# Capítulo III

# Introducción a los Gráficos (2)

## Introducción

Hasta ahora hemos aprendido a graficar grupos de datos relativamente simples. Los métodos que se han estudiado han estado restringidos a variables simples divididas por un sistema de clasificación simple (por ejemplo, medidas de peso divididas por el sexo de individuos). Debido a que los problemas biológicos resultan a menudo en situaciones donde se tienen varias variables subdivididas por más de un sistema de clasificación (ej. medidas de peso, longitud y condición general del cuerpo divididas por edad, sexo y localización de individuos) es necesario estudiar los métodos apropiados para graficar este tipo de situaciones.

Para poder manejar grupos de datos más complejos, es necesario extender muchos de los métodos aprendidos anteriormente, de dos dimensiones a tres. También se necesitará aprender algunos métodos nuevos que son específicos para ciertos problemas multidimensionales. Una vez aprendidas, estas técnicas permitirán representar una gran cantidad de información. Se debe siempre tener cuidado de no representar en un gráfico más información de la que se puede entender. Mantenga siempre en mente que a pesar de que un gráfico puede ser claro para **usted**, puede no serlo en absoluto para alguien que ve el gráfico por primera vez. Cuando sea posible se deben usar colores o sombras para identificar las diferentes variables o clases.

Exceptuando ciertas circunstancias especiales y algunas técnicas, generalmente es un error tratar de representar más de cinco variables a la vez. En la mayoría de los casos el 'costo' de emplear dos gráficos en vez de uno nos retribuye en mayor comprensión. En general, cuando se está graficando en **n dimensiones** es posible representar **n+1** (y a menudo **n+2**) variables a la vez. Por ejemplo, si se está trabajando con un gráfico bidimensional, es posible representar la relación entre tres variables. Se debe entender, sin embargo, que no siempre es posible representar una combinación de variables dada.

## Técnicas para Graficar Variables Múltiples

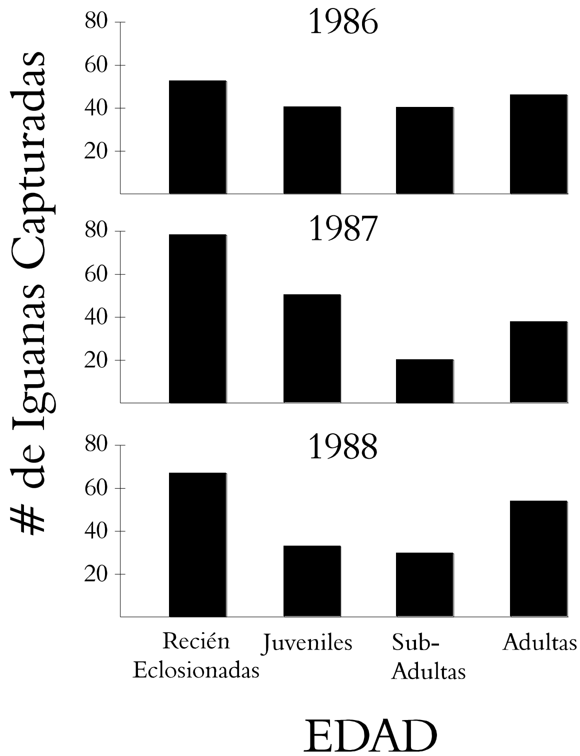
Tres variables discretas:

Pregunta: ¿Cuántas? ¿En qué proporción?

Con el Gráfico 2.1 se solucionó el problema de representar dos variables de frecuencia discreta usando diagramas de barras. Ahora se mostrará como extender el uso de estos diagramas para incluir tres variables discretas. Como ejemplo, se considerará que los datos de la tabla 2.1 representan capturas para el año 1988. A estos datos se adicionarán los datos de captura para los años 1986 y 1987. Con estos datos se desea mostrar cómo el número de iguanas en cada clase de edad ha cambiado a través de los tres años de estudio.

Antes de comenzar se debe decidir si se está interesado en cambios absolutos o relativos de los números. La respuesta a esta pregunta determinará si se usarán frecuencias relativas o absolutas. Dependiendo de la forma en que los datos fueron colectados y las preguntas que se formulen, se podrá usar una o la otra, o las dos. Por ahora se asumirá que las diferencias absolutas son importantes. La solución más fácil a este problema es usar una serie de diagramas de barras como se muestra en el Gráfico 3.1.

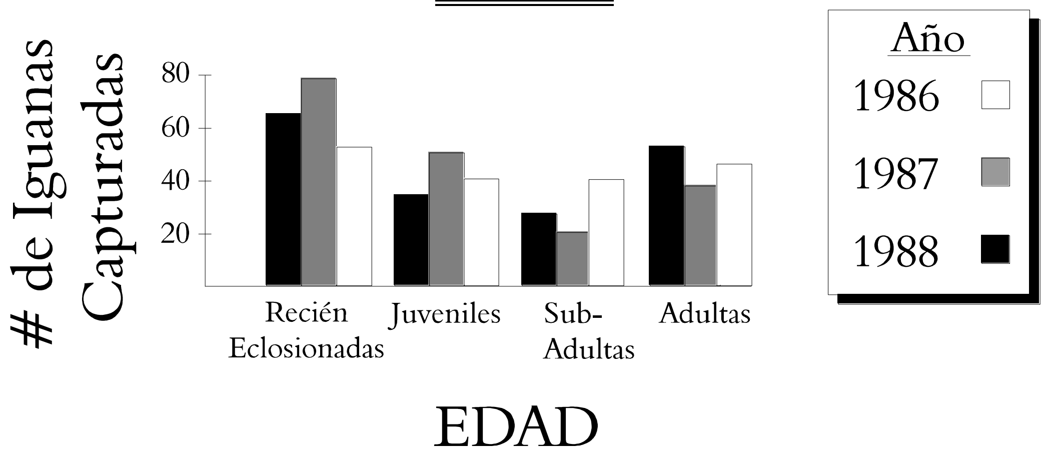
**Gráfico 3.1**




Cuando se hacen comparaciones entre gráficos como los del Gráfico 3.1, se debe recordar que las escalas de los ejes deben ser iguales, sin importar si se están usando frecuencias absolutas o relativas. Si no se mantienen las escalas iguales, las comparaciones serán, en el mejor de los casos, difíciles y en el peor, confusas. Una segunda forma de representar los mismos datos es hacer un gráfico simple como se ve en el Gráfico 3.2.

En este caso los datos están agrupados, primero de acuerdo a la edad (**clasificación primaria**), y segundo por año de captura (**clasificación secundaria**). Debe mantenerse siempre el mismo orden de las barras (ej. año de captura). Para indicar las barras con los años correspondientes, se representa cada año con un color o sombra únicos y se indica el código mediante una **leyenda.**

**Gráfico 3.2**

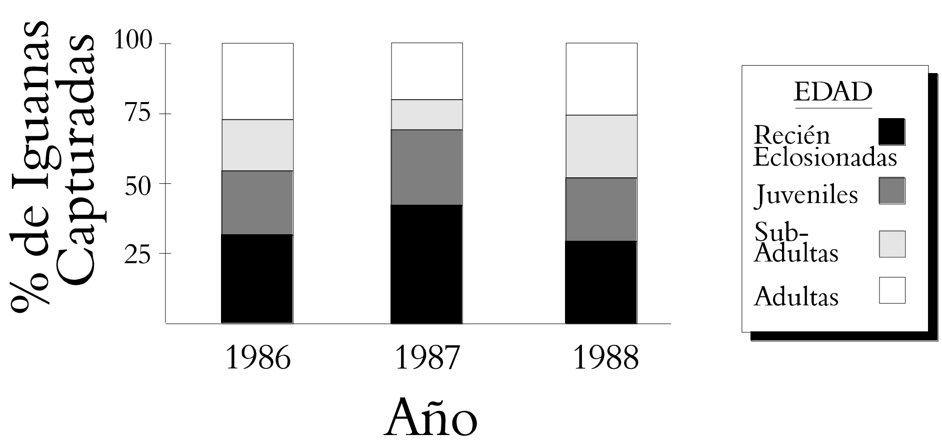



Resulta obvio que esta técnica es relativamente limitada en el número de clases primarias y secundarias que se pueden mostrar claramente. Recuérdese que nunca debe hacer un gráfico tan complicado que sea difícil de interpretar. Si se tiene que usar sombras en vez de colores se debe cuidar las combinaciones de los diferentes patrones de sombreados. Patrones como /// \\\ y ≡, son difíciles de distinguir cuando juntos.

