamos como se faz o incremento de área e como se interpretam curvas dos descritores em função do tamanho amostral. Mostramos finalmente algumas formas de expressar relações custo-benefício (pontos Molinier e Calleja, critério de CAIN & CASTRO, auto-correlação).

Abstract:

"Sample size in descriptive studies of communities of sessile and semi-sessile benthic organisms"

No one can deny the great value of adequate sample sizes to give accurate pictures of ecosystem structures and processes. The lack of a synthesis in that subject has led me to review this issue for researchers in Ecology. Initially I enumerate the main conditions required to obtain a satisfactory sample, including previous determination of study site heterogeneity and of the best shape, size and number of units (S.U.'s) to be sampled. Following an analysis of S.U. size related aspects, I review two ways for determining the ideal S.U. number in the entire sample: (1) using theoretical statistical models and their properties, and, on the other hand, (2) adopting spectral strategies. In the first case I present several techniques and algorithms used when the variables are normally distributed or when fitting Poisson or negative binomial distributions. In addition I include considerations on the premises involved and how to achieve them, at least in part, when the data do not fulfill such conditions. In the second case I consider some techniques to increase sampling area and to examine the behavior of curves obtained by plotting descriptor values against area or sample size, fitted to some common mathematical model (e.g. power, exponential or Michaelis-Menten functions). Finally, I discuss some strategies to define cost-benefit relationships as expressed by Molinier and Calleja points, CAIN & CASTRO criterion and auto-correlation.

Aspectos Gerais

O conhecimento sobre a estrutura das comunidades, especialmente sua dinâmica temporal, é essencial na avaliação da influência das perturbações, antrópicas ou não. A diversificação da biota no que tange aos hábitos, dimensões, formas de ocorrência e outras características dos organismos, entretanto, faz com que os métodos descritivos não possam ser concebidos para aplicação num contexto generalizado. Distinguem-se, por um lado, métodos destinados à investigação de comunidades de organismos fixos ao substrato, ou pelo menos pouco móveis na escala de observação adotada; por outro, há aqueles voltados para sistemas de organismos vágeis, tanto num espaço bi- (é o caso de ecossistemas terrestres e de parte da biota bentônica marinha ou límnic), como tri-dimensional (como as comunidades do plâncton e do nécton, por exemplo).

Dada a crescente procura de informação sobre técnicas de dimensionamento amostral, vem se delineando a necessidade de apresentar à comunidade científica uma síntese seletiva do assunto. Não se pretende aqui, todavia, suplantar a literatura existente exaurindo o tema, e sim satisfazer os anseios de muitos pesquisadores.

Já é bastante antiga a discussão sobre a interferência do esforço amostral nos resultados de trabalhos descritivos, pelo menos ao nível das comunidades, procurando diversos autores inclusive estabelecer meios capazes de levar à independência entre os dois aspectos (GLEASON, 1922; MARGALEF, 1951; MENHINICK, 1964; e outros). BOUDOURESQUE (1974) lamentou que, no âmbito das ciências do mar, a maioria dos fitossociólogos vinha adotando (e isso, de certo modo, perdura até hoje) critérios muito subjetivos de dimensionamento amostral, sem embasamento teórico (NIELL, 1977). Aliás, muitas decisões são baseadas em trabalhos anteriores, que não provêem necessariamente informação útil (FUENTES & NIELL, 1981).

WEINBERG (1981) revisou uma série de técnicas de amostragem de comunidades ou populações de organismos sésseis e semi-sésseis, destacando aquelas baseadas em elementos ou unidades amostrais não pontuais ou lineares, e sim com área definida (*quadrats*) - sem dúvida as mais largamente utilizadas. Trata-se, no caso, de quantificar o descritor analítico básico num conjunto de elementos que se considera representar bem a realidade do sistema-objeto sobre o qual deverão recair as conclusões.

O problema da representatividade é básico em qualquer estudo ecológico - uma boa amostra deve oferecer uma imagem se possível completa, qualitativa e quantitativa, da comunidade estudada. É justamente daí que emergem as questões mais importantes, sobre as quais repousa todo o estudo de dimensionamento amostral, particularmente da chamada *área mínima* (GREEN, 1979; GREIG-SMITH, 1983):

- qual a eficiência do aparato amostrador?
- há repartições em escala espacial maior, indicando estratificação ou zonação?
- qual a forma mais apropriada dos *quadrats*?
- qual o tamanho ideal?
- qual o número mais conveniente, num compromisso entre acurácia/precisão e custo ou esforço amostral?

Quando o descritor analítico básico já é ele próprio uma estimativa ao nível de cada elemento amostral, como é o caso da determinação do recobrimento percentual através do método dos contatos (*point quadrat method*; GOUNOT, 1969), a questão do dimensionamento torna a impor-se:

- qual é a influência da densidade e forma de registro dos contatos, ou seja, quantos devem ser e como devem estar dispersos os pontos observados em cada *quadrat*?

Fazer uma amostragem preliminar é a única forma de resolver questões que potencialmente podem converter-se em sérios problemas - não há substitutos para ela e raramente se aconselha dispensá-la; mesmo assim, esse é um dos aspectos mais subestimados nos estudos de campo. Embora muitos lembrem o dispêndio de tempo e esforço envolvido na preliminar, o aparente obstáculo normalmente é compensado mais tarde. Muitas vezes é até possível incorporar os resultados ao lote principal de dados (GREEN, 1979).

A maioria das abordagens sobre dimensionamento amostral, principalmente aquelas constantes em muitas obras de estatística aplicada ou ecologia quantitativa, enfatiza técnicas envolvidas na determinação de características das diferentes populações isoladamente. Nos estudos sinecológicos, todavia, costumam ser numerosas as populações a quantificar. Nesses casos, a partir dos efetivos específicos levantados por amostragem, procura-se representar atributos globais da comunidade na forma de descritores sintéticos, simples ou complexos.

Um primeiro ponto que surge daí é a questão de que cada população tem suas próprias características, entre as quais a densidade populacional e o tamanho e modo de dispersão dos indivíduos, exigindo uma dimensão amostral específica. Variando a densidade populacional (que é uma abundância absoluta média por unidade de espaço) os parâmetros estatísticos podem variar também; assim, um método adequado para populações mais densas pode não sê-lo para as mais rarefeitas (SOUTHWOOD, 1966). GREIG-SMITH (1983) sugeriu quantificar todas as espécies, eliminando uma após outra assim que atingido seu tamanho amostral próprio.

Outro aspecto é o fato de que diferentes descritores sintéticos, particularmente os índices, apresentam-se com distintas sensibilidades em relação

aos efetivos das espécies mais raras, via-de-regra justamente aquelas que exigem maior esforço amostral para sua quantificação. Por isso o dimensionamento deve estar diretamente vinculado não apenas ao descritor analítico básico adotado (número de indivíduos, recobrimento percentual ou biomassa das diferentes espécies, por exemplo) mas também, e principalmente, ao descritor sintético a ser calculado, repousando a validade dos resultados basicamente na representatividade dos efetivos que sobre ele se mostram mais influentes.

A heterogeneidade característica não só de sistemas de costão rochoso, como também da maioria dos ecossistemas, é um sério complicador em trabalhos quantitativos, daí a conveniência de iniciar o estudo com uma análise da repartição espacial. Além do interesse descritivo próprio desta, onde inclusive se revelam aspectos da distribuição de recursos, a delimitação de áreas homogêneas é uma condição básica para por exemplo evidenciarem-se alterações nas comunidades através de índices de diversidade e dominância. O ideal é assegurar que cada elemento amostral se refira a um e apenas um tipo de substrato ou povoamento, devendo-se evitar sua alocação em zonas de transição. Numa conceituação qualitativa, comunidades são grupos de espécies que tendem a re-ocorrer juntas (WIESER, 1960); a delimitação dos espaços ocupados por essas associações é então algo muito importante no estudo de povoamentos heterogêneos, dentro dos quais é neces-sário definir áreas mais uniformes (também no plano quantitativo), a serem tratadas como entidades idealmente homogêneas *na escala de observação adotada* (FRONTIER, 1983). GREEN (1979) enfatizou a necessidade de alocar aleatoriamente os elementos amostrais em estratos *relativamente homogêneos*, minimizando a variabilidade intraestratos e fazendo restar principalmente a inter-estratos. A própria natureza probabilística dos efetivos tomados em valores relativos, usados como base para o cálculo de muitos índices, já exige, como pressuposto, a homogeneidade na dispersão dos indivíduos pela área amostrada. PICHON (1978) propôs o reconhecimento visual das áreas homogêneas. É preciso, entretanto, alertar para os perigos inerentes à subjetividade que cerca esse tipo de conduta, sugerindo-se como alternativa a aplicação de métodos estatísticos de análise de variâncias ou então técnicas de classificação ou ordenação, desde que se mostrem compatíveis com a natureza dos dados.

A forma dos elementos amostrais pode ser até certo ponto relevante, mas pouco se fez no sentido de esclarecer melhor como ela interfere nos resultados. A adoção quase generalizada de elementos quadrados (ou retangulares) é todavia indiscutível.

