

Capítulo IX

Introducción al Análisis de Datos de Frecuencia

Introducción

Uno de los datos más comunes que un biólogo tiene la oportunidad de encontrar son los de **Frecuencia** o **Enumeración** que consisten en conteos de objetos o eventos, separados por espacio o tiempo. Las técnicas utilizadas para analizar tales datos se agrupan con el nombre de **Métodos Estadísticos de Enumeración**. Estos se dividen en dos grupos: **Pruebas de ‘Goodness of Fit’** y **Análisis de Tablas de Contingencia**. Las técnicas empleadas en ambos grupos son las mismas, lo que cambia es la manera en que son aplicadas y las preguntas que se quiere responder. De estas técnicas las dos más ampliamente conocidas son la **Prueba del Chi-cuadrado** y el **Análisis de Tablas de Contingencia**. Otras menos conocidas pero no por eso menos importantes incluyen las pruebas **G**, **Poisson**, y **Binomial**.

Prueba de ‘Goodness of Fit’

Una de las preguntas más comunes que un biólogo se pregunta es si los datos de frecuencia que han sido recolectados toman la forma de una distribución teórica/esperada o no. Como un ejemplo muy simple, suponga que se atraparon 100 ratas y se determinó el sexo para cada individuo. Se desea conocer si la proporción de sexos observada (44 machos y 56 hembras) es significativamente diferente de la proporción esperada de 50 machos y 50 hembras (1:1).

Como siempre, antes de realizar una prueba inferencial, necesitamos elaborar una hipótesis a probar. Se definen entonces las hipótesis nula y alternativa (H_0 y H_a) Como H_0 es la declaración de no diferencia, en este caso se define como “la proporción de sexos para la población de ratas no es diferente de 1:1”. H_a será de la misma forma definida como, “la proporción de sexos para la población de ratas es diferente de 1:1”.

Hay dos pruebas que pueden usarse para probar las hipótesis definidas anteriormente: la prueba del **Chi-cuadrado** y la prueba **G**. La de Chi-cuadrado se conoció alrededor del año 1900 y es tal vez la prueba estadística más usada en Biología. La prueba G (también conocida como la prueba de la **proporción de ‘log-likelihood’**) fue presentada en 1935 y es menos usada. Ambas pueden usarse indistintamente. La prueba G es un poco más complicada de calcular en forma manual que la del Chi-cuadrado. Las dos llegan generalmente a las mismas conclusiones; pero cuando no lo hacen, muchos estadísticos prefieren la prueba G y por lo tanto recomiendan su uso rutinario. Ambas se explicarán en este capítulo, pero es decisión del usuario cual emplear.

La prueba del Chi-cuadrado:

El estadístico Chi-cuadrado (χ^2) se usa como medida de la desviación de un distribución resultado de los datos de un muestreo de alguna distribución esperada/teórica. χ^2 se calcula siempre usando el mismo formato:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (9.1)$$

Dónde: **O** es la frecuencia observada: número de eventos, individuos, cosas, etc. reportada para cada clase de observación,

E es la frecuencia esperada para cada clase de observación asumiendo que la hipótesis nula es verdadera.

Es importante entender la diferencia entre las frecuencias observada y esperada. Las frecuencias observadas son aquellas que se han visto, capturado, etc. En el ejemplo dado de las ratas, hay dos clases de observación (machos y hembras) para las que fueron hechas todas las observaciones: 44 machos y 56 hembras. Las frecuencias esperadas son aquellas que esperaríamos encontrar si la hipótesis nula es verdad, dado el número total de observaciones en el experimento (n). Ya que definimos la hipótesis nula como, “la proporción de sexos para la población de ratas **no es** diferente de 1:1” o 50% machos y 50% hembras, el número esperado de machos y hembras es:

$$E_{\text{machos}} = n \times 0.50 = 100 \times 0.50 = 50$$

$$E_{\text{hembras}} = n \times 0.50 = 100 \times 0.50 = 50$$

Ahora se puede calcular el estadístico χ^2 de acuerdo a la ecuación 9.1:

$$\begin{aligned} \chi^2 &= \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} \\ &= \frac{(-6)^2}{50} + \frac{(6)^2}{50} \\ &= 0.72 + 0.72 \\ &= 1.44 \end{aligned}$$

Examinando los cálculos anteriores, debería estar claro que mientras mayor sea la diferencia entre las frecuencias esperadas y observadas (O - E) mayor será el valor de χ^2 . En otras palabras, mientras peor sea la correspondencia entre las frecuencias observadas y esperadas mayor será el valor de χ^2 . Otra manera de analizar esto sería diciendo: a mejor correspondencia entre la distribución observada y la distribución esperada/teórica, menor el valor de χ^2 . χ^2 puede variar entre 0 (una correspondencia perfecta entre los valores esperados y observados) y un número muy grande (grandes diferencias entre los valores esperados y observados). χ^2 nunca puede ser negativo, si esto ocurre se ha cometido un error en alguno de los cálculos.

Al igual que en todas las pruebas inferenciales, las observaciones que se están probando deben ser el resultado de un muestreo al azar a partir de la población entera. Como resultado, aun si H_0 es en realidad verdadera, se espera ver alguna desviación entre la distribución observada y la teórica. La significancia de una desviación observada puede calcularse directamente, pero los cálculos son muy complejos. Afortunadamente existen tablas que permiten calcular la probabilidad aproximada de obtener la desviación observada (véase la Tabla de valores críticos de Chi-cuadrado en el página 129). Para determinar la probabilidad se necesita calcular los grados de libertad (ν) para la prueba. Se calcula ν como el **número de clases - 1**. En el ejemplo, ν es: 2 (macho/hembra) - 1 = 1.

Se necesita escoger el nivel (α) que se llama ‘significativo’ (véase el Capítulo V). Si se escoge $\alpha = 0.05$ se debe encontrar el valor crítico en la Tabla de valores críticos de Chi-cuadrado correspondiente a $\nu = 1$ y $\alpha = 0.05$. Este valor es 3.841. Para poder rechazar H_0 el valor de χ^2 debe ser **mayor o igual** al valor crítico. Como el valor es de 1.44, y es menor que el valor crítico no rechazamos H_0 .

El análisis de los datos de frecuencia puede generalizarse a cualquier número de clases. Considérese el siguiente ejemplo teórico: supóngase que se desea conocer si hay preferencia por un tipo dado de cebo para rata. 100 cebos de cuatro tipos diferentes (atún, mantequilla de maní, pescado fresco y queso) se dejaron en un área. La Tabla 9.1 muestra la cantidad de cada tipo de cebo comido por las ratas:

Tabla 9.1

<u>Cebo</u>	<u>Número</u>
Atún	31
Mantequilla de maní	69
Pescado fresco	35
Queso	65

La hipótesis nula es que la preferencia es la misma para todos los cebos. El número esperado de cebos (ej. el número de cebos que se espera que sean comidos dado que la hipótesis nula sea cierta) se calcula sumando todos los cebos tomados ($31+69+35+65 = 200$) y dividiendo por el número de tipos de cebo ($200/4 = 50$). Lo que se está diciendo es que si no hubiera preferencia por un cebo se esperaría que las ratas comieran el mismo número de cebos de cada tipo (50 en este ejemplo).

El análisis de Chi-cuadrado para estos datos es como sigue:

Cebo

	atún	maní	pescado	queso	Total
Observado	31	69	35	65	200
Esperado	50	50	50	50	200

$$\begin{aligned}\chi^2 &= \frac{(31-50)^2}{50} + \frac{(69-50)^2}{50} + \frac{(35-50)^2}{50} + \frac{(65-50)^2}{50} \\ &= \frac{289}{50} + \frac{361}{50} + \frac{225}{50} + \frac{225}{50} \\ &= 22.0\end{aligned}$$

Los grados de libertad (ν) son iguales al número de clases (4) - 1 = 3. Se escoge $\alpha = 0.05$.

Revisando la Tabla de valores críticos de Chi-cuadrado se ve el valor crítico $\chi^2_{0.05,3} = 7.815$. Como el valor encontrado de 22.0 es mayor que o igual a 7.815 se puede rechazar H_0 . Se concluye por el momento que hay una preferencia diferenciada por los diferentes tipos de cebo. El próximo paso, que desafortunadamente está fuera del alcance de este libro, sería hallar cuál de los tipos de cebo (si hay alguno) es significativamente preferido a los otros.