Un tercer método es el de apilar, en lugar de agrupar las barras. Los **diagramas de barra apilados** son usados generalmente en el caso en que la suma de cada pila es la misma (ej. 100%). Como ejemplo se utilizarán los mismos datos usados para los Gráficos 3.1 y 3.2. En este caso sin embargo, se agruparán los datos primero por año de captura y luego por clase de edad. Dado que el número de iguanas capturadas en cada año es diferente se usará el % de frecuencias. Cada barra apilada representará el 100% de capturas para el año.

Así como con las barras agrupadas, es muy importante que se mantenga igual el orden de los segmentos de barra (en este caso las clases de edad). La representación de los datos de los Gráficos 3.1 y 3.2 en un diagrama de barras apiladas se muestra en el Gráfico 3.3.

**Gráfico 3.3**




Finalmente, estos datos pueden ser representados por un histograma tridimensional. Discutiremos estos gráficos más adelante.

Una variable continua subdividida en una o más clases:

Pregunta: ¿Cuántos? ¿Cómo están distribuidos?

Se vio previamente cómo representar distribuciones de frecuencia usando histogramas (ej. Gráfico 2.4). Ahora se complicará este problema dividiendo una única variable continua en sub-grupos. Como ejemplo, se considerará el siguiente problema: Se desea conocer si la distribución del peso de iguanas capturadas (descrito como la masa de iguanas) es la misma para los dos sexos (macho y hembra). Se recolectó y pesó cuidadosamente el mismo número de iguanas de cada localidad.

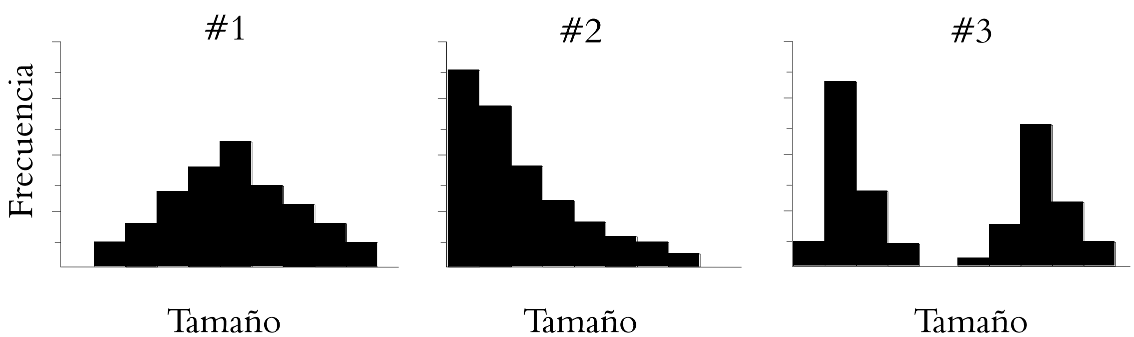
La primera solución al problema de representar los datos es separarlos de acuerdo al lugar de captura en dos Histogramas. Los gráficos pueden entonces ser colocados juntos en la misma hoja tal como se ve en el Gráfico 3.1. Para hacer más fácil las comparaciones es importante mantener las mismas escalas en los ejes de ambos gráficos. Si los datos están divididos en más de dos clases ésta es la forma más común de tratarlos.

La segunda solución a este problema se muestra en el Gráfico 3.4. Se crea un histograma de todos los datos como se hizo para el Gráfico 2.4. Para indicar que los datos pertenecen a diferentes clases se añade color o sombras a esas barras, o parte de barras que pertenecen a una de las dos clases. En el ejemplo siguiente los datos para los machos han sido sombreados.

La ventaja del Gráfico 3.4 es que hace muy fácil detectar diferencias en las medias de las distribuciones (si existen). Este también permite comparar las formas de las distribuciones, pero esta comparación se realiza de mejor manera usando la primera solución. Muy a menudo, como en el ejemplo anterior, se aumentan las soluciones al indicar alguna medida de tendencia central, tal como la media.

**Gráfico 3.4**

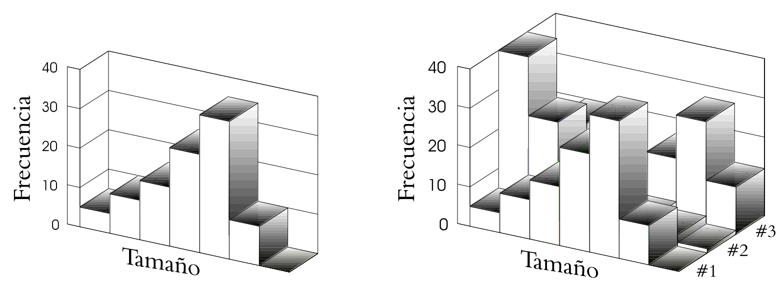
**Gráfico 3.5a**




Aunque los diagramas 3D pueden parecer impresionantes y por ello 'mejores' que cualquiera de los otros métodos, hay ciertos puntos negativos que hacen su uso restrictivo. El primer problema es que los histogramas 3D son muy difíciles de dibujar a mano, y para muchos grupos de datos, a menos que se disponga de una computadora el tiempo que se gastaría no sería justificable. Aún más importante, todos los gráficos tridimensionales sufren algún grado de distorsión debido a la perspectiva. Estas distorsiones hacen difícil comparar con precisión las alturas de las diferentes barras. Además, los histogramas tridimensionales sufren por el hecho de que las barras en la parte de adelante oscurecen las barras en la parte de atrás. Como resultado de estos dos últimos problemas, los histogramas tridimensionales nunca deben ser usados para comparar cuantitativamente datos que pertenecen a clases diferentes. Estos son impresionantes a la vista y proveen indicaciones sobre líneas generales de comportamiento de datos y diferencias entre ellos, pero en general no deben preferirse a las técnicas bidimensionales.

## **Gráfico 3.5**

(b) (c)



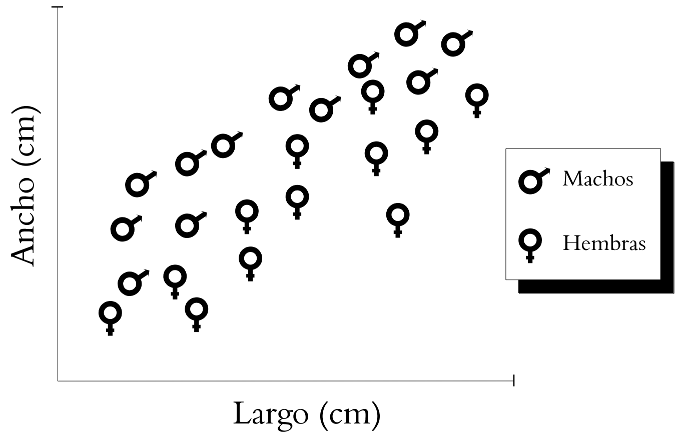

Una variable independiente y/o una o más variables dependientes:

Pregunta: ¿Cuál es la relación entre estas variables?