O tamanho dos elementos, está aí um aspecto muito importante, deve ser com-patível com a escala do distanciamento e tamanho dos indivíduos ou agregados, e com a escala do *patchiness* que se queira revelar - elementos de $0,25m^2$ não evidenciam mosaico na escala de decímetros (BUZAS, 1970; PICHON, 1978; GREIG-SMITH, 1983; SANTELICES, 1980).

O caráter geralmente gregário dos valores, que faz a variância exceder e ser função da média, depende muito do tamanho dos organismos. A área ocupada por eles deve ser bem menor que a dos elementos amostrais - o ideal é 1/20 (GREEN, 1979). Se uma população com indivíduos não aleatoriamente dispersos for amostrada por *quadrats* de tamanho muito menor que a área média dos agregados, então a variância supera pouco, ou nem supera, a média. Com o incremento do tamanho dos elementos, aproximando-se da dimensão dos *patches*, a variância em relação à média aumenta nitidamente - isso reduz a precisão das estimativas. Se o *patchiness* for regular, então espera-se uma subsequente redução até se atingir ou ultrapassar o nível da média. No caso de *patches* não uniformemente dispersos, a variância deve permanecer alta (um complicador para se visualizar essas relações é o fato de que o mosaico geralmente se verifica em diversas escalas). Num contexto estatístico, *quadrats* maiores produzem médias numericamente maiores, o que convenientemente pode reduzir a assimetria verificada em distribuições agregadas e que é tão problemática para os testes de hipótese. Na prática, o melhor tamanho é o menor possível, ao mesmo tempo capaz de fazer com que os elementos vazios não superem em número aqueles com apenas um indivíduo (GREEN, 1979; GREIG-SMITH, 1983). Adotar elementos menores (e mais numerosos) também ajuda a revelar com maior clareza os diferentes micro-habitats. No caso de variáveis com dispersão aleatória, o tamanho dos elementos amostrais não é um fator importante; já nos sistemas agregados os elementos maiores produzem variabilidades relativas muito superiores (GREEN, 1979). É bom lembrar, entretanto, que com a redução dos *quadrats* vai se ampliando a dimensão perimetral em relação à área, ficando gradativamente mais marcante o efeito de problemas de borda, por exemplo a decisão do observador sobre que indivíduos devem ser considerados dentro e fora de cada elemento (quando os "indivíduos" forem grandes ou de contornos mal definidos, o ideal é não adotar elementos amostrais muito pequenos).

Uma opção bastante objetiva para dimensionar os elementos, talvez mais complicada do ponto de vista operacional é, como regra geral em casos onde há *patchiness* e quando os resultados estatísticos devam independe do tamanho da amostra, fazer com que cada elemento amostral (quadrado) tenha lado medindo o mesmo que a distância entre elementos a partir da qual a variância deixa de ser função do afastamento. Inicialmente reduz-se o tamanho dos *quadrats* até que fique bem aquém da escala do mosaico e então aplica-se a regra.

Mais crítica, no entanto, é a determinação do número de elementos da amostra.

Aí, a subjetividade pode ser minimizada por duas vias diferentes. Uma delas é a estimativa com base nas propriedades de certas distribuições estatísticas, a partir de uma variabilidade determinada preliminarmente. Outra, é a que envolve técnicas mais ou menos empíricas, independentes do ajuste da distribuição da variável a este ou aquele modelo estatístico em particular. Nesse sentido, calculam-se os valores dos descritores em situações, simuladas ou não, de

esforço crescente. No caso, como se produzem informações espectrais, o ganho vai muito além do dimensionamento propriamente dito - o conhecimento dos padrões de variação dos diferentes descritores, por exemplo em função da área, mostra grande valor descritivo em si mesmo (PIELOU, 1975; BALLESTEROS, 1986).

Aplicação dos Modelos Estatísticos Teóricos.

Existe proporcionalidade entre o número de elementos amostrais constituinte da amostra e a precisão da estimativa das médias. Embora o dimensionamento dependa da finalidade, se os dados vão passar por um *screening* inicial (por exemplo numa amostragem preliminar) então um mínimo razoável de valores é 30, e o ideal, por volta de 60 (GREEN, 1979).

Normalmente estima-se médias populacionais a partir de médias amostrais. A diferença entre os dois valores é o erro de amostragem, incrementado em maior ou menor grau por erros humanos e pela heterogeneidade ambiental (POOLE, 1974). As médias amostrais que seriam obtidas a partir de um certo número de réplicas, *tomadas de uma mesma população estatística*, seguem sempre uma distribuição normal, independentemente da distribuição da variável original (teorema do limite central): a média, nesse caso, para n grande, coincide com a média populacional que se quer estimar: o desvio-padrão correspondente chama-se erro-padrão da média (Δ), e se relaciona diretamente com a precisão da estimativa:

$$\Delta = \frac{\sigma}{\sqrt{n}}, \text{ onde}$$

σ = desvio-padrão da distribuição da variável original;
 n = número de valores da variável.

Como pode se constatar a partir da expressão, o erro-padrão da média decresce na proporção da raiz quadrada de n . Daí, PRINGLE (1984) propôs a fórmula:

$$n = \left(\frac{\sigma}{\mu \cdot D} \right)^2 \cong \left(\frac{s}{\bar{x} \cdot D} \right)^2 = \left(\frac{s}{\Delta} \right)^2, \text{ onde:}$$

n = número de elementos amostrais;

σ = desvio-padrão populacional;

μ = média populacional;

s = desvio-padrão amostral;

\bar{x} = média amostral;

D = coeficiente de variação ou erro-padrão relativo (Δ / média).

No sentido de impor níveis de confiabilidade $(1-\alpha)$ para o erro-padrão das estimativas em função de n para variáveis com distribuição pelo menos aproximadamente normal, surge a noção de intervalo de confiança da média, que deriva da introdução de um fator t de Student:

$$i.c. = \bar{x} \pm t_{\left(1-\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}}, \text{ onde}$$

\bar{x} = média amostral;

t = valor da tabela de Student para $p=1-(\alpha/2)$;

s = desvio-padrão amostral;

n = número de elementos amostrais;

A determinação do número de elementos amostrais pode então ser vinculada à precisão expressa na forma do intervalo de confiança. Como exemplo, poder-se-ia fazer o dimensionamento para estimar a densidade média de uma população normalmente distribuída, com uma precisão de $\pm 10\%$ e $1/20$ ou 0.05 de chance de erro, ou seja, da média populacional situar-se fora do intervalo fixado ou de a média amostral, em 95% das eventuais réplicas, situar-se dentro do mesmo. Supondo estarem calculados a média \bar{x} e o desvio-padrão amostrais a partir de uma amostragem preliminar de tamanho n_{prel} , bastaria tomar o valor t_p de Student para $p=0,975$ de confiança $(1 - 0.05/2)$ e $(n_{prel} - 1)$ graus de liberdade, considerar que o erro absoluto Δ corresponde a $0,1\bar{x}$, e aplicar esses valores na expressão:

$$\frac{i.c.}{2} = t_p \frac{s}{\sqrt{n}},$$

ou mais diretamente naquela apresentada por SOUTHWOOD (1966), POOLE (1974) e BAKANOV (1985):

$$n = t_p^2 \left(\frac{s}{\Delta} \right)^2.$$

Feito isso, seria necessário repetir a operação, porém usando agora o valor de t correspondente a $(n - 1)$ graus de liberdade.

POOLE (1974), aliás, considerou a proporção entre n (número de valores na amostra) e N (tamanho da população toda), incorporando na fórmula do erro-padrão um fator de correção:

$$\Delta = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}. \text{ Para dados binários, a expressão apresentada pelo autor é:}$$

$$\Delta = \sqrt{\frac{N-n}{(n-1).N}} \cdot p \cdot q, \text{ onde}$$

p = probabilidade associada ao primeiro atributo;

q = probabilidade do atributo complementar.

Se $n = N$, então o erro torna-se nulo, já que não há mais possibilidade de desvio devido ao acaso (a amostra é integral). Quando n excede 10% de N , o mesmo autor aconselha a correção do resultado inicial do dimensionamento, calculando-se n' através da expressão:

$$n' = \frac{n}{1 + \frac{n}{N}}$$

Quando se vai amostrar um sistema com interesse em algo que se apresenta com probabilidade definida anteriormente, pode-se determinar o número adequado de elementos amostrais pela expressão indicada por SOUTHWOOD (1966):

$$n = t^2_{\left(1-\frac{\alpha}{2}\right)} \left(\frac{p \cdot q}{D^2} \right), \text{ onde}$$

t = valor da tabela de Student para margem de erro α

p = probabilidade associada ao primeiro atributo;

q = probabilidade do atributo complementar.

D = erro-padrão relativo.