La Corrección de Yates para la Continuidad

Existe un problema al probar hipótesis usando datos de frecuencia con la distribución del Chi-cuadrado. Este consiste en que mientras los datos de frecuencia son discretos por su naturaleza (por ejemplo, nunca se verá $\frac{1}{2}$ ó $\frac{1}{4}$ de una observación), el Chi-cuadrado es una distribución continua. Como resultado, la prueba es solo una aproximación y no es exactamente válida. Afortunadamente esta aproximación es muy buena a menos que solo haya un grado de libertad. En tal caso se recomienda usar la **Corrección de Yates para la Continuidad**. La corrección de Yates tiene la siguiente forma:

$$\chi^2 = \sum \frac{(|O-E|-0.5)^2}{E} \quad (9.2)$$

Considerando esto, el análisis del primer ejemplo en este Capítulo (un caso donde $\nu = 1$) sería:

$$\begin{aligned}\chi^2 &= \frac{(|44-50|-0.5)^2}{50} + \frac{(|56-50|-0.5)^2}{50} \\ &= \frac{(5.5)^2}{50} + \frac{(5.5)^2}{50} \\ &= 1.2\end{aligned}$$

Sesgo en el cálculo de Chi-cuadrado

Como se mencionó anteriormente, el estadístico Chi-cuadrado es una buena aproximación a la distribución teórica a menos que $\nu=1$. Desafortunadamente, el estadístico Chi-cuadrado NO es una buena aproximación cuando cualquier valor esperado es menor que uno, o cuando más del 20% de los valores esperados son menores que 5. La prueba del Chi-cuadrado NO debe ser usada si uno o ambos de estos casos ocurre. Si esto sucede, se puede hallar alguna ayuda en la prueba G.

Prueba G ‘Log-likelihood Ratio’:

Una nueva prueba se ha propuesto recientemente para los tipos de datos que se acaban de discutir. Es la conocida como ‘**Log-likelihood Ratio test**’ o la **prueba G**. Esta se usa en las mismas ocasiones que la del Chi-cuadrado aunque tiene la desventaja de que es un poco más difícil de calcular. Sin embargo, tiende a ser más confiable, especialmente cuando el número de observaciones es bajo. Esta prueba no puede ser usada si cualquier valor esperado es igual a 0 o cuando más del 20% de los valores esperados son menores que 3.

La prueba G se calcula como sigue:

$$G = 2 \times (\sum O \cdot \ln(O/E)) \quad (9.3)$$

Lo que es igual a

$$G = 2 \times (\sum O \cdot \ln(O) - \sum O \cdot \ln(E)) \quad (9.4)$$

dónde: **O** y **E** son, respectivamente, las frecuencias observadas y esperadas,
ln es el logaritmo natural.

La prueba G para los datos del ejemplo de las ratas y los cebos (página 102):

$$\begin{aligned} G &= 2 \times [(31 \ln(31) + 69 \ln(69) + 35 \ln(35) + 65 \ln(65)) - (31 \ln(50) + 69 \ln(50) + 35 \ln(50) + 65 \ln(50))] \\ &= 2 \times [(106.45 + 292.15 + 124.44 + 271.34) - (121.27 + 269.99 + 136.92 + 254.15)] \\ &= 13.92 \end{aligned}$$

El valor crítico $\chi^2_{0.05,3} = 7.815$. Como el valor 13.9 es mayor que o igual a 7.815 se puede rechazar H_0 . Así, se concluye que por el momento, con la evidencia que nos brindan los datos recolectados, hay una preferencia diferenciada por los diferentes tipos de cebos.

Pruebas con Tablas de Contingencia

En los ejemplos previos se trató con datos de frecuencia divididos en clases discretas por una variable (tipo de cebo, por ej.). Muchas veces, sin embargo, se necesita analizar diferenciados simultáneamente por dos variables. Se hace la pregunta “¿Son las frecuencias observadas en una variable, independientes de las frecuencias en la otra?” Como ejemplo se verá otro estudio del imaginario Pájaro Turista de Pecho Rojo. En este estudio se desea conocer si el tiempo de estadía en Galápagos es independiente del origen del ave.

La Tabla 9.2 es un ejemplo de una **Tabla de Contingencia**. Al **número de filas**, correspondiente al número de clases para una de las dos variables, se le da el símbolo **F**. Al número de columnas, correspondiente al número de clases para la otra variable, se le da el símbolo **C**. No importa cuál variable se asigne a las filas y cuál a las columnas, los resultados de la prueba serán los mismos en cualquier caso. Las tablas se describen por el número de filas y de columnas que tiene. Así, la Tabla 9.2 es un ejemplo de una tabla de 3 x 3 porque tiene tres filas y tres columnas. La clase correspondiente a las columnas es “duración de la estadía” y la clase representada por las filas es “origen”. Cada uno de los datos está encerrado en una “celda”. Por ejemplo, la columna C_1 tiene tres celdas, una para el número 11, otra para el 4 y otra para el 55.

Tabla 9.2

Duración de la estadía (Semanas)

Origen	<1	1 a 4	>4	total	
Norte América	11	38	13	62	(F ₁)
Europa	4	42	26	72	(F ₂)
Sudamérica	55	127	76	258	(F ₃)
total	70	207	115	392	(n)
	(C ₁)	(C ₂)	(C ₃)		

La hipótesis nula para todos los análisis de Tablas de contingencia es que las frecuencias de las filas son independientes de las de las columnas (o que las frecuencias de las columnas son independientes de las de las filas). En el caso que se estudia, la hipótesis nula (H_0) podría escribirse como “la duración de la estadía del ave en Galápagos es igual para todos los orígenes”, o “el origen del ave es el mismo para todas las duraciones de estadía”. La hipótesis alternativa (H_a) podría escribirse como, “la duración de la estadía es diferente para cada origen del ave” o “el origen no es el mismo para cada duración de estadía”. También se pueden plantear las hipótesis nula y alternativa, respectivamente, de la siguiente manera: “la duración de la estadía del ave es independiente del origen (o el origen del ave es independiente de la duración de la estadía)” y “la duración de la estadía del ave depende del origen (el origen del ave depende de la estadía)”.

Con la excepción de las Tablas de contingencia de 2 x 2 (véase los Capítulos X y XI), es necesario calcular las frecuencias esperadas para cada celda (combinación de origen y duración) si se supone que H_0 es verdadera.

El primer paso es calcular las sumas para cada una de las columnas y filas. Por ejemplo, la suma de la fila Origen = Norte América (F_1) es igual a: $11 + 38 + 13 = 62$. El mismo proceso se sigue para las columnas. La suma de la primera columna Duración = 1 semana (C_1) es: $11 + 4 + 55 = 70$. La Tabla 9.2 muestra las sumas de todas las filas y columnas. El número total de observaciones (n) es simplemente la suma de todos los totales de las columnas, o del total de las sumas de las filas. En este caso n es igual a 392. Como chequeo se pueden buscar los totales de las sumas de las filas y de las columnas. Si estas dos sumas no son iguales se ha cometido un error en alguna parte.

Antes de calcular las frecuencias esperadas es necesario entender perfectamente que es lo que se quiere decir con “independiente”. En estadística si se dice que una variable es independiente de otra, se quiere decir que “la probabilidad de observar una variable no está en ninguna manera relacionada con la otra”. Como se vio en el capítulo sobre probabilidad, la probabilidad de observar dos variables independientes juntas es igual al producto de las probabilidades de cada variable.

Si el origen y la duración de la estadía son independientes, como la hipótesis nula establece, ¿cuántas veces se espera ver cada combinación de origen y duración? ¿Cuál es la probabilidad esperada de aves de Norte América (NA) que se quedan menos de una semana? Si origen y duración son independientes, entonces la probabilidad de observar ambas simultáneamente es igual a la probabilidad de un ave de Norte América multiplicado por la probabilidad de cualquier ave que se quede menos de una semana.

$$\begin{aligned} \text{La probabilidad de un ave que venga de NA es: } & \frac{\# \text{ Aves de NA}}{\# \text{ total de aves}} \\ & = \frac{62}{392} \\ & = 0.1582 \end{aligned}$$

La probabilidad de que un ave se quede menos de una semana es:

$$\begin{aligned} & \frac{\# \text{ aves que se quedan menos de una semana}}{\# \text{ total de aves}} \\ & = \frac{70}{392} \\ & = 0.1786 \end{aligned}$$

La probabilidad de que un ave venga de NA y se quede por menos de una semana es:

$$\begin{aligned} P & = 0.1582 \times 0.1786 \\ & = 0.0282 \end{aligned}$$

El número esperado de aves que vengan de NA y se queden por menos de una semana es:

$$0.0282 \times 392$$

$$= 11.07$$

En general, la frecuencia esperada para una celda es:

$$\frac{F_i}{n} \times \frac{C_j}{n} \times n \quad (9.5)$$

cancelando dos de las n's se tiene:

$$F_i \times C_j / n \quad (9.6)$$

La Tabla 9.3 muestra los valores esperados para la Tabla 9.2.

Tabla 9.3

Duración de la estadía (Semanas)

Origen	<1	1 a 4	>4	total	
Norte América	11.07	32.74	18.19	62.0	(F ₁)
Europa	12.86	38.02	21.12	72.0	(F ₂)
Sudamérica	46.07	136.24	75.69	258.0	(F ₃)
	70.0	207.0	115.0	392.0	(n)
	(C ₁)	(C ₂)	(C ₃)		

Nótese que las sumas de las filas y las columnas, así como el número total de observaciones son los mismos en las Tablas 9.2 y 9.3. Si no son iguales, entonces se ha cometido un error.

Ambas pruebas, el Chi-cuadrado y G, pueden ser usadas para probar hipótesis nulas.