Se conoce como crear gráficos bidimensionales para dos variables continuas. En Biología hay muchas ocasiones en las cuales es necesario mostrar la relación entre varias variables continuas y discretas al mismo tiempo. Hay muchas versiones acerca del tipo de gráficos que se van a discutir a continuación. Muchas de las técnicas pueden ser combinadas exitosamente para producir nuevos gráficos. Con el aprendizaje de unos cuantos “trucos” se pueden graficar un amplio rango de problemas.

El primer problema que se enfrentará es aquel que relaciona dos variables continuas, tales como las señaladas en el Gráfico 2.10, las cuales están separadas en dos distintos sub-grupos, por ejemplo medidas de hembras y machos. Para solucionar este problema se representan todos los puntos en el mismo gráfico, pero usando diferentes símbolos para identificar a los diferentes sub-grupos. Como ejemplo, el Gráfico 3.6 muestra los datos para longitud y ancho del cuerpo, los cuales están divididos en medidas para hembras y machos.

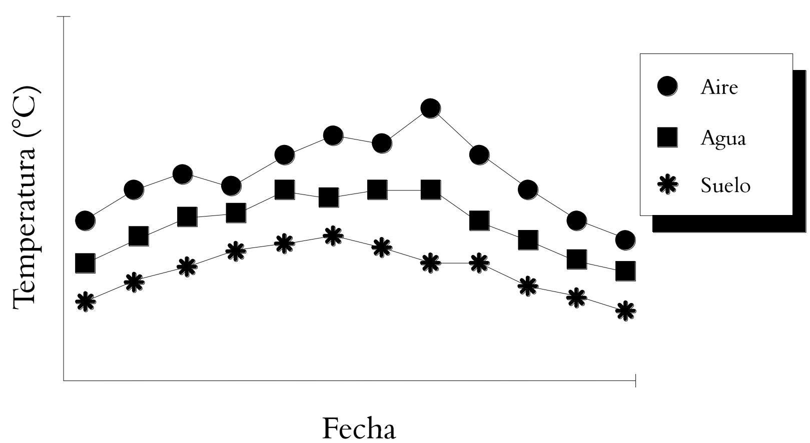
**Gráfico 3.6**




Como con los histogramas, cuando se presentan resultados para más de un grupo de datos, se debe incluir una leyenda cercana a la parte superior del gráfico para indicar los símbolos corresponden a cada subgrupo. Teóricamente no hay límite al número de sub-grupos que se pueden presentar en un gráfico. Sin embargo, un gráfico con muchos sub-grupos y/o con muchos puntos puede ser muy difícil de interpretar. A veces es mejor dividir los sub-grupos entre dos o más gráficos, especialmente si los sub-grupos pueden separarse en grupos.

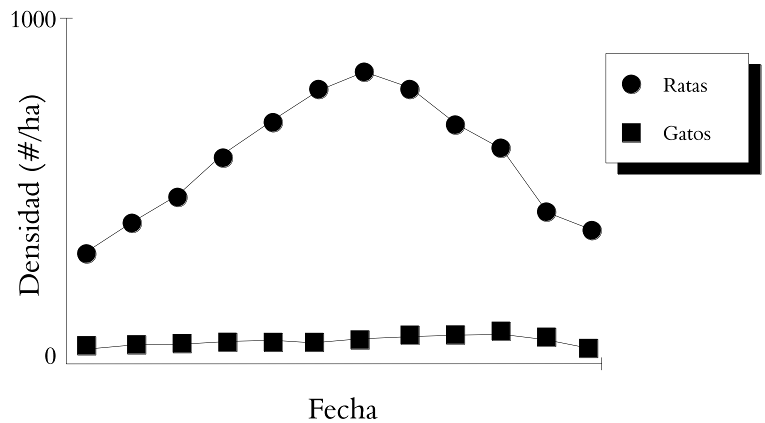
En el ejemplo anterior se trató con dos grupos de datos que, a pesar de estar separados en hembras y machos, son del mismo tipo, ej. Observaciones de longitud y ancho. Frecuentemente, sin embargo, se desea contrastar tipos de datos que son de diferente naturaleza. Como ejemplo, imaginemos que se han tomado medidas de temperatura del aire, tierra y agua durante algunos días. Ahora se tienen tres grupos de datos de diferente naturaleza, todos basados en la misma escala de medida y todos relacionados con una variable común (fecha). Es perfectamente legítimo combinar los tres grupos de datos en el mismo gráfico usando el mismo grupo de ejes. El Gráfico 3.7 es un ejemplo.

**Gráfico 3.7**



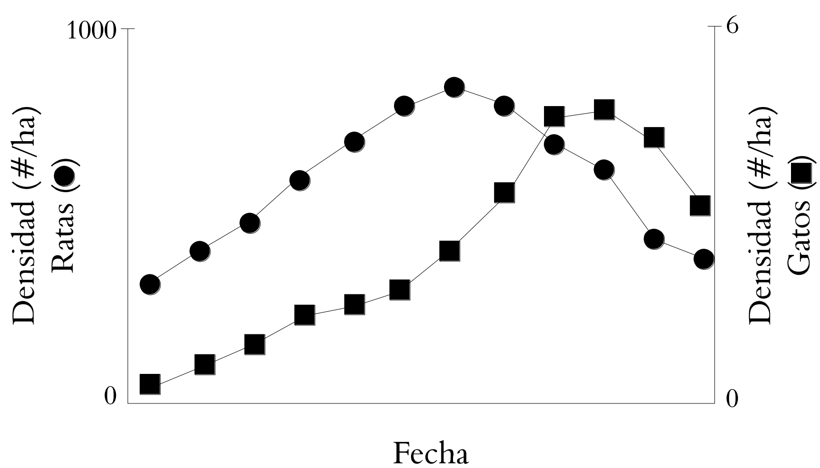
Soluciones como las del Gráfico 3.7 son correctas, a menos que la escala de medidas de los diferentes datos sea muy diferente en magnitud, tal como en el caso mostrado en el Gráfico 3.8. Nótese que debido a que la magnitud de la densidad de ratas es mucho mayor que la densidad de gatos, es imposible ver los cambios en la densidad de gatos durante el período de estudio.

**Gráfico 3.8**




Para resolver este problema se puede usar dos gráficos, uno para la densidad de gatos y otro para la de ratas, o se puede usar el lado derecho del gráfico y añadir un nuevo eje. En este último se podría, por ejemplo, graficar la densidad de ratas con respecto al eje de la izquierda y la densidad de gatos con respecto al de la derecha (Gráfico 3.9). Nótese que ahora se pueden seguir fácilmente las variaciones en las densidades de ambos y cómo están relacionados temporalmente. No hay restricción a la variedad de combinaciones que se puedan hacer, con la excepción de que todos deben compartir una variable común que debe ser graficada en el eje de la base ($\overline{X}$). Es posible incluso combinar los Gráficos 3.8 y 3.9 para representar varias variables al mismo tiempo. Como siempre, tenga cuidado de no hacer sus gráficos muy complicados.