POOLE (1974) e FRONTIER (1983) apresentaram a técnica para quando se quer amostrar um atributo disperso de modo heterogêneo no espaço amostral, caso em que é preciso levar em conta o fato do erro-padrão da média estimada não ser apenas devido ao desvio médio entre valores da estimativa e média populacional, e sim derivado também das diferenças entre as médias esperadas nos distintos setores do domínio. A amostragem estratificada propriamente dita serve para compensar essas variações entre os setores ou estratos ao se determinar o erro-padrão da média. Uma vez definidos os estratos (seja por meio de análise simples de variância, em estudos auto-ecológicos ou com descritores sintéticos já calculados para cada elemento amostral, seja por técnicas multi-variadas, ou ainda ambas as coisas) retira-se, de cada um, uma amostra constituída por um número razoável de elementos aleatoriamente alocados. A estimativa da média populacional para um estrato *qualquer* é dada por:

$$\bar{Y} = \frac{\sum_{h=1}^{st} N_h \cdot \bar{y}_h}{N}, \text{ onde}$$

st = número total de estratos;

N_h = tamanho do estrato h (número total de indivíduos, por

exemplo);

\bar{y}_h = estimativa da média para o estrato h ;

N = tamanho total geral dos estratos.

Daí se depreende que a estimativa para um estrato qualquer é uma média ponderada (pelos tamanhos dos estratos) das estimativas parciais.

Considerando o peso de cada estrato no contexto de todo o domínio como sendo $\bar{W}_h = \frac{N_h}{N}$, então o erro-padrão da média \bar{Y} estimada é:

$$\Delta(\bar{Y}) = \sqrt{\sum_{h=1}^{st} W_h^2 \cdot \frac{s_h^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right)}, \text{ onde}$$

N_h = tamanho do estrato h ;

N = tamanho do domínio;

st = número total de estratos;

s_h = desvio-padrão amostral para o estrato h ;

n_h = tamanho amostral correspondente ao estrato h .

Quando a amostra nos estratos é menor que 10% do tamanho total dos mesmos então o termo de correção torna-se desprezível e a equação se simplifica.

Uma forma prática de determinar o percentual do esforço amostral (SW) a ser alocado em cada estrato h quando o custo da amostragem é sempre o mesmo é aplicar a expressão:

$$SW_h = \frac{N_h \cdot s_h}{\sum_{h=1}^{st} N_h \cdot s_h}$$

Com respeito a esses instrumentos convém frisar, entretanto, aspectos muitas vezes um tanto desconsiderados:

a) o valor de s deve ser uma boa estimativa do desvio σ populacional;

b) a variável a ser estudada deve ser normalmente distribuída.

Avaliar a qualidade de s como estimador de σ é difícil, além disso, pelo menos quando se trata de efetivos das espécies, a condição da normalidade na maioria das vezes não se verifica (a dispersão dos indivíduos numa população raramente é normal). Assim, ao contrário das distribuições de Poisson (dispersão aleatória) e binomial negativa (dispersão agregada), a normal como descritor de dispersão é pouco interessante para o ecologista. Sua relevância vem do fato de que a maioria dos métodos, dentro do que se chama estatística paramétrica, pressupõe distribuição normal (simétrica, entre outras coisas) e variância com as propriedades de independência em relação à média e aditividade de seus componentes ou fontes.

GREEN (1979) apresentou uma interessante argumentação no que se refere à atenção que se deve dar à validade dos pressupostos dos testes estatísticos de significância. Segundo ele, os que usam estatística em estudos ambientais tendem a cair em um dentre dois extremos:

- ignorar a existência de pressupostos para os métodos em uso;
- super-valorizar os pressupostos, e cair quase sempre na

estatística não paramétrica.

Os métodos não paramétricos raramente são necessários, pelo menos antes de se conceber a amostragem. O mais comum é, após realizada a tomada de dados, eles passarem a ser a única alternativa. Pressupostos como normalidade e homogeneidade da variância (independência da média) quase sempre não se verificam completamente, mas quase sempre quase se verificam (HARRIS, 1975). Além disso, o fato de um certo pressuposto ter sido considerado ao se idealizar um teste não implica em que, não se verificando, tal teste se torne inútil. Aliás, às vezes conta-se com razoável robustez, mesmo em condições bastante desfavoráveis - as maiores exceções são valores de n muito pequenos ou desiguais (no caso de testes comparativos) e aplicação de testes monocaudais, bem mais exigentes. SOUTHWOOD (1966) alertou para o fato de que embora a análise de variância seja de certo modo mais robusta que o teste de χ^2 , dados com variância dependendo da média e em distribuições assimétricas não podem ser analisados sem riscos de erro. Por outro lado, segundo GREEN (1979), testes de significância para coeficientes de correlação são válidos para qualquer população unimodal de x e y com $n > 10$. Testes bicaudais como F e t são geralmente válidos mesmo em populações extremamente não normais.

No estabelecimento de intervalos de confiança, para distribuições ligeiramente assimétricas qualquer n maior que 2 serve (isso não é relacionado com a amplitude dos intervalos, e sim com sua validade); para as moderadamente assimétricas, maior que 10; para as fortemente assimétricas, maior que 40. Na verdade, a maior parte das análises estatísticas tendem a ser robustas face às violações dos pressupostos desde que o número de graus de liberdade seja grande, caso em que passa a valer o teorema do limite central. BARRETT & GOLDSMITH

(1976) fizeram estudos para determinar valores de n que produzem intervalos de confiança válidos, a despeito da não normalidade dos dados.

Um recurso para amenizar as violações é aplicar técnicas de transformação.

Primeiramente, durante o *screening* dos dados, deve-se reconhecer como a variância se relaciona com a média, seja determinando-se o tipo de dispersão dos dados, seja por meio da análise de regressão, com a média correspondendo à variável independente.

A melhor revisão sobre dispersão foi sem dúvida a de ELLIOTT (1977). Em termos gerais, quando não há nenhuma hipótese preliminar, a aderência à distribuição de Poisson, característica da dispersão aleatória, pode ser avaliada pelo teste de χ^2 :

$$\chi^2 = \frac{s^2 \cdot (n-1)}{\bar{x}}$$

Os valores críticos que delimitam o mínimo e o máximo da faixa de aceitação para uma dada confiabilidade $p = (1-\alpha)$ são dados pela tabela da distribuição de χ^2 com $(n-1)$ graus de liberdade e probabilidades respectivamente $\alpha/2$ e $1-\alpha/2$. Quando o valor calculado através da expressão se situa nesse intervalo, a hipótese de que a variância e a média sejam distintas deve ser rejeitada; abaixo daí, a média supera a variância; acima, a variância supera a média.

A probabilidade (p) de encontrar um certo número (esperado) de indivíduos (x) numa amostra de população com média \bar{x} e distribuição de Poisson, é dada por:

$$p(x) = e^{-\bar{x}} \cdot \frac{\bar{x}^x}{x!} \quad (\text{Southwood, 1966}).$$

A aderência, para um grupo de dados, pode então ser diretamente verificada pelo teste do χ^2 , comparando-se valores obtidos com os esperados.

Outro recurso é aplicar o índice de dispersão de Morisita (1959, 1962, 1964), cujo valor é dado por:

$$I_{\delta} = \frac{\sum_{i=1}^n x^2 - \sum_{i=1}^n x}{\left(\sum_{i=1}^n x \right)^2 - \sum_{i=1}^n x}$$

O índice é relativamente independente da distribuição, número de elementos e magnitude da média. Quando os dados são aleatórios, seu valor é 1; quando são uniformemente dispersos, o valor é menor que 1; quando agregados, ele supera 1. A significância do desvio em relação à distribuição de Poisson é dada pelo teste F , com valor crítico F_c referente a $n_1 = n - 1$ e $n_2 = \infty$, para dada confiabilidade. O valor de F é dado por:

$$F = \frac{I_{\delta} \left(\sum_{i=1}^n x - 1 \right) + n - \sum_{i=1}^n x}{n - 1}$$

Na distribuição de Poisson, a variância é igual à média; quando ela é menor, então a dispersão é mais regular (uniforme) do que seria esperado pela Poisson. Mais freqüentemente a variância supera a média e então a dispersão é contagiosa (termo derivado da epidemiologia, menos adequado em ecologia) ou agregada, aproximando-se sua distribuição da binomial negativa, descrita pelos parâmetros \bar{x} e k , este último um índice de agregação.

No caso de populações com dispersão de Poisson, a precisão das estimativas passa a depender exclusivamente do tamanho total da amostra, qualquer que seja o número de elementos constituintes da mesma. Em se tratando de números de indivíduos, por exemplo, o erro-padrão será o mesmo sempre que tiver sido alcançada uma certa contagem total, não importando se em poucos *quadrats* maiores ou vários menores. Pode-se escolher um tamanho conveniente para os *quadrats*, replicando-os até obter um total adequado (GREIG-SMITH, 1983).

Por outro lado, para médias muito baixas, a distribuição de Poisson é consideravelmente assimétrica (médias aproximadamente entre 0 e 9 unidades); a partir daí, ela aproxima-se bastante da normal, mostrando-se praticamente simétrica em torno da média. O pressuposto de independência desta em relação à variância, entretanto, permanece insatisfeito, já que persiste a igualdade. Para testar a significância das diferenças entre médias de distribuições de Poisson, pode-se usar as transformações:

$$z = \sqrt{x + 0,5}, \text{ para } n < 10, \text{ ou então}$$

$$z = \sqrt{x}, \text{ para } n \text{ maior.}$$

Essa questão será, todavia, tratada adiante de modo mais

Quando a dispersão não é aleatória, então não só a variância deixa de ser igual à média, como também mostra-se não proporcional a esta. Dos modelos matemáticos que se adaptam a essa situação, a distribuição binomial negativa é o mais comum. Nesse caso, a probabilidade esperada para certo valor de x numa amostra é dada por:

$$p(x) = \frac{\Gamma(k+x)}{x! \Gamma(k)} \left(\frac{\bar{x}}{\bar{x}+k} \right)^x \left(\frac{k}{k+\bar{x}} \right)^{-k}. \text{ Tal como se assina-}$$

lou para dispersão aleatória, mede-se a aderência aos valores obtidos usando o teste χ^2 (SOUTHWOOD, 1966).