La prueba de Chi-cuadrado es:

$$\chi^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(O_{ic} - E_{ic})^2}{E_{ic}} \quad (9.7)$$

$$= \frac{(11 - 11.07)^2}{11.07} + \frac{(38 - 32.74)^2}{32.74} + \frac{(13 - 18.19)^2}{18.19} + \frac{(4 - 12.86)^2}{12.86} + \frac{(42 - 38.02)^2}{38.02}$$

$$+ \frac{(26 - 21.12)^2}{21.12} + \frac{(55 - 46.07)^2}{46.07} + \frac{(127 - 136.24)^2}{136.24} + \frac{(76 - 75.69)^2}{75.69}$$

$$= 0.0004 + 0.845 + 1.481 + 6.104 + 0.418 + 1.129 + 1.731 + 0.627 + 0.001$$

= 12.329

Para probar este valor en contra del valor crítico apropiado se tienen que calcular los grados de libertad (ν). Para cualquier Tabla de Contingencia ν se calcula por la fórmula:

$$\nu = (f-1) \cdot (c-1) \quad (9.8)$$

$$\nu = (3 - 1)(3 - 1)$$

$$= 4$$

El valor crítico para $\alpha = 0.05$ y $\nu = 4$ es 9.448. Dado que el valor hallado es \geq que el valor crítico se rechaza H_0 y se concluye por ahora, en base a la evidencia que proveen los datos, que la duración de la estadía no es igual para los diferentes orígenes del ave turística de pecho rojo.

La prueba G correspondiente es:

$$G = 2[\sum \sum O \ln(O) - \sum F \ln(F) - \sum C \ln(C) - \sum n \ln(n)] \quad (9.9)$$

$$G = 2 \left[\begin{array}{l} 11 \ln(11) + 38 \ln(38) + 13 \ln(13) + 4 \ln(4) + 42 \ln(42) + 55 \ln(55) + 127 \ln(127) + 76 \ln(76) \\ 62 \ln(62) + 72 \ln(72) + 258 \ln(258) + 70 \ln(70) + 207 \ln(207) + 115 \ln(115) + 392 \ln(392) \end{array} \right]$$

$$= 14.545$$

Los grados de libertad (ν) son los mismos calculado que para la prueba del Chi-cuadrado (9.8). El valor crítico para $\nu = 4$ y $\alpha = 0.05$ es como antes = 9.448. Como 14.545 es ≥ 9.448 se rechaza H_0 y se concluye igualmente que por ahora y en base a la evidencia de los datos usados, que la duración de la estadía no es igual para los diferentes orígenes del imaginario pájaro turista de pecho rojo.

Sesgo en el Análisis de Contingencia:

El análisis de los datos de una Tabla de Contingencia está sujeto a las mismas restricciones que las pruebas del Chi-cuadrado y G. Esto significa que cuando el estadístico Chi-cuadrado va a ser usado, NO debe ser usado si cualquier valor esperado es menor que 1, o si más del 20% de los valores esperados son menores que 5. Si el estadístico G va a ser usado no es válido si cualquier valor esperado es igual a 0 o cuando más del 20% de los valores esperados son menores que 3.

Capítulo X

Introducción a los Métodos de Muestreo

Introducción

Con excepción de las comunidades más simples (y posiblemente artificiales) todas las demás comprenden interacciones complejas entre un gran número de organismos representantes de muchas especies. Así como no existe un individuo que viva independiente de los otros individuos de la comunidad, no hay una comunidad que viva independiente de los millones o aún billones de organismos que comparten la misma localización (ej. bacteria del suelo, nematodos, gusanos, insectos, aves, mamíferos, etc.). Tampoco existe una comunidad que viva en aislamiento de los elementos no vivientes de su ambiente, tales como el suelo, la lluvia, el viento y la temperatura. Todas las comunidades vegetales son el producto final de un número virtualmente infinito de interacciones que ocurren entre una comunidad y su ambiente.

Los ecólogos han hecho un gran esfuerzo para describir las comunidades. El objeto de esta descripción es permitir que personas diferentes al observador puedan construir una imagen mental de los organismos del área en cuestión. Ninguna investigación seria acerca de una comunidad puede comenzar sin determinarse primero que especies están presentes, su abundancia y distribuciones. El objetivo de este Capítulo es presentar una serie de técnicas analíticas y descriptivas que pueden ser usadas para investigar una comunidad vegetal típica. Se examinarán primero tanto las ventajas como las desventajas de cada técnica con respecto a su precisión, confiabilidad y costos relativos. A pesar que en este Capítulo (y el siguiente) se hablará sobre comunidades de plantas, mucho es aplicable igualmente a comunidades de animales.

Abundancia

Una de las piezas de información que se puede tener acerca de una comunidad vegetal es la lista de las especies que existen en ella y sus abundancias. Idealmente sería magnífico contar cada planta en el área de estudio, pero incluso en un área de un metro cuadrado se pueden encontrar cientos de individuos y docenas de especies. Cuando el área de estudio es de muchos metros cuadrados o kilómetros cuadrados, esta técnica normalmente no es práctica. Existen muchas especies de plantas y algunas de animales inferiores (anémonas, esponjas, etc.) para las que es muy difícil determinar qué es un individuo. En lugar de contar cada individuo, un observador puede hacer una estimación visual subjetiva de la abundancia relativa, o puede hacer una estimación cuantitativa, basada en datos obtenidos a partir de muestreos al azar tomados en el área de estudio.

Es importante tener presente que no existe la técnica o el método perfecto. Cada técnica tiene sus ventajas y sus desventajas. En general, mientras más precisa es una técnica, mayor es el esfuerzo que se debe poner para usarla. Por ejemplo, el método más confiable (al menos para áreas pequeñas) es el conteo total de todos los individuos. Esto requiere una mayor inversión de tiempo

y esfuerzo. Por último, cuando se escoge una técnica o grupo de técnicas, el investigador tiene que hacer un balance entre su necesidad de confiabilidad y precisión y el tiempo, esfuerzo, dinero y herramientas requeridas para llevarla a cabo.

Cálculo Subjetivo de la Abundancia

Ordenamiento:

El método más simple y rápido para describir la abundancia es hacer una lista de las especies presentes y asignar a cada una un valor subjetivo relativo. Este método ha sido usado por mucha gente y existen varios sistemas de clasificación.

El más fácil, y posiblemente el menos útil, consiste en ordenar especies de acuerdo a su abundancia relativa aparente. Se da un valor de 1 a la especie más abundante a discreción del observador, un valor de 2 a la que le sigue en abundancia y así sucesivamente para todas las especies presentes. Este método no suministra información acerca de las abundancias absolutas: El valor de 1 se le puede dar a una especie con 10 ó 1000 individuos, dependiendo de la abundancia de las otras especies presentes.

Escala de Dominancia:

Un mejoramiento del método anterior es la escala de dominancia. Esta usa clases de abundancia mucho más comparables. Las cuatro clases principales de este sistema de clasificación son:

Dominante, Abundante, Ocasional y Raro

con los prefijos **muy** y **local** para proveer más información sobre abundancia y distribución. Así, una especie de vegetación distribuida en manchas, esparcida por el área de estudio, puede ser referida como **localmente abundante**. Es infortunado que se haya escogido la palabra **dominante** para representar la clase más abundante, ya que la palabra indica algún tipo de dominancia sociológica, tal como la usada para describir el papel del macho dominante o alfa en comunidades de babuinos. En el contexto presente, la palabra **dominante** es usada en el sentido de **muy abundante**. Una especie que crece en manchas esparcidas y excluye a las demás especies del sitio donde crece, puede describirse localmente como dominante.

Escalas Braun-Blanquet y Domin:

Un sistema de clasificación más elaborado llamado Sistema de Braun-Blanquet, usa estimaciones de la abundancia así como del área que cada especie parece cubrir. Una variación de esta técnica, la Escala Domin, añade algunas clases adicionales al sistema Braun-Blanquet.

Escala Domin		Braun-Blanquet	
Cobertura del 100%	10	5	Cobertura >75%
Cobertura del 75-100%	9		
Cobertura del 50-75%	8	4	Cobertura del 50-75%
Cobertura del 33-50%	7	3	Cobertura del 25-50%
Cobertura del 25-33%	6		
Abundante, cobertura del 20%	5	2	Muy numeroso o cobertura del 5-25%
Abundante, cobertura del 5%	4		
Esparcido, poca cobertura	3		
Muy esparcido, poca cobertura	2	1	Bastante, poca cobertura
Escaso, poca cobertura	1		
Aislado, poca cobertura	X	X	Escaso o muy escaso muy poca cobertura.

Problemas con el cálculo de abundancias relativas:

El mayor problema con el cálculo subjetivo es la tendencia a que hayan variaciones entre diferentes personas y diferencias en la interpretación de lo que cada clase significa. Lo que una persona puede juzgar como abundante otra puede juzgar solo como frecuente. Para una persona es a menudo difícil escoger entre dos clases adyacentes. Mientras más clases tenga un sistema (por ejemplo, mientras más finas sean las distinciones hechas sobre la abundancia) más grande será la incertidumbre acerca de qué clase escoger en un momento dado. Si el sistema de clasificación que se está usando tiene más clases que las que se pueden usar con confianza, se debe reducir el número de clases.