**Gráfico 3.9**



Tres variables continuas al mismo tiempo:

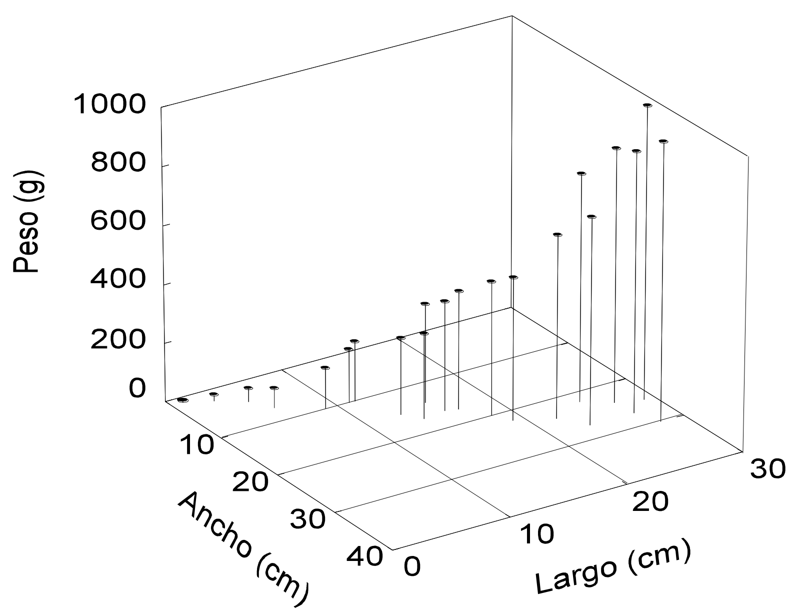
Pregunta: ¿Cuál es la relación entre estas variables?

Previamente se ha tratado solo con datos que se presentan como un par de observaciones simultáneas. Ahora se desea trabajar con el problema de representar la relación existente entre tres o más variables simultáneamente. Se debe entender por “simultáneos” datos que fueron obtenidos del mismo individuo, múltiples medidas abióticas tomadas el mismo día, estimativos de densidad de varias especies en el mismo momento. En cada caso se verá cómo pares de variables se relacionan y también cómo se relacionan todas las variables juntas.

Cuando hay tres variables, es posible graficar los datos en un gráfico tridimensional. El Gráfico 3.10 muestra un ejemplo de un gráfico tridimensional usando datos de tres medidas corporales tomadas del mismo individuo de cada una de 23 personas.

Los puntos son ubicados en el gráfico en la siguiente manera. Primero, se ubica el valor de una de las medidas (ancho o largo en el Gráfico 3.10) en uno de los ejes abajo. Segundo, se ubica el valor de la otra medida del otro eje abajo. Tercero, se localiza el punto correspondiente en la superficie reticulada de abajo dibujando líneas imaginarias en ángulos rectos desde los puntos ubicados en los dos ejes de abajo. Cuarto, se ubica el valor de la medida del eje vertical (peso en nuestro caso). Use esta altura para dibujar una línea vertical desde el punto localizado en el retículo de abajo. La punta de esta línea es la ubicación del punto que representa el dato en el gráfico tridimensional.

**Gráfico 3.10**




No siempre es necesario incluir las líneas verticales en el gráfico, pero éstas ayudan al lector a localizar los puntos. Cuando hay muchos puntos para mostrar, las líneas a menudo se sobreponen y dejan de ser prácticas. Cuando hay muchos puntos y la parte más importante del gráfico es la forma que toman los datos, no es aconsejable usar las líneas verticales.

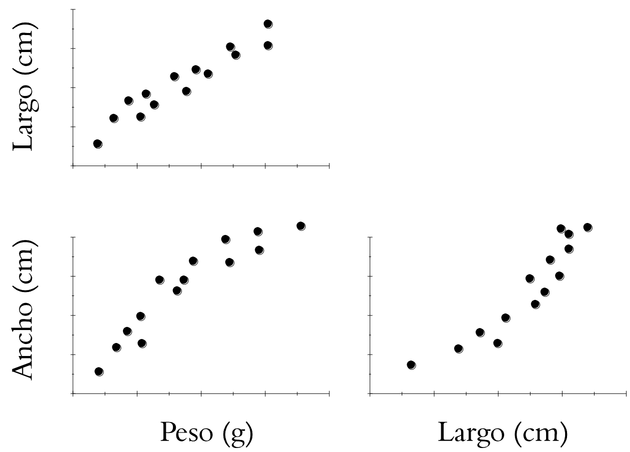
Los gráficos tridimensionales se hacen mejor en una computadora, primero porque son muy tediosos de hacer a mano, y lo que es más importante, porque muchos programas permiten cambiar la perspectiva de visión del gráfico, lo cual puede mejorar inmensamente la interpretación de los datos.

Finalmente, como en el Gráfico 3.6, es posible indicar sub-agrupaciones usando diferentes símbolos en una leyenda.

Cuando el número de variables continuas simultáneas pasa de tres, no hay forma de representar los datos en un gráfico multidimensional. Una de las mejores soluciones es la de graficar todos los pares posibles de variables en una forma conocida como **Graficación de ‘Draftsman’**. Esta es una manera efectiva de representar la relación entre tres o más variables continuas. Los diagramas de ‘Draftsman’ son limitados porque no muestran relaciones multivariables porque en ellos se grafican solo interacciones entre dos variables.

Los mismos datos usados en el Gráfico 3.10 fueron utilizados para crear el gráfico de ‘Draftsman’ del Gráfico 3.11. Cada gráfico de dos variables representa una rotación del Gráfico 3.10, es como si estuviéramos viendo todos los ángulos en cada par de ejes.

**Gráfico 3.11**




El número de gráficos necesario para crear un gráfico de 'Draftsman' se incrementa rápidamente con el número de variables. Si **n** es el número de variables, entonces el número total de posibles combinaciones de dos variables es $\mathbf{(n^2-n)/2}$.

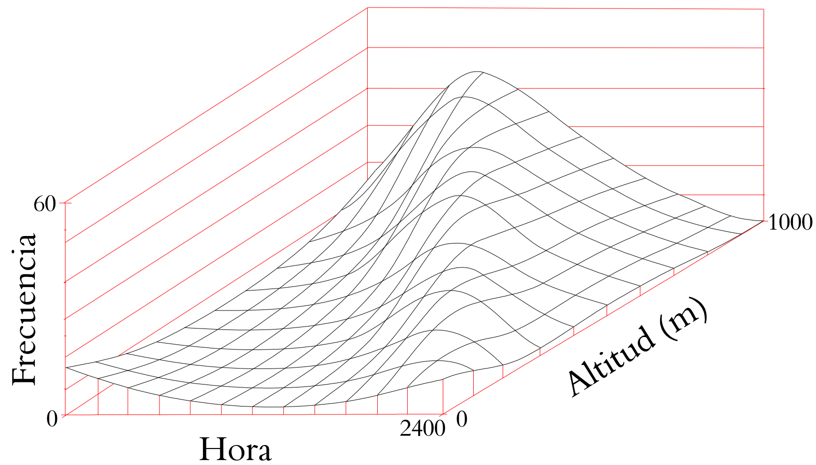
Un caso especial de tres variables continuas ocurre cuando dos de éstas forman un retículo con intervalos regularmente espaciados; el valor del tercero depende solamente de las combinaciones de las otras dos variables. Las primeras dos variables deben ser independientes la una de la otra (ej. el valor de una variable no influye en el valor de la otra). Por ejemplo, estas variables pueden representar localizaciones en un mapa (espacios regulares de longitud y latitud) o combinaciones regularmente espaciadas de dos variables independientes (ej. disponibilidad de alimento y presión de depredación, con la abundancia como tercera variable).

El Gráfico 3.12 usa como ejemplo las observaciones de tortugas a intervalos regulares de tiempo y altitud en una isla imaginaria. Los puntos graficados en la superficie son localizados de la misma manera que en el Gráfico 3.10. Sin embargo, en vez de dibujar líneas verticales, los puntos se conectan en un retículo donde hay cuatro vecinos.