ROJAS (1964) demonstrou que quando a dispersão da população é bem descrita pela distribuição binomial negativa, o número ideal de elementos é:

$$n = \frac{\frac{1}{\bar{x}} + \frac{1}{k}}{D^2}$$

Daí, BAKANOV (1985) propôs uma alternativa:

$$n = \frac{t_p^2 \bar{x}^2}{D^2} \left(\frac{1}{\bar{x}} + \frac{1}{k} \right), \text{ onde:}$$

n = número ideal de elementos amostrais;

t_p = valor da tabela de Student para confiabilidade p

\bar{x} = média amostral;

D = erro relativo aceito (coeficiente de variação).

k = parâmetro da distribuição binomial negativa, calculado segundo técnica iterativa descrita pelo autor.

O ponto crítico, com respeito às distribuições binomiais negativas é a determinação do parâmetro k . SOUTHWOOD (1966) discutiu algumas técnicas para isso; BAKANOV (1985) apresentou um nomograma para dimensionamento do número de elementos amostrais para médias e precisões diversas. As fórmulas de Southwood são:

$$a) \quad k = \frac{\bar{x}^2}{s^2 - \bar{x}} \text{ (método dos momentos - pouco eficaz com médias}$$

pequenas);

$$b) \log\left(\frac{n}{n_0}\right) = k \cdot \log\left(1 + \frac{\bar{x}}{k}\right), \text{ onde } n_0 \text{ é o número de valores nulos}$$

(mais adequado justamente quando as médias são pequenas);

$$c) n \cdot \ln\left(1 + \frac{\bar{x}}{k}\right) = \sum_{i=1}^n \left(\frac{A_{x_i}}{k - x_i}\right), \text{ onde } A_{x_i} \text{ é o número de valores}$$

maiores que x_i

(método considerado o melhor, embora já se tenha demonstrado que mesmo essa boa estimativa de k é sujeita a desvios quando a média é pequena e k é grande). Neste último caso, procede-se de modo iterativo, procurando qual o valor de k que valida a igualdade; quando o valor inicial é muito alto, o membro esquerdo supera o direito, e *vice-versa*. Esta forma é a mais similar à apresentada por BAKANOV (1985).

Em termos gerais, a relação entre média e variância pode ser descrita por uma função potencial conhecida como lei potencial de Taylor (TAYLOR, 1961), válida para dispersões desde uniformes até fortemente agregadas:

$$s^2 = a\bar{x}^b.$$

A série de médias e respectivas variâncias a partir da qual, por análise de regressão, determinam-se a e b , deve ser obtida de lotes de elementos de tamanhos originalmente distintos, ou então compostos por simulação de modo a constituir unidades de tamanho variado a fim de produzir médias de diferentes magnitudes. A expressão potencial de Taylor lineariza-se por transformação log-log, ficando a função linear expressa por:

$$\log s^2 = \log a + b \log \bar{x}.$$

Já que pelo menos a análise de variância só é invalidada em casos extremos de assimetria (não normalidade), o pressuposto mais delicado é sem dúvida a independência entre média e variância, o qual uma vez satisfeito também assegura aquele requisito de aditividade. Em se tratando de intervalos de confiança, quando $n \geq 30$, a não normalidade pode ser ignorada; caso n seja menor, não se provando discordância com a distribuição de Poisson, pode-se obter os intervalos a partir de tabelas; o maior problema vem no caso de poucos valores e dispersões agregadas. Aí, o parâmetro b é um índice de agregação muito útil para estimar o tamanho da amostra e obter a adequada transformação de dados capaz de eliminar ou quase eliminar a assimetria da distribuição, tornando s^2 e \bar{x} independentes:

$$z = x^n, \text{ onde } y = 1 - \frac{1}{2}b \text{ (MCINTYRE et al., 1984).}$$

Como essa transformação é específica de cada amostra, complicam-se as comparações entre amostras cujos dados foram diferentemente transformados. Assim sendo, o ideal é apenas atenuar a violação dos pressupostos seguindo como regras gerais (SOUTHWOOD, 1966):

dispersão uniforme: $z = x^2$;

dispersão levemente agregada: $z = \sqrt{x}$;

dispersão fortemente agregada: $z = \log(x+1)$.

Quando a lei de Taylor é válida, se $b=0$ então a variância pode ser considerada independente da média, daí a possibilidade de testar tal independência com base na horizontalidade da função linearizada. MCINTYRE *et al.* (1984) apresentaram algoritmos baseados na lei de Taylor para estimativa do tamanho amostral nos casos de distribuições não normais:

$$n = t_p^2 \alpha \bar{x}^{(b-2)} D^{-2}, \text{ ou } \log n = 2 \log t_p + \log \alpha + (b-2) \log \bar{x} + \frac{1}{2 \log D}.$$

GREEN (1979) avançou com a lei potencial de Taylor, obtendo $D^2 = a n^{1-b} \cdot A^{b-2}$, que se lineariza como:

$$\log A = \left[\frac{\log \left(\frac{D^2}{a} \right)}{b-2} \right] + \left(\frac{b-1}{b-2} \right) \log n, \text{ onde } A \text{ é o efetivo}$$

total acumulado (por exemplo o número total de indivíduos já incluído no conjunto dos elementos).

Se a população se distribui conforme a série geométrica, $b = 2$; quando $b = 1$ então trata-se da distribuição de Poisson. Num caso prático, resolvida a expressão linear, define-se uma reta chamada *reta de decisão* ou **stop-line**, correspondendo a valores de A e n que satisfazem a precisão desejada D . Se a representação de A contra n , ambas em escala logarítmica, constituir uma reta horizontal então conclui-se que a amostragem simplesmente deve prosseguir até que se atinja determinado efetivo total A , fixo, a despeito do número de elementos amostrais incluídos na amostra (dispersão aleatória: $b = 1$). Caso a reta seja vertical (dispersão fortemente agregada: $b = 2$), apenas deve-se acumular determinado número n fixo de elementos, independentemente da magnitude da média. Finalmente, para os casos intermediários ($1 < b < 2$) obtém-se uma reta mais ou menos inclinada.

Estratégias Espectrais

As técnicas descritas na seção anterior, como ficou bastante claro, exigem um número considerável de valores a serem obtidos através da amostragem preliminar, a qual, então, se torna particularmente laboriosa. Além disso, prestam-se mais à determinação de atributos ao nível de população, basicamente densidades populacionais, que em primeira instância correspondem a efetivos médios por elemento amostral, podendo ser os valores convertidos para unidades de área, preferivelmente menores que a área dos elementos. De qualquer modo, parece claro que a base é o elemento amostral. Nos casos em que se vai tratar de descritores sintéticos, ao nível de comunidade, as mesmas técnicas podem ser aplicadas, desde que se queira estimar médias também baseadas na área dos elementos amostrais. Esse é um ponto delicado, pois a maioria dos descritores sintéticos, por exemplo os de dominância e diversidade, é mais ou menos fortemente influenciada pela área de referência.

Os métodos espectrais são uma alternativa para contornar tal complicador, servindo para dimensionar amostras em estudos tanto de populações como de comunidades.

PIELOU (1974), tratando da diversidade de dominância, expressa pelo índice de Shannon (H'), abordou essa questão; suas observações são certamente válidas, pelo menos em princípio, no caso de outros descritores. Em primeiro lugar, constituir um *pool* de todos os elementos amostrais e então sobre ele calcular o valor do índice, no mínimo impossibilita a determinação do erro-padrão da estimativa e portanto as comparações estatísticas. Em segundo lugar, é obviamente inadequado determinar H' para cada elemento amostral e então calcular a média, já que os elementos individualmente apresentam um conteúdo informativo menor que o todo da amostra e assim a diversidade desta fica subestimada. Isso é especialmente verdadeiro em comunidades faceteadas com mosaico grosseiro, onde cada elemento não inclui mais que pequena parte da estrutura e fração do número total de espécies.

O principal objetivo de se expressar a diversidade através de índices é viabilizar comparações entre comunidades distintas. Não obstante, a dependência dos mesmos em relação à área impõe dificuldades para as próprias comparações. Uma maneira de resolver esse problema da dependência, e portanto da desigualdade entre os tamanhos das amostras que se quer comparar, foi proposta por SANDERS (1960) com o nome de *método da rarefação*. O método de Sanders corresponde à seleção aleatória de indivíduos em meio à comunidade, contando o número total de espécies representadas no conjunto dos indivíduos já escolhidos. A curva obtida, referente à riqueza em função do efetivo acumulado é chamada curva de rarefação.