Los cálculos subjetivos están también expuestos a un gran número de sesgos por observación. Generalmente las especies raras son subvaloradas y las comunes sobrevaloradas. Especies con colores brillantes o muy conspicuas por razones varias (ej. especies con flores) son a menudo sobrevaloradas en comparación con especies que no poseen estas características. Difícilmente se puede evitar este tipo de sesgos en cálculos subjetivos, aunque con práctica se pueden reducir.

Cálculo cuantitativo de la abundancia

Los cálculos cuantitativos ofrecen la ventaja de estar relativamente libres de sesgos causados por observación y sus valores están bien definidos. Como resultado, se pueden comparar medidas tomadas en diferentes localidades. La herramienta principal que se utiliza para hacer todos los cálculos cuantitativos es el cuadrante. Un cuadrante es cualquier área fija usada para tomar

sub-muestras de un área mayor. Un cuadrante puede tener cualquier tamaño o forma y estar hecho de cualquier material. Sin embargo la mayoría son de forma cuadrada y están hechos con marcos de madera o metal. Algunas veces los cuadrantes se limitan con cuerda, cuando los espacios a ser muestreados son demasiado grandes para usar los marcos. La mejor forma y tamaño para el cuadrante dependerá de varios factores que se discutirán luego.

Densidad:

Es el conteo del número de individuos en un área. Las densidades se dan en términos del número promedio de individuos por unidad de área (ej. el número de plantas por metro cuadrado). Las estimaciones de densidad se obtienen generalmente a partir de cuadrantes establecidos al azar. Se obtiene a partir de la muestra total un número de individuos relativo al tamaño del cuadrante usado. Cuando los tamaños de las muestras son grandes (véase la sección acerca del tamaño de muestra) el método es exacto y una medida absoluta de la varianza.

Cobertura:

Las medidas de densidad pueden presentar problemas cuando es difícil distinguir a los individuos. Muchas especies de plantas (y algunas de animales como, anémonas, esponjas, corales, etc.) crecen en grupos densos donde puede ser imposible o impráctico identificar individuos, o donde el término individuo puede ser difícil de aplicar. Muchas especies de pastos poseen troncos o rizomas bajo tierra, que envían troncos a la superficie a ciertos intervalos. No es realista pensar en sacar todo el sistema de raíces y buscar las plantas individuales. En este caso, sería mejor usar la cobertura como medida de la abundancia.

La cobertura está definida como la porción del suelo cubierta por la proyección perpendicular de las partes aéreas de las plantas en él presentes. Las estimaciones de cobertura ignoran los problemas de individualidad al medir el área de superficie cubierta por una especie particular. La cobertura se expresa normalmente como un porcentaje del área total cubierta por una especie particular y se estima tomando un número de muestras. Las estimaciones de cobertura pueden ser obtenidas a través de una estimación visual subjetiva. Esta técnica es usada en las medidas para Braun-Blanquet/Escala Domin y está por lo tanto sujeta a los problemas de estimación ya discutidos. La cobertura puede ser estimada objetivamente usando una herramienta que se llama **marco de agujas ('pin frame')**.

El principio en el que se basa el marco de agujas es el de examinar un gran número de puntos (literalmente el área debajo de una aguja de gran tamaño) en el área de estudio y determinar en cada punto cuales especies, si hay alguna, están presentes. Si, por ejemplo, después de examinar 100 puntos la especie 1 se encontró 36 veces, la estimación de cobertura para esa especie es 36%.

Hay dos clases principales de marcos de agujas, que difieren en la manera en que se localizan los puntos en la superficie. La primera clase es conocida como **marcos de agujas ópticas** y se forman al dividir el área de un cuadrante cuadrado con un retículo de cuerdas espaciadas regularmente que se encuentra en ambas superficies del cuadrante (véase la Figura 10-1). Ambos retículos, el superior y el inferior, deben estar directamente uno encima del otro. Idealmente el cuadrante debe tener unos centímetros de grosor para que haya un espacio entre los retículos. El

cuadrante es colocado en el suelo y el observador debe mirar directamente sobre la intersección de las cuerdas y registrar la especie que está exactamente debajo de ese punto (véase la Figura 10.2). Una estimación del porcentaje de cobertura para cada cuadrante se obtiene sumando por cada especie, el número de observaciones y dividiendo por el número total de intersecciones. Idealmente el porcentaje total de cobertura para todas las especies más el suelo desocupado será igual a 100%.

Los marcos de agujas ópticos son muy prácticos cuando los individuos se encuentran en superficies casi planas y donde existe poca superposición de la vegetación. En la naturaleza, sin embargo, tales ejemplos son relativamente raros. Con más frecuencia las comunidades de plantas son altamente estratificadas con muchas capas de vegetación (véase la Figura 10.3). Por ejemplo, un punto en el suelo puede estar cubierto por musgo, pasto, arbusto y árbol, todos a diferentes alturas. Incluso las plantas individuales tienen varias capas de hojas que se superponen. En estos casos puede ser necesario usar un **marco con agujas** que se puedan ajustar a la altura de la vegetación (este es nuestro segundo tipo de marco de agujas). La Figura 10.4 muestra un marco de agujas típico. El número, tamaño y espacio entre las agujas varía grandemente entre diferentes marcos de agujas. Las agujas deben bajarse una a la vez y la especie tocada por cada aguja, registrada. El porcentaje de cobertura se estima en la misma manera que con el marco óptico (número de veces especie debajo de la aguja/número de agujas usadas). En el caso de vegetación estratificada el porcentaje total de cobertura para todas las especies puede superar 100% (¿por qué?).