El Gráfico 3.12 muestra una superficie muy lisa donde se ve que la mayoría de las observaciones de las tortugas ocurre a medio día en las altitudes más grandes. Las observaciones se incrementan suave y monótonamente con la altitud para todas las horas del día cuando hay observaciones. Con respecto a la hora del día, las observaciones se incrementan a un máximo alrededor de las 1200hr. en todas las altitudes, y luego decrece otra vez.

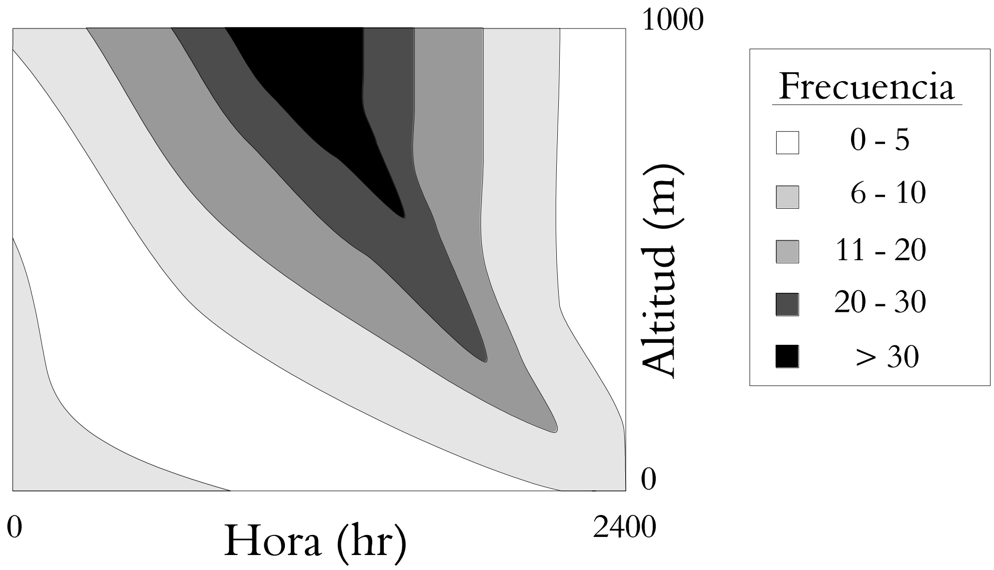
Dado que, en este caso la superficie en el Gráfico 3.12 es lisa, el Gráfico es fácil de interpretar. Sin embargo, muchas veces la superficie es irregular con muchas colinas y valles. Dependiendo del punto de visión, a veces alguna información en los extremos del gráfico se pierde detrás de alguna colina. Esto se puede arreglar al cambiar el punto de visión, pero muchas otras veces tenemos que perder algo de información.

**Gráfico 3.12**




Los mismos datos usados para estos **gráficos de superficie** pueden ser representados en otros llamados **gráficos de contorno**Error! Bookmark not defined.. La mayoría de la gente ha visto éstos en forma de mapas topográficos. En ellos, las líneas de nivel muestran la altitud. Todos los contornos tienen esta misma forma general. Las dos variables (en mapas éstas son longitud y latitud) espaciadas regularmente se localizan en un retículo bidimensional. La tercera variable se convierte en intervalos discretos. A todas las áreas del retículo que pertenecen al mismo intervalo se les asigna el mismo color o sombreado**.** Si la superficie es razonablemente lisa, puede ser suficiente con indicar la magnitud del intervalo de los contornos, sin usar color o sombra. Si los intervalos de contorno son del mismo tamaño, se puede dibujar sólo las líneas de contorno e indicar que cada línea representa un incremento constante en magnitud, como se hace con muchos mapas. Los mismos datos usados en el Gráfico 3.12 se presentan en un gráfico de contornos en el Gráfico 3.13.

**Gráfico 3.13**



Los gráficos de contornos son mejores que los de superficie porque nunca se pierden datos debido a las colinas. Sin embargo se pierden datos porque para hacerlos se debe convertir una variable continua en intervalos discretos. Si los datos son discretos desde un comienzo, obviamente esto no sucederá. Algunas personas encuentran que los gráficos de contorno son más difíciles de entender que los de superficie (mientras otros piensan al revés). Como regla general se deben graficar los datos de las dos formas y después decidir cuál es la mejor para representar los datos con los que se trabaja.

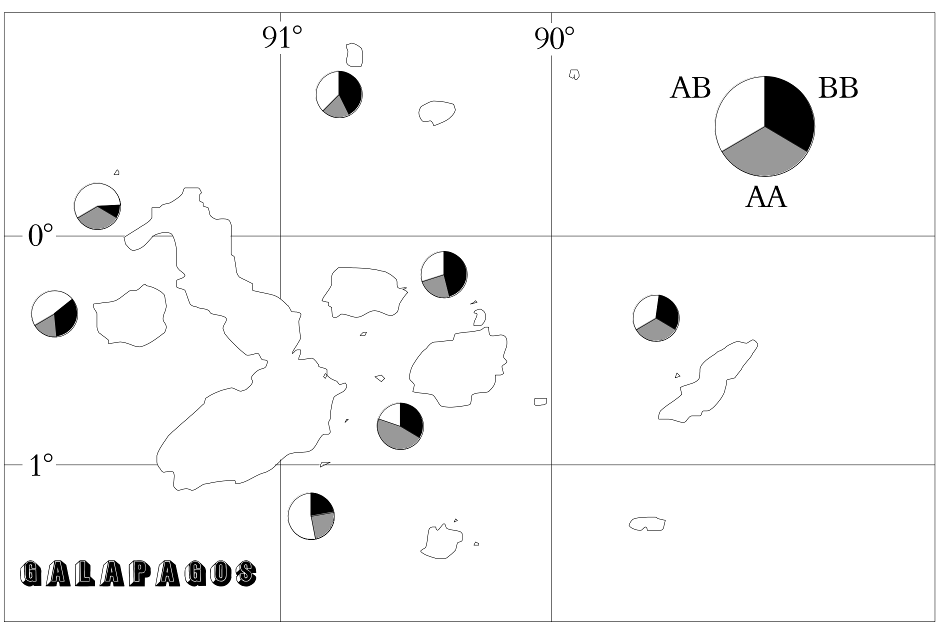
Variables con componentes espaciales:

Pregunta: ¿Cómo están distribuidos en el espacio?

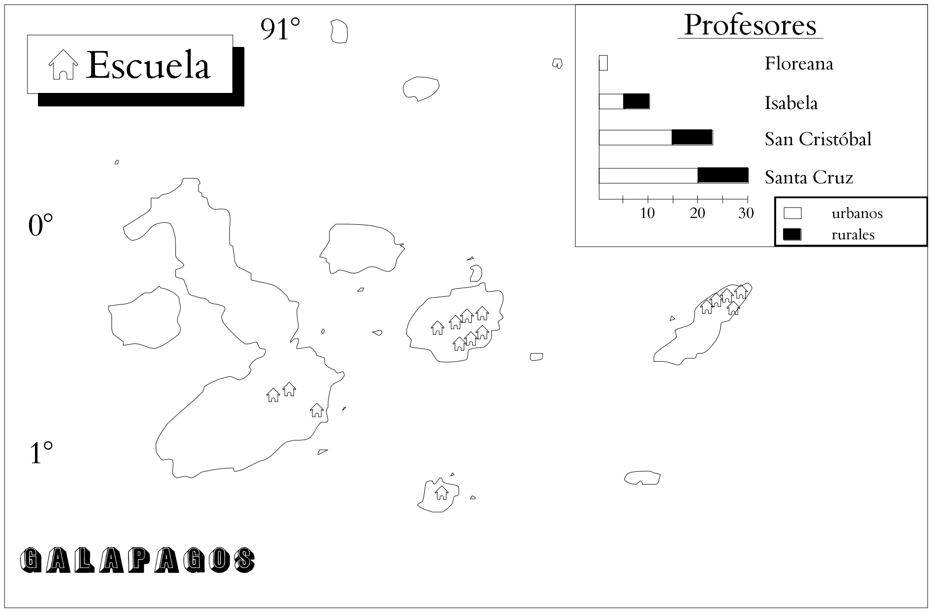
Finalmente se considerará una combinación especial de información geográfica (ej. mapas) y gráficos de todas las formas que se han estudiado. De tiempo en tiempo, se desea mostrar cómo ciertas variables cambian dependiendo de su situación geográfica. Uno de los casos más simples es mostrar la mera ocurrencia de un evento, como la localización de un animal. Si los puntos están bien separados, entonces es posible colocar un punto en un mapa para cada observación. Si los puntos están muy cerca, o varios ocurren en el mismo lugar, se puede indicar el número de observaciones mediante un símbolo (2, 3, 4, etc.) o por el tamaño del punto. Si hay un amplio rango de valores a indicar de esta manera, entonces se pueden usar rangos discretos de números representados por un símbolo dado.