HURLBERT (1971) mostrou que o método de Sanders geralmente superestima a riqueza S , introduzindo seu estimador para cada tamanho n de amostra em populações conhecidas de tamanho N :

$$\hat{S}_n = \sum_{i=1}^s \left\{ 1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right\}, \text{ onde}$$

n = tamanho da amostra (efetivo total na mesma);

N = tamanho da população;

N_i = efetivo da espécie i na população.

A expressão de Hurlbert é especialmente interessante, pois a partir uma dada amostra (considerada como população) permite retomar os dados e obter o resultado em forma espectral, com tamanhos amostrais crescentes. O problema é que métodos como esse e o da rarefação desconsideram totalmente a heterogeneidade devida à repartição espacial dentro da própria amostra e portanto qualquer variabilidade calculada a partir dos mesmos quase que certamente será uma subestimação.

Outra forma de estimar a contribuição do $k^{\text{ésimo}}$ elemento amostral em termos do número médio de espécies foi apresentada por GAUFIN *et al* (1956), sendo a variância correspondente estimada segundo HARRIS (1957).

A forma espectral ou complexa dos descritores vem de encontro à noção de *área mínima*, há muito propalada porém discutível. Através dos espectros pode-se determinar o tamanho amostral além do qual o descritor deixa de variar significativamente.

Além disso, a forma dos espectros ou curvas riqueza-área, diversidade de Shannon-área, e mesmo variabilidade-área, pode ser excelente descritor das comunidades (BALLESTEROS, 1986). Sob esse enfoque, a utilidade dos espectros transcende o dimensionamento amostral.

No caso dos levantamentos faunísticos/florísticos, o tamanho ideal da amostra é aquele que inclui um determinado percentual do total de espécies do sistema estudado. Para determiná-lo pode-se partir da relação espécies-área válida em cada situação específica.

Diversos autores procuraram estabelecer modelos matemáticos como descritores da relação espécies *versus* área ou riqueza *versus* número de elementos amostrais. Entre as funções mais comuns estão:

a) função potencial: $S_{(n)} = an^b$;

b) função exponencial: $S_{(n)} = ab^n$;

c) função, baseada no modelo de Arrhenius: $S_{(n)} = \sum_{i=1}^S \left[1 - \left(1 - \frac{n_i}{N} \right)^n \right]$.

A esta última pode associar-se um fator de correção para compensar o erro introduzido por espécies com frequência muito baixa na amostra ($< 0,1$):

$$S_{(n)} = \sum_{i=1}^S \left[1 - \left(1 - \frac{n_i}{N} \right)^n \right] \cdot \left\{ \frac{S_0}{1 - \left(1 - \frac{n_i}{N} \right)^n} - S_0 \right\}, \text{ onde}$$

$S_{(n)}$ = riqueza esperada numa área correspondendo a n elementos;

n_i = número de elementos incluindo a espécie i ;

N = número total de elementos na amostra;

S_0 = número de espécies raras (frequência relativa nos elementos menor que 0,1).

No caso da função potencial, o parâmetro a estima a riqueza na área de um elemento amostral (área unitária). O expoente b , por seu lado, independe do tamanho dos elementos, constituindo-se num atributo da comunidade, inclusive mantendo-se entre 0,25 e 0,35 mesmo em situações bem diversas (KILBURN, 1966).

Uma função exponencial relacionando o efetivo total da amostra (A) com o número de espécies (S) serviu de base para Gleason e Margalef estabelecerem seus índices de diversidade, correspondendo ambos à inclinação das curvas linearizadas por transformação semi-log aplicada respectivamente às expressões:

$$A = ab^S, \text{ e } A = ab^{S-1}.$$

KILBURN (1966) mostrou que a relação potencial, que se lineariza por transformação log-log, é melhor que a exponencial. Todavia, esta última é mais indicada quando as comunidades apresentam abundâncias distribuídas segundo a série logarítmica:

$$S_{(n)} = \alpha \cdot \log \left(1 + \frac{n}{\alpha} \right).$$

HOLGATE (1969) demonstrou, por sua vez, que nesse caso S aumenta com log n sem estabilizar. Ao mesmo tempo, S_1/S (proporção de espécies

com um só indivíduo, ou efetivo unitário) aumenta até estabilizar. Se a distribuição de abundâncias for binomial negativa, a taxa de incremento de S gradualmente diminui, até que a riqueza amostral se aproxima assintoticamente da riqueza populacional. Paralelamente S_I/S aumenta, atinge um máximo, e então cai tendendo a 0.

Quanto à diversidade de dominância (H'), via-de-regra aplicam-se as funções potencial ou exponencial ou, muito mais frequentemente, a de

Michaelis-Menten:
$$H' = \frac{an}{b+n}$$

Essa função, uma vez linearizada fica sendo
$$\frac{1}{H'} = \frac{1}{a} + \frac{b}{a} \cdot \frac{1}{n}$$

Aí, $1/a$ é um estimador da diversidade global da comunidade (valor assintótico, correspondendo à área infinita); b/a , por sua vez é a área na qual já se verifica metade da diversidade global, tanto maior quanto mais diagonal for a curva.

O problema é que as relações previstas nos modelos citados não se verificam de modo generalizado, o que os torna apenas relativamente úteis. Em alguns casos os valores aumentam continuamente com o incremento da área amostral (espectro diagonal), devido à adição de espécies não cada vez mais raras, e sim com proporções flutuantes. Isso caracteriza sistemas mais estruturados, com *patchiness* mais acentuado. Quando o grau de estruturação é menor, apresentando-se os componentes muito misturados e o sistema homogêneo, o espectro tende a ser retangular, crescendo a curva a princípio, mas estabilizando logo num patamar. A função potencial descreve bem as relações entre riqueza e diversidade de dominância e área nos sistemas mais afins do primeiro padrão. A função de Michaelis-Menten, por sua vez, é mais adequada quando o padrão é assintótico (BALLESTEROS, 1986).

A estratégia mais comum para utilização dos métodos espectrais de dimensionamento amostral baseia-se inicialmente na acumulação de área (aumento no número de elementos). Há entretanto numerosas maneiras de proceder tal incremento, algumas propiciando a determinação da variabilidade, e outras não. O ponto principal é se as áreas crescentes são independentes uma da outra, ou não - no caso de serem imbricadas, há dependência.

No que se refere à riqueza S , só se obtém as verdadeiras curvas espécies-área quando as áreas crescentes são independentes; quando não o são, trata-se das chamadas *curvas de coletores* (*collector's curves*) (PIELOU 1977).

PICHON (1978), por exemplo, propôs coletar amostras independentes, de tamanhos crescentes (superfície ou volume), contando nelas o número de espécies (analogamente pode-se calcular qualquer outro descritor sintético). Coletando um número conveniente de réplicas para cada dimensão, calcula-se

as variabilidades associadas aos valores médios. No caso da riqueza, o autor sugeriu eliminar todas as numerosas espécies representadas por um só indivíduo, por considerá-las de pequena importância ecológica. Segundo o autor, esse artifício é especialmente conveniente por propiciar estabilização de curvas não assintóticas.

A partir de uma área sistematicamente amostrada por um reticulado de *quadrats*, preparam-se espectros de dois tipos (PIELOU, 1977):

- por acumulação de novos elementos selecionados ao acaso no reticulado (nesse caso os componentes da área total não são obrigatoriamente contíguos);

- por acumulação de elementos segundo um padrão centrífugo definido: parte-se do centro ou de outro ponto do reticulado, incorporando novos elementos de modo a aumentar o raio (áreas imbricadas).

Basicamente podem ser obtidos dois tipos de resultados:

- o espectro se estabiliza;
- o espectro é sempre crescente

Combinando-se os modos de acumulação e os resultados, reconhecem-se 4 variantes:

a) estabilização tanto na acumulação aleatória quanto na imbricada - a amostra é suficiente para estimar a riqueza da comunidade; o povoamento é homogêneo e, para ele, o modelo da série logarítmica de Fisher não é apropriado;

b) estabilização apenas na acumulação aleatória - a riqueza da comunidade pode ser estimada, porém o povoamento não é homogêneo; assim, não se justifica o ajuste de algum modelo de abundâncias aos dados;

c) estabilização apenas na acumulação imbricada - resultado considerado impossível pela autora;

d) espectro sempre crescente tanto na acumulação aleatória quanto na imbricada - a riqueza da comunidade não pode ser estimada; é impossível decidir sobre a característica do povoamento, podendo tratar-se de um sistema heterogêneo ou então homogêneo distribuído segundo a série logarítmica de Fisher.

BORJA (1986) propôs constituir, a partir de um reticulado de 4×4 *quadrats*, todas as combinações de elementos contíguos de tamanhos $n \times n$, ($n \times n-1$, $n \times (n-1)$) e $[n \times (n-1)] - 1$, com n variando de 1 a 4. Os elementos são nesse caso parcialmente imbricados, já que se verifica considerável superposição.

BALLESTEROS (1986) acumulou área segundo 3 padrões:

- a) sub-amostras imbricadas de tamanho crescente;
- b) não imbricadas de tamanho crescente, e

c) sub-amostras compostas segundo a técnica de BORJA (1986), muito mais numerosas para um mesmo reticulado inicial, porém com a vantagem de, permitindo diversas combinações, viabilizar o cálculo da variância associada a cada sub-área.