Figura 10.1

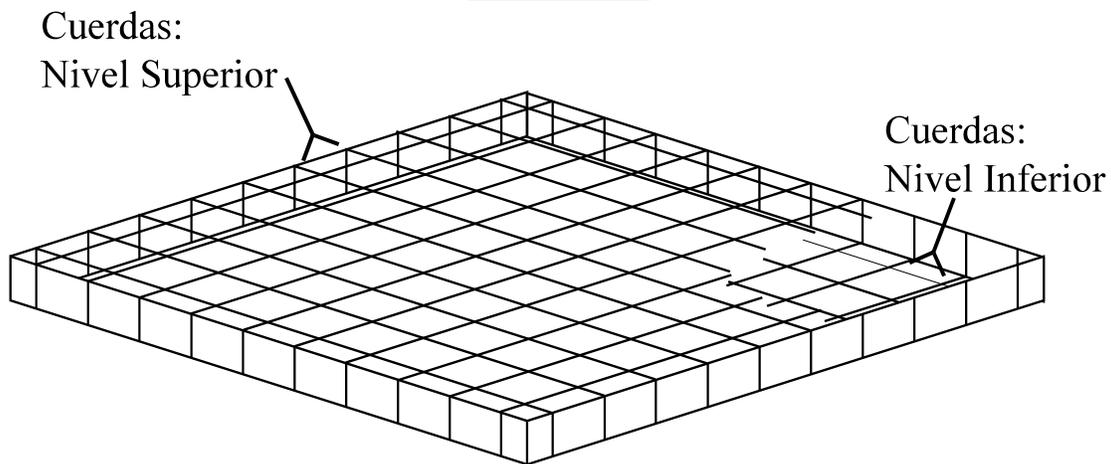


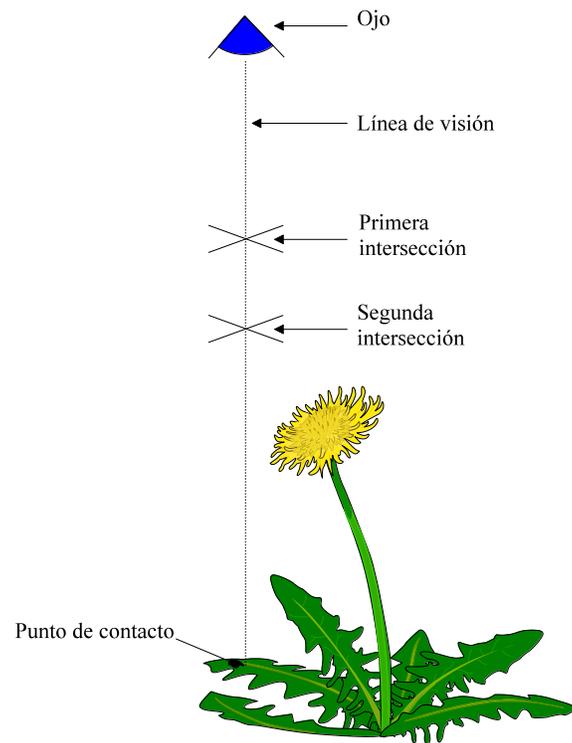
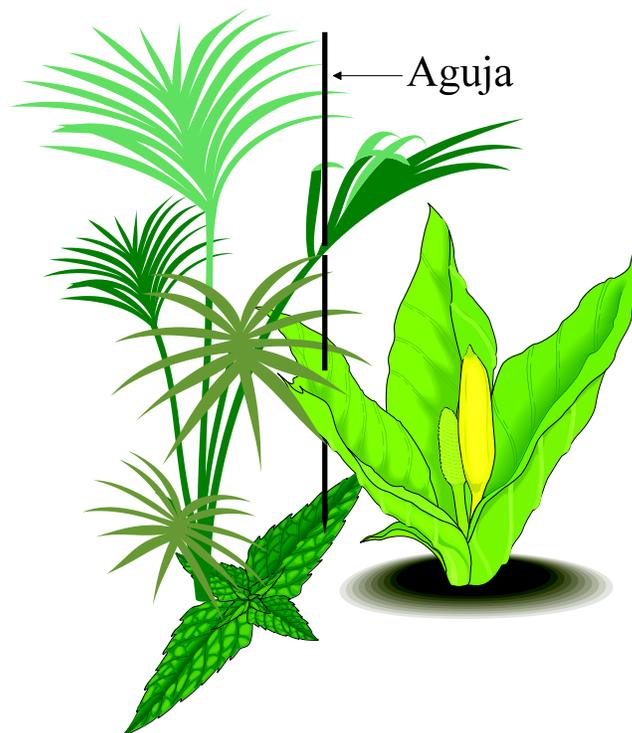
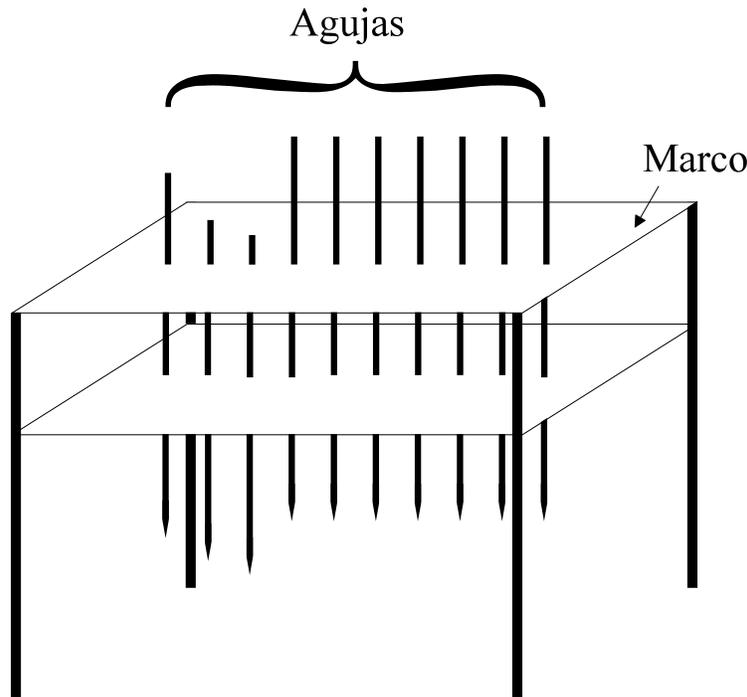
Figura 10.2**Figura 10.3**

Figura 10.4**Frecuencia:**

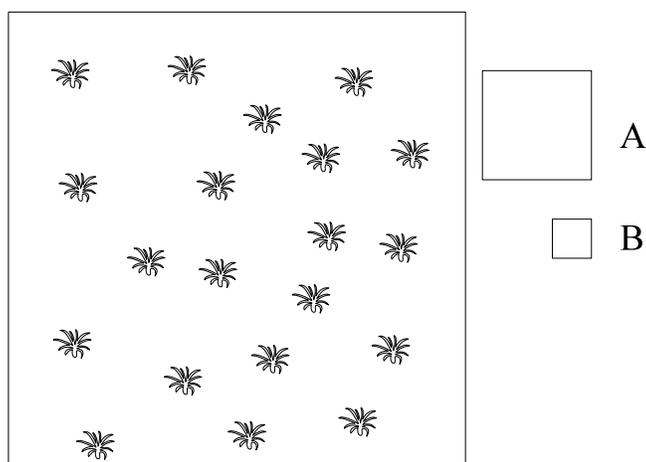
a frecuencia de una especie es la medida de la probabilidad de encontrarla al colocar un cuadrante al azar en un área. Así, si una especie tiene una frecuencia de 10% debería en promedio ocurrir cada 10 cuadrantes examinados. La frecuencia se obtiene muy fácilmente anotando si una especie está presente o no en una serie de cuadrantes puestos al azar. Por ejemplo, si 20 de 25 cuadrantes contienen la especie “roja”, entonces su frecuencia es: $20/25 \times 100 = 80\%$.

La única ventaja de este método de medida de abundancia es la velocidad y facilidad con que un área puede ser muestreada. Por esto es muy usada, olvidando a menudo sus muchas desventajas. Hay tres problemas principales relacionados con la interpretación de la estimación de la frecuencia:

1. Efecto del tamaño del cuadrante: Es claro a partir de la Figura 10.5 que los cuadrantes **A** y **B** de tamaños diferentes darán valores de frecuencia diferentes para la especie representada en el cuadro de al lado.

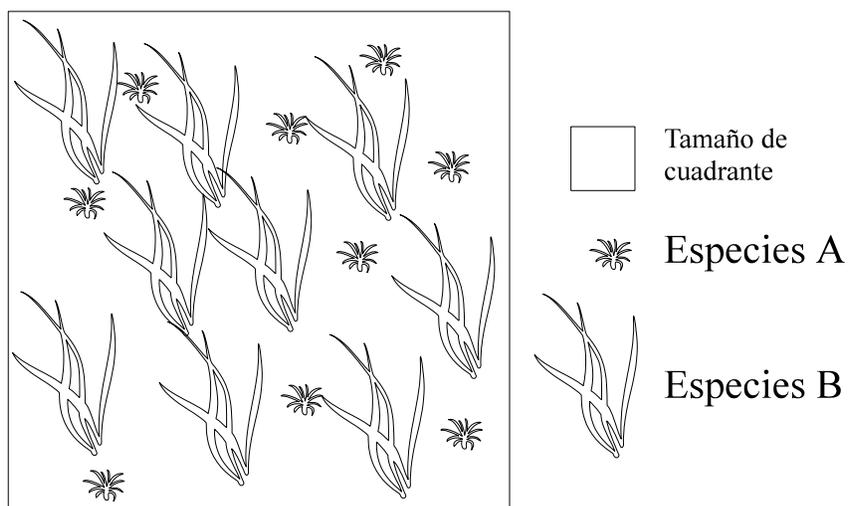
Con el cuadrante **A** se obtendrán valores de frecuencia de cerca del 100%, mientras el **B** dará valores intermedios. Así, los valores de frecuencia dependen en gran proporción del tamaño del cuadrante. Consecuentemente si se desea comparar datos de diferentes parcelas con el mismo tipo de vegetación, es esencial usar el mismo tamaño de cuadrante. Por lo mismo, **siempre** que se den valores de frecuencia se debe mencionar el tamaño del cuadrante.

Figura 10.5



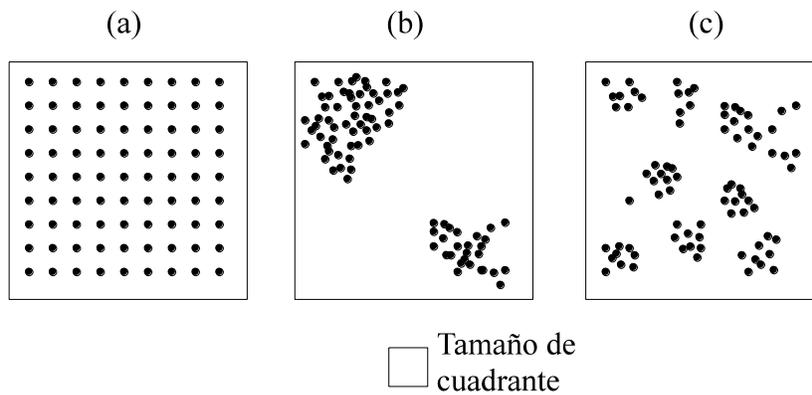
2. Efecto del tamaño de la planta: De nuevo, es obvio que los valores de frecuencia para las especies **A** y **B** en la Figura 10.6 diferirán aunque ambos tengan la misma densidad. Con el cuadrante que se muestra, la especie **A** tendrá una baja frecuencia y la **B** una muy alta.

Figura 10.6



3. Efecto de la distribución espacial de individuos: En la Figura 10.7 se muestran tres distribuciones, todas con la misma densidad y cobertura, pero diferente patrón. Con el tamaño de cuadrante mostrado la frecuencia estimada para la especie **A** será alta, **B** muy baja, y **C** intermedia. La distribución uniforme de **A** hace casi imposible tomar una muestra con cero de presencia, mientras el arreglo en montón de **B** nos presenta el otro extremo.

Figura 10.7



En general, la frecuencia es una medida útil de la abundancia cuando las comparaciones se van a hacer a gran escala, siempre y cuando se tomen en cuenta las limitaciones de la técnica. Este método será poco usado como medida de la abundancia a menos que por razones de tiempo y equipo sea completamente imposible usar otro método. La suma de la cobertura y la frecuencia para una especie dada se llama **valor de importancia**. De nuevo, es preciso indicar el tamaño del cuadrante cuando se escriben valores de frecuencia (o valores de importancia).