Otras aplicaciones ocurren cuando, por ejemplo, ciertas variables son medidas en localizaciones discretas y es importante mostrar cómo varían estas variables. Los Gráficos 3.14, 3.15 y 3.16, muestran algunas de las formas cómo estos tipos de datos pueden ser combinados con mapas.

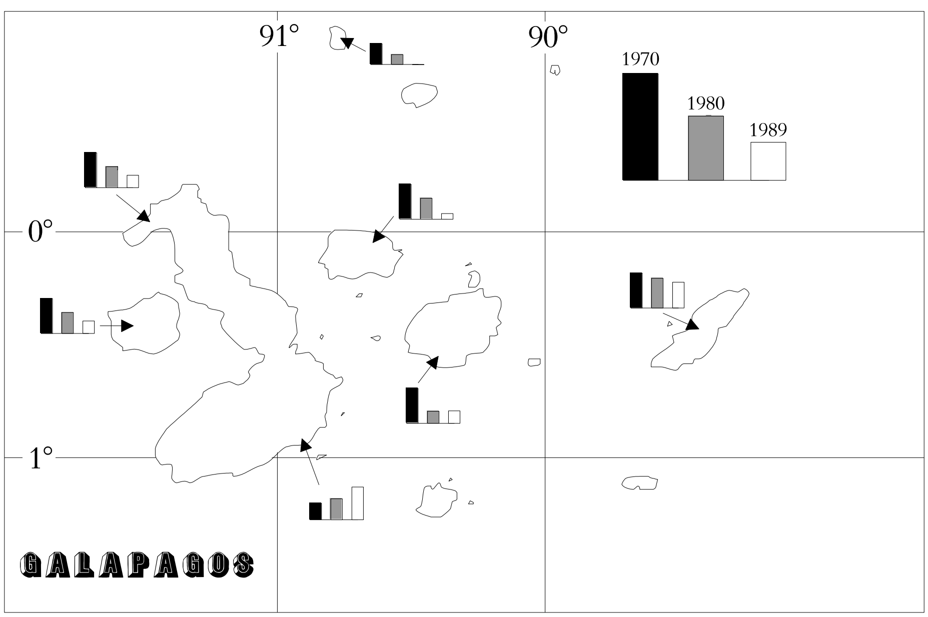
**Gráfico 3.14**



## Gráfico 3.15




## Gráfico 3.16



# Capítulo IV

# Introducción a la Probabilidad y a las Distribuciones de Probabilidad

## Introducción

Muchas veces a lo largo de este curso se han encontrado términos de aleatoriedad (números al azar, muestreo al azar, etc.) y probabilidad. Para entender mejor la estadística es necesario comprender estos conceptos y cómo se relacionan entre ellos y con las teorías estadísticas. La estadística debe su existencia a muchas ramas de las matemáticas, pero sin duda una de las más importantes es la del estudio de la probabilidad. Aunque una discusión completa de las reglas y teorías de probabilidad se extendería por cientos de páginas, hay unas pocas reglas principales que son básicas para el entendimiento de la estadística. En este capítulo se hará una breve introducción a estos importantes conceptos.

## Probabilidad

¿Qué tienen en común el lanzamiento al aire de una moneda, el rodar un dado, el escoger un papelito con un nombre de dentro de un sombrero, el pesar una iguana cualquiera, la edad de la octava persona encontrada en la calle cada día, la calidad de una torta seleccionada al azar? La respuesta es que todas esas cosas son ejemplos de variables aleatorias.

Las variables aleatorias, como toda variable, vienen en muchas formas y tamaños (véase el Capítulo I). La característica común de las variables aleatorias es que el resultado de cada muestra (experimento, prueba o evento) es un valor aleatorio. Es decir que solo se lo conoce cuando se obtiene la muestra.

Aunque el valor exacto de una variable aleatoria no se puede conocer con anterioridad, se puede buscar cierta información que permita intentar predecirlo. Por ejemplo, con frecuencia conocemos el rango de posibles valores que pueden resultar de una prueba (por ejemplo, se conoce que cada vez que rodamos un dado, el rango de valores posibles es 1, 2, 3, 4, 5 y 6). También algunas veces se sabe, o al menos se tiene una idea, de la posibilidad de que un valor dado ocurra. Se sabe que la probabilidad para cada número debe ser igual.[1]

Por el momento se considerarán solo variables aleatorias basadas en posibilidades discretas con rangos finitos. Estas están representadas por situaciones como lanzar una moneda, rodar un

---

[1] El hecho de que cada cara del dado debe tener la misma probabilidad de salir en una lanzada permite probar la ‘honestidad’ del dado mediante la comparación de las frecuencias observadas y esperadas para cada lado. Véase el Capítulo IV para mayor información sobre esta prueba.

dado, escoger de un sombrero un papel con un nombre, etc. Además, se asumirá por el momento que la probabilidad de obtener cualquier resultado es la misma. Así se tiene la primera regla:

> Si hay **n** posibles resultados con la misma oportunidad de ocurrir, la probabilidad de obtener cualquiera de ellos es:

$$\mathbf{1/n} \tag{4.1}$$

Por ejemplo, para el dado, el número de posibilidades (n) es 6. Por lo tanto, la probabilidad de obtener un lado cualquiera es 1/6 = 0.1667 (16.67%).

En estadística los resultados para variables aleatorias discretas se expresan en términos de "éxito" o "fracaso". Estos se pueden prestar a malos entendidos porque no se refieren a la calidad o al deseo de obtener un resultado en particular. Es más, son términos relativos que dividen los resultados en aquellos de interés ('éxito') y todas las demás ('fracaso'). Por ejemplo, cuando se consideran los lanzamientos de una moneda se puede estar interesado solamente en la probabilidad de obtener caras. De acuerdo a lo dicho, en este caso, sería 'éxito' obtener una cara y 'fracaso' obtener sellos al contrario.

Se puede generalizar el concepto de 'éxito' definiéndolo como la ocurrencia del resultado de interés entre todos los posibles resultados. Por ejemplo, el caso de lanzar un dado, un 'éxito' podría estar indicado por un 1, un 2, un 3, o igualmente un 4, un 5 o un 6, o cualquier otra combinación arbitraria en la que se esté interesado.