A estratégia dos elementos imbricados não é totalmente correta, dada a possibilidade de distorção introduzida pelo aparecimento de espécies raras já nas sub-áreas iniciais. NIELL (1977) já havia lembrado que dada a agregação característica de muitas populações, caso já se inclua no primeiro elemento uma espécie rara ou agregado denso pouco frequente, isso se reflete em todo o restante da curva. Por outro lado, a acumulação de área segundo um padrão imbricado, seja ele centrífugo, seja ao longo de um ou mais gradientes (amostra do tipo *transecção*), permite apresentar a chamada γ -diversidade (PIELOU, 1977). SANTELICES (1980) evidenciou dessa forma uma série de heterogeneidades espaciais.

BOUDOURESQUE (1974) montou curvas baseadas numa única série de áreas imbricadas. Outros, como NIELL (1977) e FERNANDEZ & NIELL (1981), procederam a ordenações distintas, a fim de avaliar aproximadamente a variabilidade inerente aos diversos pontos das curvas.

Além da incorporação de *quadrats* contíguos aos previamente considerados (acumulação imbricada) e constituição de sub-áreas contínuas parcialmente imbricadas, técnicas que levam à determinação do *tamanho de um quadrat representativo do sistema estudado*, pode-se incrementar área simplesmente constituindo grupos de elementos não obrigatoriamente contíguos (replicados ou não; com ou sem reposição), selecionados aleatoriamente em meio a um reticulado inicial, caso em que se determina o *número de quadrats de tamanho pré-estabelecido que devem integrar uma amostra representativa* (GOODALL, 1952; GREIG-SMITH, 1983). NIELL (1977), tendo acumulado área deste modo, com réplicas, e também de modo imbricado, constatou que a seleção aleatória produziu áreas mínimas maiores, com ampla variabilidade. Isso se justifica pelo fato de aumentar a distância média entre os elementos constituintes dos grupos e portanto a heterogeneidade entre mesmos. MCINTYRE *et al.*, (1984) indicaram esta técnica quando é necessário testar a significância das diferenças entre amostras ou determinar o padrão de dispersão das espécies.

LIVINGSTONE *et al.* (1976) descreveram uma forma peculiar de obter o espectro espécies-área. Aparentemente as ocorrências das espécies foram randomizadas ao longo de séries estabelecidas de 16 ou 40 elementos amostrais, determinando-se por computador, e para cada configuração, o número de elementos acumulado quando a riqueza *S* atingia 90% do total da amostra. Por exemplo, se um grande número de espécies casualmente se fizesse incluído já nas primeiras posições, o número acumulado de elementos ao ser atingida a riqueza crítica seria menor. Os autores tomaram a média dos valores das diversas configurações (séries) como número ideal de elementos amostrais para incluir 90% da riqueza total. Esse esquema foi aplicado para valores crescentes de riqueza entre 1 e *S* representando-se as médias resultantes ($\pm \sigma$) em espectros. Essa estratégia, embora parecendo similar à da constituição de grupos com elementos selecionados aleatoriamente, difere pelo fato de que em princípio o primeiro elemento de uma série já pode

conter todas as espécies da amostra. Parece ser um pressuposto da técnica de LIVINGSTONE *et al.* (1976) a ausência de associação entre espécies, isto é, suas ocorrências não podem ser correlacionadas. Como isso não ocorre na prática (pelo menos de modo generalizado), não se justifica sortear suas ocorrências, uma independentemente da outra.

PIELOU (1966, 1974) apresentou sua técnica, especificamente concebida para constituir espectros da diversidade de dominância de Shannon (podendo ser adaptada para o caso de outros descritores) e, a partir deles, determinar um *ponto de corte* adequado. LLOYD *et al.* (1968) aplicaram-na no estudo de uma comunidade de anfíbios e répteis.

Inicialmente coloca-se os elementos em seqüência aleatória. Seja H_1 a diversidade do elemento 1 da série, calculada pelo índice de Brillouin (e portanto em unidades de informação por indivíduo). Seja, agora, H_2 a diversidade dos elementos 1 e 2 fundidos. Calcule-se sucessivamente H_3, H_4, \dots , até H_n . A representação desses valores em função do número de elementos constitui uma curva que cresce rapidamente a princípio, estabilizando-se depois. Suponha-se então que depois de terem sido acumulados n_0 elementos a curva se apresente horizontal (a não ser por pequenas irregularidades), formando um platô. Nessas condições, para qualquer $n \geq n_0$, pode-se calcular a diversidade de Brillouin usando a expressão:

$$\bar{H} = \frac{(A_{n+1} \cdot H_{n+1}) - (A_n \cdot H_n)}{A_{n+1} - A_n}, \text{ onde } A_n = \text{efetivo total nos } n$$

elementos já acumulados.

Se $H_{n+1} \cong H_n = H_0$, o que se tem é:

$$H = \frac{H_0 \cdot (A_{n+1} - A_n)}{A_{n+1} - A_n} = H_0, \text{ ou seja, } H_0 \text{ é o valor estabilizado de } H.$$

A estimativa de H' para o domínio inteiro é a média dos valores observados de H referentes a $n \geq n_0$.

LLOYD *et al.* (1968) notaram que, embora os sucessivos valores de H não sejam independentes um do outro, os incrementos A_n, \bar{H}_n são. Assim, as variações que se verificam após a "estabilização" devem ser consideradas aleatórias. A variância da estimativa de H' para n_q ($n_q = n - n_0$) *quadrats* adicionados após a estabilização é:

$$\text{var}_{(\hat{H}')} = \frac{1}{n_q \cdot (n_q - 1)} \left[\sum_{i=1}^{n_q} H_i^2 - \frac{\left(\sum_{i=1}^{n_q} H_i \right)^2}{n_q} \right]$$

O intervalo de confiança, segundo a autora, então é dado por

$$i.c = \hat{H}' \pm t_p \cdot \sqrt{\text{var}_{(\hat{H}')}}.$$

PIELOU (1974) completou sua abordagem definindo que seu índice de equitatividade ou diversidade relativa então pode ser estimado por

$$\hat{J}' = \frac{\hat{H}'}{\log S}, \text{ e a respectiva variância, por } \text{var}_{(\hat{J}')} = \frac{\text{var}_{(\hat{H}')}}{(\log S)^2}.$$

Nesses casos, S é a riqueza populacional, que precisa ser conhecida.

Uma segunda etapa no dimensionamento pelo método espectral envolve a determinação de um *ponto de corte*, o qual define a discutível área mínima.

Foi BOUDOURESQUE (1971) que cunhou os termos *área mínima específica* ou *qualitativa* e *área mínima estrutural* ou *quantitativa*. Na medida em que a área mínima, seja ela qual for, representa o sistema do qual é parte, pode refletir complexidade estrutural: as áreas mínimas tendem a aumentar em povoamentos mais ricos e intrincados. Nos mais homogêneos, via-de-regra surge a dominância e então a complexidade se reduz juntamente com a diminuição da repartição (NIELL, 1977; SIERRA & FERNANDEZ, 1984). A área mínima qualitativa se baseia em espectros espécies-área, tendo sido usada primeiro em estudos de fitossociologia em sistemas terrestres (CAIN & CASTRO, 1959); mais tarde passou a ser determinada também em estudos do fitobentos (BOUDOURESQUE, 1971, 1974).

HAWKINS & HARTNOLL (1980) estudaram a relação espécies-área, concluindo pela inadequação do próprio conceito de área mínima como unidade capaz de incluir todas as espécies de um sistema. Afinal, em muitos casos, o incremento da riqueza S em função da área não se mostra assintótico ou tendendo à estabilização. FUENTES & NIELL (1981) criticaram as afirmações de HAWKINS & HARTNOLL (1980), lembrando primeiro que elas não se aplicariam à área mínima estrutural. BALLESTEROS (1986), todavia, concordou com a idéia propugnada por HAWKINS & HARTNOLL (1980), de que cada pesquisador deve eleger seu próprio critério de estabilização. Por outro

lado, é necessário admitir que devem ser assegurados subsídios que permitam alterar o critério *a posteriori*, a fim de viabilizar comparações entre resultados obtidos por diferentes pesquisadores. Tais critérios para estabelecimento dos pontos de corte nos espectros dos descritores em função da área podem ser agrupados em diversas categorias:

a) reconhecimento simplesmente visual ou gráfico do ponto de estabilização dos valores do descritor em si ou de sua variabilidade;

b) estabelecimento de um critério arbitrário, tal como um valor crítico da relação custo/benefício, incluindo-se aí os diversos pontos Molinier, Calleja, Vestal, etc.;

c) estabelecimento de um limite crítico máximo para a variabilidade dos valores do descritor, adotando-se como área amostral aquela na qual ela é inferior ao estabelecido;

d) teste estatístico de significância da independência do descritor em relação à área a partir de um ponto estabelecido *a priori*.