Capítulo XI

Introducción a los Métodos de Campo y su Análisis

Técnicas Analíticas y de Muestreo

Después de la presentación de medidas cuantitativas de abundancia hecha en el capítulo anterior, se debe recordar que antes de comenzar un programa de muestreo, es necesario diseñar cuidadosamente el plan de acción para obtener la máxima cantidad de información de las muestras tomadas. Una preparación cuidadosa anterior al inicio de la toma de muestras puede eliminar muchas horas perdidas en el campo. Este es el tema de este capítulo.

Muestreo al Azar:

Ya se ha usado la técnica de muestreo al azar para obtener medidas de abundancia, pero ¿por qué debemos usar muestreo al azar? Primero, las muestras al azar son muestras cuyas localizaciones en un área de muestreo definida están distribuidas al azar. La explicación de por qué se usa muestreo al azar tiene mucho que ver con la naturaleza humana. Si se deja a la discreción del investigador la localización de los cuadrantes, ésta sería casi siempre sesgada. Un sesgo común es distribuir las muestras a intervalos regulares. Si la vegetación está distribuida al azar no habría problema. Sin embargo, la vegetación es raro que esté distribuida al azar. El peor de los casos que se encontraría si se usa un muestreo a espacios regulares es si la vegetación está distribuida a los mismos espacios regulares (véase la Figura 11.1). El resultado podría ser que se sobrestime la abundancia de algunas especies y subestime la de otras.

Otro sesgo común al recoger muestras es el de sobre-colectar un área mientras se sub-colectan otras. Es común sobre-colectar en áreas de alta densidad (porque son más interesantes) y sub-muestrear las de baja densidad. Ambos sesgos pueden resultar en la representación desproporcionada de algunas especies en los resultados.

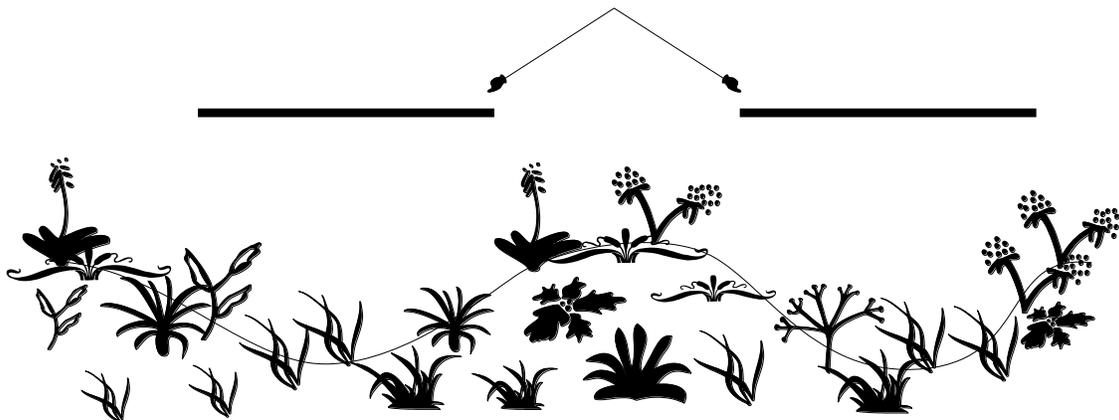
Finalmente, el muestreo al azar es un requerimiento de muchas pruebas estadísticas. Si el muestreo no es al azar, los tipos de análisis posibles son limitados en gran medida. Sin embargo, como se verá más tarde, hay situaciones donde algunos elementos de muestreo no al azar son permitidos o incluso necesarios.

Diseñar un régimen de muestreo al azar puede ser desde muy fácil hasta sumamente complicado. Como regla general, especialmente en estudios de campo, se debe mantener el diseño lo más simple posible, siempre tratando de mantener el número de variables experimentales al

mínimo. Desafortunadamente, hasta una discusión básica sobre buen diseño experimental requeriría mucho más espacio del que tenemos en este libro.

Figura 11.1

Cuadrantes espaciados en forma regular



Números al Azar:

En algún momento, cuando se desea hacer un muestreo al azar, se necesita escoger uno o más valores (localidades, horas, observadores, individuos, etc.) de una lista o un rango de posibilidades. Por ejemplo, se está usando cuadrantes de 1 m². Para tomar muestras al azar de un área de 100 m² necesitaremos una forma de seleccionar un número dado de entre las 10,000 (100²) localizaciones posibles en el área. Debido a que se tiene que escoger frecuentemente al azar a partir de un amplio número de posibilidades, se necesita una forma de generar series de números al azar.

Hay muchas formas de obtener números al azar (una moneda, un dado, escoger números de dentro de un sombrero, etc.) No importa qué método se use es necesario que cada posibilidad (cara o sello; 1, 2, 3, 4, 5, ó 6; 1 a 100, etc.) tenga la misma probabilidad de ser escogido. Aún más, que la elección de cualquier valor no cambie la posibilidad de escoger valores futuros.

Los métodos más comunes como fuente de números al azar son las computadoras (y algunas calculadoras) y las Tablas de números al azar. Las computadoras y las Tablas de números al azar usualmente producen números con un rango de valores posibles muy diferentes de aquellos que un experimento típico requiere. Para convertir estos números al azar a un rango más útil, se debe primero determinar el valor máximo que puede ser obtenido del computadora o Tabla (**NA_{max}**) y los valores más grande (**MAX**) y más pequeño (**MIN**) requeridos para el experimento. Por cada número al azar (**NA**) se calcula un nuevo valor (**NV**) necesario para el experimento:

$$NV = \frac{NA}{NA_{MAX}} \times (MAX - MIN) + MIN \quad (11.1)$$

Normalmente se requieren números enteros, así que será necesario redondear **NV** al entero más cercano.

Número Mínimo de Muestras:

¿Cómo se determina el número de muestras que deben ser usadas en los experimentos? Siempre, cuando se recoge un muestreo, se está tomando una sub-muestra del área total. Es importante que las estimaciones que se hacen de la abundancia de esas sub-muestras representen con exactitud (lo más exacto posible) los valores reales del área total. Si se fuera a tomar muestras en toda el área, se podría estar seguro de tener una estimación exacta. Sin embargo, cuando las áreas son muy grandes, este tipo de muestreo no es práctico. Se debe establecer un punto medio entre la exactitud del muestreo, el número de muestras que se desean hacer y el tiempo y dinero que se quiere invertir. No existen reglas que nos digan el número de muestras a tomar: cada caso debe ser estudiado independientemente.

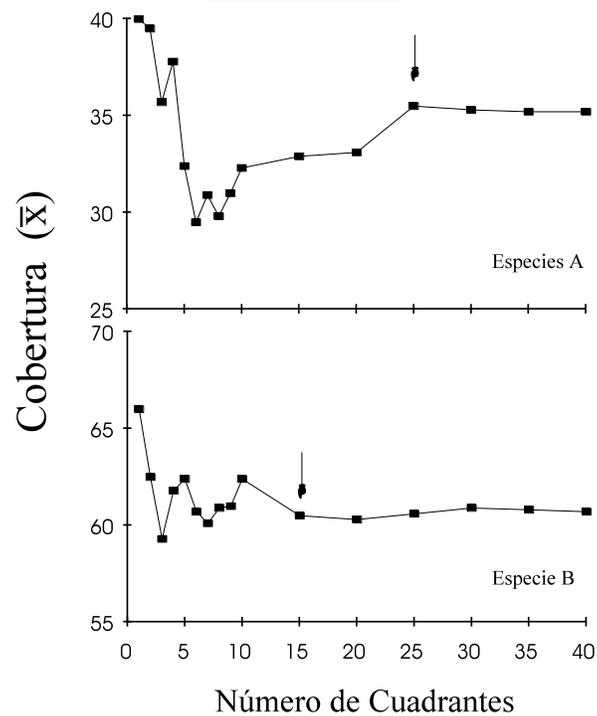
A menudo la pregunta sobre cuántas muestras a tomar se convierte en ¿cuántas muestras se debe tomar para obtener cierto nivel de exactitud? Hay algunas pruebas estadísticas que se pueden emplear para responder esta pregunta. La mayoría de estos métodos dependen de un número pequeño inicial de muestras, usadas para calcular el número mínimo de muestras adicionales necesario para alcanzar cierto grado de confianza. Estas técnicas están fuera del alcance de este libro.

Se debe considerar un método más subjetivo para estimar el tamaño mínimo de muestra. Este método se puede llevar a cabo en el campo, al mismo tiempo que se recogen los datos. El método incluye el cálculo de la media de la abundancia para cada especie (densidad, cobertura, frecuencia, etc.) después de que se recolecta el muestreo de cada nuevo cuadrante. Por ejemplo, los datos de la Tabla 11.1 representan el porcentaje de cobertura de dos especies diferentes obtenido a partir de un cierto número de cuadrantes. Para cada cuadrante se calcula la media del porcentaje de cobertura para ese y todos los cuadrantes previos.