Esto lleva a la segunda regla:

> si existen **n** resultados igualmente posibles de los cuales uno deba ocurrir y **s** de ellos son considerados como un 'éxito', entonces la probabilidad de obtener un 'éxito' es:

$$\mathbf{s/n} \tag{4.2}$$

En estadística se usa la letra **p** para indicar la probabilidad de 'éxito'. La letra **q** es usada para indicar la probabilidad de 'fracaso'. La probabilidad de "éxito' más la de 'fracaso' (p + q) <u>debe</u> ser igual a 1. Dado esto, se tiene:

$$p + q = 1 \tag{4.3}$$

Arreglando:

$$q = 1 - p \qquad p = 1 - q \tag{4.4}$$

En otras palabras, la probabilidad de que un evento no suceda (q) es igual a 1 menos la probabilidad de que suceda (1-p); la probabilidad de que un evento suceda (p) es igual a 1 menos la probabilidad de que no suceda.

<u>El Triángulo de Pascal</u>:

La discusión previa describe las probabilidades para una sola muestra o ensayo. Ahora se considerará lo que pasa cuando se repite un experimento varias veces.

Primero se debe expandir la definición de una variable aleatoria para incluir el concepto de **independencia estadística**Error! Bookmark not defined.. Dos eventos son estadísticamente independientes cuando el resultado de un ensayo no se ve afectado (o viceversa) por el otro. Esto conduce a otra regla de probabilidad:

> Si dos resultados provienen de ensayos que son aleatorios e independientes, la probabilidad de observar exactamente estos dos resultados juntos es el producto de las probabilidades separadas de los dos resultados. (4.5)

Por ejemplo, si la probabilidad de capturar una tortuga hembra ('éxito')[2] es 0.2 la probabilidad de capturar dos hembras seguidas sería:

$$p \text{ x } p = p^2$$

$$0.2 \text{ x } 0.2 = 0.04$$

Lo mismo sería verdad para capturar dos machos[3]

$$q \text{ x } q = q^2$$

$$0.8 \text{ x } 0.8 = 0.64$$

La probabilidad de capturar un macho y una hembra sería:

$$2pq = 0.32$$

Pero ¿por qué se multiplica pq (la probabilidad de capturar una hembra multiplicada por la probabilidad de capturar un macho) por 2? La respuesta es porque hay dos formas de capturar un macho y una hembra. Se puede capturar un macho y luego una hembra, o una hembra y luego un macho. Matemáticamente lo que se obtiene es:

$$qp + pq = 2pq$$

Nótese que la suma de las probabilidades $p^2 + 2pq + q^2$ (0.04 + 0.32 + 0.64) es igual a 1. Después de todo, una de esas tres posibilidades DEBE ocurrir.

Se puede extender esta teoría desde dos a cualquier número de ensayos independientes. Considérese un solo dado con números del 1 al 6. La probabilidad de observar un número dado en una lanzada particular es 1/6 (véase la ecuación 4.1). ¿Cuál es entonces, la probabilidad de obtener el mismo número en cada una de las seis lanzadas?

De acuerdo con la regla 4.5 la probabilidad es:

---

2 La probabilidad se expresa bien como un valor entre 0 y 1, o como porcentaje entre 0 y 100%. Las dos formas son lo mismo. Si se usa porcentajes no se debe olvidar el escribir el signo %.

3 En este caso se debe referir a la probabilidad de capturar un macho como **q** menos la probabilidad de obtener una hembra. Dado que q = 1 - p, entonces q = 1 - 0.2 = 0.8.

$$\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = 1/46656$$
$$= 2.14 \text{ x } 10^{-5}$$

Como un ejemplo más, considérese que el imaginario pájaro turista de pecho rojo tiene dos fases: blanca y roja. Si se define la probabilidad de encontrar un ave blanca como **p** y la de encontrar un ave roja como **q**, se pueden escribir las probabilidades de encontrar todas las posibles combinaciones de fases a partir de una muestra de **n** aves.

| | | |
|---|---|---|
| Para n = 1 | Hay solo dos posibilidades: una blanca (probabilidad = **p**) o una roja (probabilidad = **q**). | |
| Para n = 2 | Hay tres posibilidades: | |
| | dos blancas | $\mathbf{p^2}$ |
| | una blanca + una roja (dos formas) | $\mathbf{2pq}$ |
| | dos rojas | $\mathbf{q^2}$ |
| Para n = 3 | Hay cuatro posibilidades: | |
| | tres blancas | $\mathbf{p^3}$ |
| | dos blancas + una roja (3 formas) | $\mathbf{3p^2q}$ |
| | una blanca + dos rojas (3 formas) | $\mathbf{3pq^2}$ |
| | tres rojas | $\mathbf{q^3}$ |
| Para n = 4 | Hay cinco posibilidades: | |
| | cuatro blancas | $\mathbf{p^4}$ |
| | tres blancas + una roja (4 formas) | $\mathbf{4p^3q}$ |
| | dos blancas + dos rojas (6 formas) | $\mathbf{6p^3q}$ |
| | una blanca + tres rojas (4 formas) | $\mathbf{4pq^3}$ |
| | cuatro rojas | $\mathbf{q^4}$ |

Nótese que el coeficiente (formalmente conocido como el **coeficiente binomial**) es igual al número de maneras de formar cada combinación. Aún más, la potencia a la que p y q se elevan es igual al número de aves de cada tipo en el grupo (ej. $\mathbf{6p^2q^2}$ significa que hay 6 maneras de formar grupos que contengan 2 aves blancas y 2 rojas). Una forma simple de determinar los coeficientes para cada combinación, en vez de tratar de encontrar todas las combinaciones, es construir el triángulo que se muestra en el Gráfico 4.1.

Este triángulo se conoce como el **Triángulo de Pascal**. Cada uno de los números corresponde al coeficiente de cada posible resultado de **n** observaciones. El triángulo comienza con n=0 lo que quiere decir que hay una sola forma de no obtener capturas. El próximo nivel (n=1) dice que hay dos formas posibles de obtener una captura (ambas con igual probabilidad). El próximo nivel (n=2) tiene tres posibilidades (2 machos, 1 macho/1 hembra, 2 hembras) con probabilidades relativas de 1, 2 y 1, respectivamente. Para n=3 hay 4 posibilidades, y así sucesivamente. Nótese que los coeficientes se pueden calcular fácilmente sumando el par de números en la línea de arriba, y después adicionando uno (1) en cada extremo.

## Gráfico 4.1

n

0 ......................................... 1
1 ................................... 1 1
2 ............................. 1 2 1
3 ....................... 1 3 3 1
4 ................. 1 4 6 4 1
5 ............. 1 5 10 10 5 1
6 ........ 1 6 15 20 15 6 1

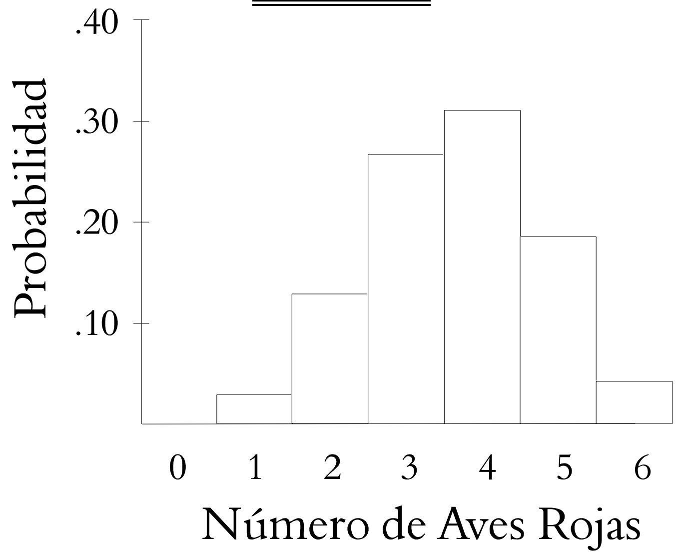
Al usar el Triángulo de Pascal se puede calcular la probabilidad de observar un resultado específico, entre un total de **n** posibilidades dados los valores de **p** y **q** (recuérdese que **p + q = 1**).