PICHON (1978) afirmou que a curva ou espectro deve apresentar um *ponto singular de estabilização*, sugerindo ser este conspícuo e de fácil reconhecimento visual. NIELL (1974, 1977) e FERNANDEZ & NIELL (1981) também reconheceram o ponto de estabilização diretamente sobre a curva. Uma forma menos subjetiva é aquela onde se fixa uma determinada relação custo/benefício. No caso dos chamados pontos MOLINIER e do critério de CAIN & CASTRO (1959), tanto o custo quanto o benefício, respectivamente correspondendo ao incremento de área e à variação no valor do descritor considerado, se expressam em termos relativos. A notação ponto Molinier 20:2, por exemplo, corresponde a não obter diferença percentual maior que 2% após incrementar a área amostral em 20%. Quando os dados originais, normalmente apresentando certas instabilidades, são ajustados à função potencial, é impossível determinar o ponto Molinier a partir dos parâmetros da equação; no caso de exponenciais e da função de Michaelis-Menten, não há problema. BORJA (1986) apresentou todos os seus espectros como curvas superpostas índice-área e variabilidade relativa-área. O corte foi feito segundo os pontos Molinier 20:1 e 20:2, além do critério 10%:10% de CAIN & CASTRO (1959). FUENTES & NIELL (1981) expressaram a diversidade em função da área, associando a cada valor médio, o desvio correspondente. Apresentaram também os espectros do desvio-padrão e coeficiente de variação. O corte das curvas também se fez segundo o critério de CAIN & CASTRO (1959).

BALLESTEROS (1986) apresentou uma abordagem das mais interessantes quanto aos critérios de corte em espectros formados a partir do ajuste dos dados às funções exponencial e de Michaelis-Menten. Discutiu então o que chamou método do *K*, aplicável a espectros exponenciais (basicamente espécies-área), sobre os quais se quer determinar pontos Molinier.

O parâmetro K vem da forma exponencial com que pode ser descrito o espectro linearizado por transformação semi-log ($I = a \cdot \ln A + b$, onde I é o valor do índice, correspondente à área A):

$$A = K \cdot e^{\lambda I}, \text{ onde}$$

$$K = e^{-\frac{b}{a}}, \text{ e } \lambda = \frac{1}{a}.$$

Esse valor de K , não só serve como descritor da forma da curva, como também é uma constante direta para determinar pontos já estabelecidos:

- ponto Molinier 20:1 = 82805082 K ;
- ponto Molinier 20:2 = 9100 K ;
- ponto Molinier 20:4 = 95,4 K ;
- ponto Molinier 20:5 = 38,3 K .

Valores elevados de K implicam em áreas específicas maiores, por exemplo devido à presença de espécies raras e esparsas; valores pequenos correspondem a áreas menores, normalmente em consequência do fato de as espécies se acharem mais concentradas ou bem representadas (pequeno número de espécies pouco frequentes na escala de amostragem). O parâmetro K é então uma forma elegante de descrever a comunidade sob o ponto de vista qualitativo.

Os pontos Calleja, por sua vez, correspondem à expressão do custo e do benefício em termos absolutos. BALLESTEROS (1986) usou o número de espécies representado em certo ponto Calleja (1:0,05) como índice de riqueza. No caso de espectros de diversidade de Shannon, o critério de corte adotado pelo autor correspondeu a um acréscimo de não mais que 0,001 no índice, como resultado da adição de um novo elemento amostral sobre a área já acumulada. A esse ponto, mais exatamente a essa área, chamou S , atribuindo-lhe grande valor como descritor da forma da curva de Michaelis-Menten, função assintótica à qual esse tipo de espectro via-de-regra se ajusta muito bem.

HAWKINS & HARTNOLL (1980) criaram seu próprio critério para corte das curvas, correspondendo a uma relação custo/benefício expressa em termos de um incremento absoluto de área ($0,25\text{m}^2$) ao qual se associa um determinado acréscimo relativo no número acumulado de espécies.

MIYARES & ANADÓN (1981) também representaram índices em função da área, com espectros montados a partir de um reticulado inicial de 8 x 6 elementos de 1 dm^2 . Definiram áreas mínimas para determinação de efetivos específicos empregando a variabilidade relativa dos valores em diferentes áreas. SIERRA & FERNANDEZ (1984) expressaram riqueza e diversidade de

dominância em termos de valores médios por sub-área, juntamente com os respectivos coeficientes de variação, considerando satisfatório, para estes, um nível máximo de 10%.

A análise conjunta dos espectros de médias e de variabilidades, sejam estas expressas em termos de erros-padrão, intervalos de confiança ou simplesmente desvios-padrão, freqüentemente em valores relativos, é sem dúvida mais indicada que a simplesmente baseada em médias (e isso é especialmente pertinente quando se apresenta curvas ajustadas aos dados médios originais, sobre cuja variabilidade deixa de se ter informação clara). O maior cuidado a tomar, no entanto, é com a superposição entre elementos quando a acumulação é imbricada, parcialmente imbricada ou então aleatória com reposição. Nesses casos, como as sub-amostras vão se constituindo com tamanho cada vez mais próximo do total do reticulado inicial, a variabilidade vai se reduzindo progressivamente e de modo artificioso, até tornar-se nula. Com certeza muitas das variabilidades associadas às maiores sub-áreas são fortemente subestimadas. A seleção de elementos de modo aleatório e sem reposição não garante a independência entre os mesmos e portanto não se pode a rigor utilizar os grupos de mesma área como réplicas, além do que o número de variantes é drasticamente diminuído. O ideal seria selecionar elementos constituindo grupos de áreas bem menores que a área total, minimizando a superposição e assegurando certa independência entre eles. Isso, por outro lado, tornaria a amostragem preliminar excessivamente laboriosa. Uma forma de contornar esse declínio artificioso da variabilidade, é observar não um limite crítico, e sim o ponto a partir do qual o espectro correspondente declina de modo uniforme, com declividade mais ou menos constante, ou então onde se dá o cessamento das oscilações.

ROSSO (1991b) apresentou *softwares* apropriados não só para preparar os espectros de médias e outras estatísticas de diversos descritores sintéticos em função da área (aplicativo MINAREA), como também determinar *pontos de corte* segundo relações custo/benefício fixadas (em termos relativos ou absolutos) pelo pesquisador, partindo dos parâmetros da função que, dentre várias, melhor tiver se ajustado aos dados (aplicativo MOLINIER). Esses instrumentos vêm sendo utilizados numa série de trabalhos descritivos, mostrando-se muito úteis (BALDRESCA & ROSSO, 1991; BORGES & ROSSO, 1991; ROSSO, 1991a,b; BORGES & ROSSO, 1992; ROSSO *et al.*, 1992; ROSSO & MENDES, 1992).

Para PIELOU (1966, 1974), com respeito à sua técnica para estimar H' , o melhor modo de julgar se houve estabilização é fazer um reconhecimento preliminar do *ponto de corte*, testando se a partir daí os n valores de H estão serialmente correlacionados (auto-correlacionados). Inicialmente estabelece-se o passo da auto-correlação (p), constituindo-se 2 vetores, o primeiro incluindo valores de H_1 a H_{n-p} e outro, de H_{p+1} a H_n . A seguir, calcula-se:

$$F_1 = \sum_{i=1}^{n-p} H_i \quad /, \quad F_2 = \sum_{i=1}^{n-p} H_{p+i} \quad \text{e} \quad F_3 = \sum_{i=1}^{n-p} H_i \cdot H_{p+i}.$$

A covariância dos dois vetores é $C = \frac{1}{(n-p)} \cdot F_3 - \frac{1}{(n-p)^2} \cdot F_1 \cdot F_2$.

$$\text{A variância da seqüência toda é } V = \frac{\sum_{i=1}^n H_i^2}{n} - \frac{\left(\sum_{i=1}^n H_i \right)^2}{n^2}.$$

O coeficiente de auto-correlação para o passo p é $r_p = \frac{C}{V}$.

Finalmente determina-se a significância dessa auto-correlação pelos métodos usuais.

A técnica da auto-correlação é válida na medida em que os valores após a estabilização apresentem tão somente variações aleatórias em domínio homogêneo. Caso este seja heterogêneo, então haverá oscilações não randômicas dos valores de H_i , afetando a significância real. A própria seqüência de acumulação pode interferir na estimativa de H' . Assim, é recomendável tratar diferentes seqüências aleatórias, tomando-se como estimativa final de H' a mediana das diversas estimativas calculadas (tal como fizeram LLOYD *et al.*, 1968).

Bibliografia

- BAKANOV, A.I. 1985 A nomogram for estimating the required number of samples of an aggregated distribution. Hydrobiol. J. 6:85-8.
- BALDRESCA, R.; ROSSO, S. 1991 Estudo da estrutura de um costão rochoso intermareal da estação ecológica da Juréia (SP). In: Simpósio sobre Oceanografia, 2. São Paulo, 21-25 Out. 1991. IOUSP Resumos. São Paulo, p.2
- BALLESTEROS, E. 1986 Métodos de análisis estructural en comunidades naturales, en particular del fitobentos. Oecol. Aquat. 8:117-31.
- BARRETT, J.P.; GOLDSMITH, L. 1976 When is n sufficiently large? Am. Stat. 30:67-70.