El Gráfico 11.1 muestra la relación entre la media de la abundancia y el número del cuadrante. Inicialmente las medias muestran una gran cantidad de variabilidad, pero eventualmente se estabilizan. Para cada especie el punto en el cual la media deja de fluctuar es la estimación subjetiva del tamaño mínimo de muestra para esa especie. En el ejemplo que se ha propuesto, sería aproximadamente 35 y 15 cuadrantes para las especies 1 y 2 respectivamente. Este método resulta bueno para todas las medidas de abundancia.

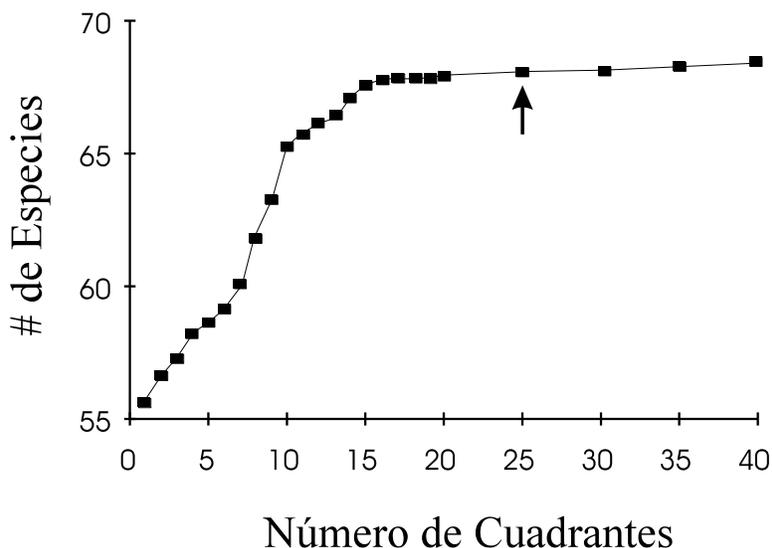
Tabla 11.1

<u>Cuadrante</u>	<u>Especie A</u>		<u>Especie B</u>	
	<u>Cobertura</u>	<u>Cobertura Media</u>	<u>Cobertura</u>	<u>Cobertura Media</u>
1	40	40.0	66	66.0
2	39	39.5	59	62.5
3	28	35.7	53	59.3
4	44	37.8	53	61.8
5	11	32.4	65	62.4
6	15	29.5	52	60.7
7	39	30.9	57	60.1
8	22	29.8	66	60.9
9	41	31.0	62	61.0
10	44	32.3	75	62.4
15	43	32.9	53	60.5
20	47	33.1	58	60.3
25	46	35.5	54	60.6
30	27	35.3	60	60.9
35	31	35.2	64	60.8
40	40	35.2	65	60.7

Gráfico 11.1

Además de las medidas de abundancia, al usar el número total de especies encontradas en la muestra, se puede estimar el número mínimo de cuadrantes necesarios. El “número de especies” se refiere al número de especies que se han encontrado (no el número de individuos) después de revisar “x” número de cuadrantes. El Gráfico. 11.2 muestra una relación típica entre el número total de especies y el número acumulativo de cuadrantes. Inicialmente el número de especies incrementa muy rápido, pero a una tasa decreciente. En algún punto (alrededor de 20 - 25, véase al Gráfico. 11.2) el número de las especies nuevas que se añaden a cada cuadrante se vuelve muy pequeño. Este punto es el estimativo que se busca del número mínimo de cuadrantes que se deben hacer.

Gráfico 11.2



Muestreo no Aleatorio:

El muestreo al azar es una regla general, pero hay algunas situaciones donde se prefiere el muestreo no aleatorio o muestreo **pseudo azar**. Una de estas excepciones ocurre cuando se desea investigar el influjo de un gradiente medioambiental en la abundancia y distribución de las especies. En este caso sería necesario un transecto, que es una serie de cuadrantes orientados en la dirección del gradiente. Un segundo caso ocurre cuando se necesita establecer en un mapa la vegetación de un área en términos de algún factor medioambiental como elevación, pH, temperatura, etc. En ambas situaciones se busca determinar la abundancia en relación a una localización geográfica (ej. cada cuadrante podría estar a una distancia específica de la playa). Los transectos son una herramienta muy útil para investigar los cambios de vegetación en relación con cambios medioambientales complejos. Imagínesse que se quiere investigar el efecto del agua salada sobre la vegetación. Un transecto que vaya desde la línea de marea baja hacia tierra firme, puede mostrar cambios vegetacionales en la marca de marea alta, en la de marea de sicigia, en las de líneas de tormenta, etc. La posición de los cuadrantes se relaciona directamente con la fuerte influencia del ambiente oceánico. Los cambios en la vegetación pueden reflejar diferencias en tolerancia a la sal, la desecación y la abrasión mecánica por las olas.

Algunas veces no es posible o no es particularmente deseable hacer un conteo no sesgado de la abundancia. Muchas veces se emplean técnicas de muestreo sesgadas porque no es posible hacer un verdadero muestreo al azar. Esto puede ocurrir porque el organismo en cuestión (y este es el caso general de los pequeños mamíferos) no puede ser muestreado al azar. Algunas veces el tipo de información que se desea sobre la abundancia no justifica el tiempo o gasto que requiere un sistema no sesgado de muestreo. Hay muchas maneras de describir la abundancia usando conteos no al azar y muchas de estas pueden envolver evidencia indirecta (ej. huellas, excretas) en vez de observaciones directas del organismo estudiado. Generalmente estas técnicas son llamadas **índices**.

Todos los índices tienen la forma de observaciones por unidad de esfuerzo, donde las observaciones son solo medidas relativas útiles de la abundancia (organismos vivos, huellas, excretas, vocalizaciones, etc.) y el esfuerzo se da en unidades de tiempo, muestras, volumen, etc. Todos los índices son medidas relativas de la abundancia porque nunca se sabe cómo se relacionan los índices con la abundancia real. Se asume que un incremento en un índice dado refleja un incremento en abundancia. No se puede asumir, sin embargo, que este incremento es constante para todos los valores de este índice o para la abundancia. A causa de estas limitaciones, los índices nunca se deben usar como estimadores de la abundancia absoluta. Solo pueden ser usados para indicar cambios relativos en abundancia (ej. incremento o disminución) y no se pueden comparar entre sitios, aun cuando se use el mismo índice. Por ejemplo, consideremos la siguiente situación: se desea conocer si la abundancia de una planta con flores en una localización particular está disminuyendo o permanece estable. En este caso no interesa el tamaño absoluto de la población, sino saber si está disminuyendo (posiblemente como resultado de la intervención humana). Se puede escoger como índice el número de flores (si tiene sólo una flor por planta) observado en un paseo de dos horas. El índice estará, por lo tanto, basado en un conteo totalmente no aleatorio y por lo tanto será, muy probablemente, sesgado (puede suceder que se ha escogido caminar solo por los senderos). Nunca se podrá saber qué fracción de las plantas se ven, pero si se repite el muestreo varias veces al año por varios años, se tendrá una indicación bastante buena de la estabilidad de la población.

Existen muchos índices posibles, a menudo varios para el mismo tópico. El usado será mejor si se acerca a los valores reales de abundancia. La selección del índice dependerá de la experiencia y conocimiento sobre la biología del organismo que se está estudiando.

Pruebas de Asociación

Como se mencionó anteriormente, las plantas individuales no viven aisladas de su medioambiente. Este incluye muchos elementos vivos y no vivos. La presencia de otras plantas, de la misma o diferentes especies, es a menudo un factor importante. Muchas especies crecen en grupos muy apretados compuestos de cientos o millares de individuos de la misma especie. Estas agrupaciones de plantas pueden cubrir grandes áreas y excluir la mayor parte de otras especies. Tales agrupaciones pueden reflejar una adaptación para la supervivencia (ej. protección contra depredadores); un patrón de crecimiento en el que las plantas nuevas comienzan en troncos vegetativos; o simplemente un patrón de dispersión de semillas. Las plantas pueden también

formar asociaciones mixtas de especies. Tales asociaciones pueden resultar de un gran número de situaciones. Diferentes especies pueden crecer juntas porque tienen más éxito cuando lo están que cuando no; porque una especie es parásita de la otra; o porque comparten requerimientos de *hábitat* muy parecidos. Hay muchas otras causas para la formación de asociaciones. Se dice que especies de plantas están asociadas positiva o negativamente si la probabilidad de encontrarlas en la misma área es mayor (o menor) que la que se esperaría si estuvieran distribuidas al azar la una con respecto a la otra.

Muchas asociaciones son obvias dadas su fuerza. Sin embargo, otras asociaciones son muy débiles y no son aparentes luego de una inspección detallada. En esos casos las asociaciones se pueden ver solo a través del análisis estadístico de muchas muestras. Mientras más débil es la asociación, mayor es el número de muestras que se requiere para poder demostrarla.