- BORGES, R.P.; ROSSO, S. 1991 Dimensionamento amostral e avaliação preliminar da estrutura de um costão rochoso da Praia da Tatuira. São Sebastião, São Paulo. In: Simpósio sobre Oceanografia, 2, São Paulo, 21-25 Out. 1991. IOUSP Resumos. São Paulo, p. 296.
- _____. 1992 Estudo da área amostral em um costão de São Sebastião (SP). In: Reunião anual da Sociedade Brasileira para o Progresso da Ciência, 44, São Paulo, 12-17 Jul. 1992. Anais. SBPC São Paulo, p 65
- BORJA, A. 1986 Estudio del área mínima de muestreo en una población intermareal de pequeños moluscos. Invest. Pesq., 50(1):5-22.
- BOUDOURESQUE, C.F. 1971 Méthodes d'étude qualitative et quantitative du benthos (en particulier du phytobenthos). Tethys, 3(1):79-104.
- _____. 1974 Aire minima et peuplements algaux marins. Soc. Phycol. Fr. Bull., 19:141-57.
- BUZAS, M.A. 1970 Spacial homogeneity: statistical analysis of unispecies and multispecies populations of foraminifera. Ecology, 51:874-9.
- CAIN, S.A.; CASTRO, G.M. 1959 Manual of vegetation analysis. New York, Harper Ed. 325p.
- ELLIOTT, J.M. 1977 Some methods for the statistical analysis of benthic invertebrates. 2nd ed., (Scientific Publications, 25) Freshwater Biological Association. 156p.
- FERNANDEZ, C.; NIELL, F.X. 1981 Distribución espacial del fitobentos en los horizontes inferiores del sistema intermareal rocoso de Cabo Peñas (Asturias). invest. Pesq., 45(2):309-26.
- FRONTIER, S. 1983 Stratégies d'échantillonnage en écologie. Paris, Masson. 494p.
- FUENTES, J.M.; NIELL, F.X. 1981 Spatial structure of a mid-level intertidal community: some comments on sampling. Bot. Mar., 24:135-8.
- GAUFIN, A.R.; HARRIS, E.K.; WALTER, H.J. 1956 A statistical evaluation of stream bottom sampling data obtained from three standard samples. Ecology, 37:643-8.
- GLEASON, H.A. 1922 On the relation between species and area. Ecology, 3:158-62.
- GOODALL, D.W. 1952 Quantitative aspects of plant distribution. Biol. Rev., 27:194-245.
- GOUNOT, M. 1969 Méthodes d'étude quantitative de la végétation. Paris, Masson. 314p.

- GREEN, R.H. 1979 Sampling design and statistical methods for environmental biologists. New York, Wiley-Interscience. 257p.
- GREIG-SMITH, P. 1983 Quantitative plant ecology. 3rd ed. (Studies in ecology, 9). Oxford, Blackwell. 359p.
- HARRIS, E.K. 1957 Further results in the statistical analysis of stream sampling. Ecology, 38:463-8.
- HARRIS, R.J. 1975 A primer of multivariate statistics. New York, Academic Press.
- HAWKINS, S.J.; HARTNOLL, R.G. 1980 A study of the small-scale relationship between species number and area on a rocky shore. Estuarine Coastal Mar. Sci., 10:201-14.
- HOLGATE, P. 1969 Species frequency distributions. Biometrika, 56:651-60.
- HURLBERT, S.H. 1971 The nonconcept of species diversity: a critique and alternative parameters. Ecology, 52:577-86.
- KILBURN, P.D. 1966 Analysis of the species area relation. Ecology, 47:831-43.
- LIVINGSTONE, R.J. *et al.* 1976 Determination of sampling strategy for benthic macrophytes in polluted and unpolluted coastal areas. Bull. Mar. Sci., 26(4):569-75.
- LLOYD, M.; INGER, R.F.; KING, F.W. 1968 On the diversity of reptile and amphibian species in a Bornean rain forest. Am. Nat., 102(928):497-515.
- MARGALEF, R. 1951 Diversidad de especies en las comunidades naturales. Publ.Inst.Biol.Apl., 9:15-27.
- MCINTYRE, A.D.; ELLIOTT, J.M.; ELLIS, D.V. 1984 Introduction: design of sampling programmes. In: HOLME, N.A.; MCINTYRE, A.D. (eds.) Methods for the study of marine benthos. Oxford, Blackwell. 387p.
- MENHINICK, E.P. 1964 A comparison of some species-individuals diversity indices applied to samples of field insects. Ecology, 45:859-61.
- MIYARES, P.; ANADÓN, R. 1981 Área mínima de muestreo en poblaciones animales de los niveles superiores del intermareal: ejemplo en poblaciones de *Patella spp.* Oecol. Aquat., 5:185-93.
- MORISITA, M. 1959 Measuring of the dispersion of individuals and analysis of the distributional patterns. Mem.Fac.Sci.Kyushu Univ., E(Biol.) 2:215-35.
- _____. 1962 I_g -index, a measure of dispersion of individuals. Res.Popul.Ecol., 4:1-7.
- _____. 1964. Application of I_g -index to sampling techniques. Res. Popul. Ecol., 6:43-53.

- NIELL, F.X. 1974 Les applications de l'indice de Shannon a l'étude de la végétation intertidale. Bull. Soc. Phycol. France, 19:238-54.
- _____. 1977 Método de recolección y área mínima de muestreo en estudios estructurales del macrofitobentos rocoso intermareal de la Ría de Vigo. Invest. Pesq., 41(2):509-21.
- PICHON, M. 1978 Quantitative benthic ecology of Tuléar reefs. In: STODDART, D.R.; JOHANNES, R.E. (eds.) Coral reefs: research methods. Paris, UNESCO. 581p.
- PIELOU, E.C. 1966 The measurement of diversity in different types of biological collections. J.Theor.Biol., 13:131-44.
- _____. 1974 Population and community ecology: principles and methods. New York, Gordon and Breach Science Publ. 424p.
- _____. 1975 Ecological diversity. New York, Wiley-Interscience. 165p.
- _____. 1977 Mathematical ecology. New York, Wiley-Interscience. 384p.
- POOLE, R.W. 1974 An introduction to quantitative ecology. New York, McGraw-Hill. 532p.
- PRINGLE, J.D. 1984 Efficiency estimates for various quadrat sizes used in benthic sampling. Can. J. Fish. Aquat. Sci., 41(10):1485-9.
- ROJAS, B.A. 1964 La binomial negativa y la estimación de intensidad de plagas en el suelo. Fitotecnia Latinamer., 1(1):27-36.
- ROSSO, S. 1991a Estratificação e repartição vertical da dominância e diversidade em um domínio intermareal de substrato consolidado da região inferior do estuário de Santos-São Vicente e considerações sobre a adequação de diversos descritores na discriminação dos sub-sistemas mediolitorais inferior e superior. In: II Simpósio sobre Oceanografia, São Paulo, 21-25 Out. 1991. Resumos. São Paulo, IOUSP. p141.
- _____. 1991b. MINAREA e MOLINIER: novos aplicativos (IBM-PC) para dimensionamento amostral e caracterização da estrutura de comunidades de organismos sésseis e semi-sésseis. In: II Simpósio sobre Oceanografia, São Paulo, 21-25 Out. 1991. Resumos. São Paulo, IOUSP. p142.
- _____. MENDES, A.J.S.; ETEROVIC, A. 1992. Primeiras notas sobre a relação entre descritores e dimensões amostrais no estudo, pelo método dos contatos, da comunidade de um costão rochoso do estuário de São Vicente (SP). III. Diversidade e dominância. In: Reunião anual da Sociedade Brasileira para o Progresso da Ciência, 44, São Paulo, 12-17 Jul. 1992. Anais. São Paulo, SBPC. p.659.

- _____. MENDES, A.J.S. 1992. Relação entre descritores e dimensões amostrais no estudo de uma comunidade intermareal de costão rochoso. In: Simpósio sobre estrutura, funcionamento e manejo de ecossistemas. Rio de Janeiro, 27-29 Maio 1992. Resumos. Rio de Janeiro, UFRJ. p51.
- SANDERS, H.L. 1960 Benthic studies on Buzzard Bay. III. The structure of the soft bottom community. Limnol. Oceanogr., 5(2):138-53.
- SANTELICES, B. 1980 Muestreo cuantitativo de comunidades intermareales de Chile Central. Arch. Biol. Med. Exp., 13:413-24.
- SIERRA, F.; FERNANDEZ, C. 1984 El horizonte de *Corallina elongata* en la costa central de Asturias (Nde España). I. Área mínima y distribución espacial. Invest. Pesq., 48(2):255-84.
- SOUTHWOOD, T.R.E. 1966 Ecological methods: with particular reference to the study of insect populations. London, Methuen. 391p.
- TAYLOR, L.R. 1961 Agregation, variance and the mean. Nature, 189:732-5.
- WEINBERG, S. 1981 A comparison of coral reef survey methods. Bijdr. Dierk., 51(2):199-218.
- WIESER, W. 1960 Benthic studies in Buzzard Bay. II. The meio-fauna. Limnol. Oceanogr., 5(2):121-37.

Endereços:

ROSSO, S.

Depto de Ecologia Geral do Instituto de Biociências, Universidade de São Paulo.

Cx.Postal 11461 CEP 05499 São Paulo, Brasil.

BITNET <SERROSSO@BRUSP>.

Centro de Biologia Marinha (CEBIMar), Universidade de São Paulo.

Cx.Postal 11037 CEP 05499 São Paulo, Brasil.