Diagramas de Dispersión:

Si dos especies están distribuidas independientemente (ej. presencia o ausencia de una especie no tiene efecto en la presencia de la otra) entonces la abundancia de cada especie en una localización dada debe también ser independiente. Si las abundancias están relacionadas de alguna forma, entonces las dos especies no son independientes la una de la otra. Los diagramas de dispersión proveen una forma rápida (pero cruda) de determinar si las abundancias están relacionadas, y de hallar el tipo y fuerza de la relación si ésta existe.

Los diagramas de dispersión tienen, generalmente, dos ejes cada uno representando la abundancia de una especie. Cada punto en el diagrama indica la abundancia (normalmente densidad o % de cobertura) de cada especie en un cuadrante. Si dos especies están relacionadas los puntos tenderán a caer a lo largo de una línea; la fuerza de la asociación determinará cuán esparcidos están los puntos a lo largo de la línea. Si no hay asociación entre las especies, los puntos tenderán a estar esparcidos por todo el gráfico.

Hay algunas pruebas estadísticas que pueden ser usadas para calcular la significancia estadística observada en un diagrama de dispersión. Las pruebas más comunes son las de **regresión y correlación** (véase los Capítulos VII y VIII). El objetivo de estas pruebas es determinar la probabilidad de obtener la relación observada, entre las dos especies, puramente al azar. La hipótesis nula en estas pruebas siempre es que no hay correlación entre las especies.

Prueba de Chi-cuadrado para Asociaciones:

Mientras los diagramas de dispersión son usados principalmente con densidad y porcentaje de cobertura, la prueba de Chi-cuadrado es usada muchas veces para buscar asociaciones basadas en datos de frecuencia. La prueba está basada en el siguiente teorema de probabilidad:

Si **X** e **Y** son dos eventos independientes (en nuestro caso eventos pueden ser interpretados como la presencia de un individuo de la especie **X** o **Y** en un cuadrante) que ocurren con probabilidades (% de frecuencia) $P(\mathbf{X})$ y $P(\mathbf{Y})$, respectivamente, entonces la probabilidad de que

X e **Y** ocurran simultáneamente (ej. ambas especies se encuentren juntas en el mismo cuadrante) es:

$$P(\mathbf{X}) \times P(\mathbf{Y}) \quad (11.2)$$

Por ejemplo, si la especie **X** tiene una frecuencia del 10% y la especie **Y** tiene una frecuencia del 50% y **X** e **Y** son completamente independientes la una de la otra, la probabilidad de encontrar ambas en el mismo cuadrante es:

$$(10/100) \times (50/100) = 0.05 \text{ (5\%)}$$

Esto significa que se espera encontrar ambas especies juntas en uno de cada 20 cuadrantes. Si dos especies no son independientes, la fórmula anterior no tendrá valor. Si tenemos datos de frecuencia de **n** cuadrantes en los cuales **a** cuadrantes contienen ambas especies **X** e **Y**, **b** cuadrantes que contienen solamente **X**, **c** cuadrantes que contienen solo **Y** y **d** cuadrantes que no contienen ni **X** ni **Y** (note que $a + b + c + d = n$) entonces se puede usar la Tabla de Contingencia de 2 x 2 (pronunciada “dos por dos”):

		Especie X		
		+	-	
Especie Y	+	a	b	a+b
	-	c	d	c+d
		a+c	b+d	n

+ presencia
- ausencia

La prueba de Chi-cuadrado calcula la significancia de la diferencia entre los valores observados (a, b, c, d) y los valores esperados calculados a partir del teorema de probabilidades dado anteriormente. El número esperado de cuadrantes que contienen ambos, **X** e **Y** (asumiendo que son independientes) es igual a:

$$(a+b) \cdot (a+c) / n \quad (11.3)$$

De la misma manera, se puede calcular el número de cuadrantes que contendrían **X** sola, **Y** sola, y ninguna de las dos. Si los valores esperados de **a** y **d** son mayores a lo esperado, la asociación es positiva. Si **c** y **b** son mayores que lo esperado la asociación es negativa. El cálculo estadístico de Chi-cuadrado (χ^2) puede ser obtenido convenientemente por medio de la fórmula:

$$\chi^2 = \frac{(ad-bc)^2 \cdot n}{(a+b)(a+c)(c+d)(b+d)} \quad (11.4)$$

La Tabla de Chi-cuadrado provee la probabilidad de obtener la diferencia observada entre los resultados esperados y observados puramente por casualidad. Siempre es posible obtener en un muestreo una gran diferencia entre los valores observados y los esperados aun cuando las especies

sean completamente independientes las unas de las otras. Usualmente se escoge alguna probabilidad arbitraria como significancia estadística. La fórmula de arriba trabaja bien a menos que el número esperado de cuadrantes para una celda específica (a, b, c, d) sea menor que 5. En ese caso se debe utilizar una fórmula un poco diferente:

$$\chi^2 = \frac{(|ad-bc| - n/2)^2 \cdot n}{(a+b)(a+c)(c+d)(b+d)} \quad (11.5)$$

Como ejemplo, se considerarán los datos obtenidos a partir de un estudio sobre comunidades intermareales de algas que se realizó en una playa cualquiera. Para que el ejemplo sea más simple, se usarán sólo los datos para dos especies: un alga verde (*Ulva lactuca*) y un alga roja (*Gelidium hancockii*). Los siguientes datos se obtuvieron a partir de un total de 100 cuadrantes localizados al azar en el área de estudio:

<u>Especie</u>		<u>Número de Cuadrantes</u>	
<i>Ulva lactuca</i>	<i>Gelidium hancockii</i>		
presente	presente	(a)	30
presente	ausente	(b)	20
ausente	presente	(c)	25
ausente	ausente	(d)	25

Tabla de Contingencia de para los **Valores Observados**:

		<i>Ulva lactuca</i>		
		+	-	
<i>Gelidium hancockii</i>	+	30	20	50
	-	25	25	50
		55	45	100

El próximo paso no es estrictamente necesario a menos que se rechace la hipótesis nula de que las frecuencias de las dos especies son independientes. Sin embargo, los cálculos en este punto ayudarán a entender mejor como funciona esta prueba.

Como se mencionó en un principio, si las frecuencias de las dos especies son independientes, la probabilidad de observar una especie es independiente de la de observar la otra. Por ejemplo, la probabilidad de observar *Ulva lactuca* (P(Ul)) es igual al número de veces que fue observada (55) dividido por el número de muestras tomadas (100) = 55/100 = 0.55. Igualmente, la probabilidad de observar *Gelidium hancockii* será P(Gh) = 0.50. La probabilidad de no observar una especie es 1 menos la probabilidad de observación. La probabilidad de no observar *Ulva lactuca* (P(Ul')) es igual a 1 - P(Ul) = 0.50, y la de no observar a *Gelidium hancockii* es 0.50.

A partir de los valores anteriores se puede construir una Tabla de Contingencia de valores esperados basada en la suposición de independencia. El número esperado de observaciones de ambas especies en el mismo cuadrante es $P(UL) \times P(Gh) \times \text{número total de cuadrantes} = 0.50 \times 0.55 \times 100 = 27.5$.

Tabla de Contingencia de para los **Valores Esperados**:

		<i>Ulva lactuca</i>		
		+	-	
<i>Gelidium hancockii</i>	+	27.5	22.5	50
	-	27.5	22.5	50
		55	45	100

Para comprobar si se han hecho los cálculos correctamente, la suma de los valores esperados para cada columna y fila debe ser igual a los valores correspondientes de los valores observados.

La prueba de χ^2 es una forma de determinar si las diferencias entre los valores observados y los esperados son suficientes para justificar el rechazo de H_0 y la aceptación de H_a .

Como ninguno de los valores esperados es menor que 5 se puede usar la ecuación 11.4:

$$\begin{aligned} \chi^2 &= \frac{(ad-bc)^2 \cdot n}{(a+b)(a+c)(c+d)(b+d)} \\ &= \frac{(30 \times 25 - 20 \times 25)^2 \cdot 100}{(50)(55)(50)(45)} \\ &= \frac{6250000}{6187500} \\ &= \mathbf{1.010} \end{aligned}$$

El valor crítico para $P=0.05$, $\chi^2_{0.05,1} = 3.841$. Como el valor calculado es menor que el valor crítico, no se rechaza H_0 ; no hay evidencia para rechazar la hipótesis de que las dos especies están distribuidas independientemente.

Para tomar en cuenta:

Esta ha sido una introducción muy corta a un tema muy amplio. Sin embargo, se ha tratado de tocar los puntos más relevantes a tomar en cuenta en cualquier muestreo. Es necesario anotar nuevamente que no existe una técnica que tenga solo ventajas. Lo importante es escoger la (o las) que presente(n) las mejores posibilidades frente al problema que nos ocupe. También se debe tener siempre presente que el conocimiento que se tenga acerca de la biología del organismo que se estudia es uno de los principales factores para hacer una selección correcta del método de muestreo. He aquí la importancia de la investigación bibliográfica previa a la realización de cualquier estudio. Con este libro esperamos que tengan mayor éxito en sus muestreos. ¡¡Buena suerte!!.

***P.D.* En último caso consulte al estadístico más cercano...**