## Tabla 4.1

| Número | | Probabilidad | |
|---|---|---|---|
| Aves blancas | Aves rojas | simbólica | numérica |
| 6 | 0 | $p^6$ | 0.004096 |
| 5 | 1 | $6p^5q$ | 0.036864 |
| 4 | 2 | $15p^4q^2$ | 0.138240 |
| 3 | 3 | $20p^3q^3$ | 0.276480 |
| 2 | 4 | $15p^2q^4$ | 0.311040 |
| 1 | 5 | $6pq^5$ | 0.186624 |
| 0 | 6 | $q^6$ | 0.046656 |
| | | | 1.000000 |

Por ejemplo, se sabe que la probabilidad (p) de observar un ave blanca es 0.40 y que la de encontrar una roja es 0.60 (0.40 + 0.60 = 1.0). Si se capturan al azar 6 aves, la probabilidad para cada uno de los resultados es:
Gráficamente se vería lo siguiente:

**Gráfico 4.2**




Esta distribución es muy importante en estadística y se le ha dado un nombre especial: **Distribución Binomial**

# Distribución Binomial

La distribución binomial representa las probabilidades asociadas con la pregunta: "¿Cuál es la probabilidad de **X** 'éxitos' después de **n** intentos (pruebas)?", siempre y cuando la probabilidad de 'éxito' sea la misma para cada uno de un número fijo de intentos y considerando que todos los intentos son independientes. La probabilidad de 'éxito' puede ser cualquier valor entre 0 y 1.

Como se ve en la Tabla 4.1, se puede usar el número de 'éxitos'**Error! Bookmark not defined.** y 'fracasos**Error! Bookmark not defined.**' en el grupo, y el Triángulo de Pascal**Error! Bookmark not defined.** para calcular la probabilidad que tendría cada uno de los resultados posibles. Otra forma de calcular estos valores es haciendo uso de la siguiente ecuación:

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad (4.6)$$

donde: n es el número de muestras
x es el número de 'éxitos'
x! es el factorial[4]
p es la probabilidad de 'éxito'
q es la probabilidad de 'fracaso'

[4] El factorial de un número X se ha definido como: X(X-1)(X-2)...(2)(1). El factorial de cero (0!) se ha definido como 1. Por ejemplo: 8! = 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 = 40,320.

El valor $n!/[x!(n-x)!]$ es el coeficiente binomial**Error! Bookmark not defined.**, y como se dijo es solo otra manera de calcular los valores encontrados en el Triángulo de Pascal**Error! Bookmark not defined.**. En realidad es la manera matemática de formular la misma pregunta que se hizo en la página 45 de este capítulo: “Para una muestra con tamaño n, ¿Cuántas formas hay de conseguir cada combinación de ‘éxitos**Error! Bookmark not defined.**’ y ‘fracasos**Error! Bookmark not defined.**’?”. La segunda parte de la ecuación 4.6, $p^x q^{n-x}$, representa las potencias de p y q para cualquier combinación de ‘éxitos’ y ‘fracasos’ (se puede tratar de calcular la Tabla 4.1 usando la ecuación 4.6). Otra forma de calcular la Tabla 4.1 es calcular la expansión del binomial $(p+q)^n$ como se muestra a continuación:

## Tabla 4.2

| n | $(p+q)^n$ |
|---|---|
| 1 | $p + q$ |
| 2 | $p^2 + 2pq + q^2$ |
| 3 | $p^3 + 3p^2q + 3pq^2 + q^3$ |
| 4 | $p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$ |
| 5 | $p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$ |
| 6 | $p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + {}^6pq^5 + q^6$ |

Por obvias razones, la ecuación 4.6 es la forma más conveniente de calcular las probabilidades binomiales cuando n es mayor que 5 o 6.

Los datos de la Tabla 4.1 y el Gráfico 4.2 pueden ser usados para formular algunas preguntas interesantes sobre ciertos tipos de resultados consecuencia de la captura de seis pájaros turistas. Por ejemplo, ¿cuál es la probabilidad de capturar exactamente 4 aves rojas?, ¿menos de tres aves blancas?, ¿más de dos y menos de seis aves rojas?

La respuesta a la primera pregunta se determina fácilmente mirando la tabla de valores para el número de aves rojas = 4. Se puede ver que la probabilidad es igual a 0.31104 (véase el Gráfico $4.3_a$).

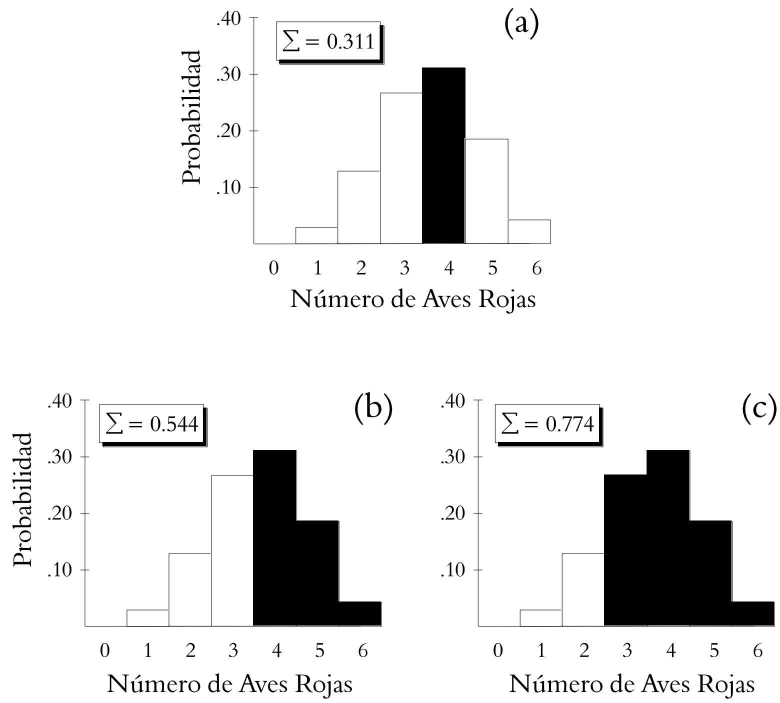
La segunda pregunta es un poco más difícil. Menos de tres aves blancas incluye las observaciones de 0, 1 y 2 aves blancas. De ahí la necesidad de sumar las probabilidades de cada una de estas posibilidades. El total es igual a 0.544320 (véase el Gráfico $4.3_b$).

De forma similar, la solución para la tercera pregunta se obtiene sumando las probabilidades para el número de aves turísticas rojas = 3, 4 y 5. El total es igual a 0.774144 (véase el Gráfico $4.3_c$).

¿Cuál sería la probabilidad de observar ya sea un ave roja o un ave blanca en un grupo de 6?

Existe un número infinito de distribuciones binomiales**Error! Bookmark not defined.**, una por cada combinación única de p, q y n.

**<u>Gráfico 4.3</u>**




La distribución binomial**Error! Bookmark not defined.** tiene algunas propiedades importantes:

1. El número promedio (media) de 'éxitos' de una distribución binomial**Error! Bookmark not defined.** ($\mu_X$) es: **np**

2. La varianza ( $\sigma_X^2$ ) es: **npq**

En el ejemplo de los pájaros turistas donde n=6, p=0.2 y q=0.8, La media $\mu_x$ = 0.12 y la varianza $\sigma^2$ = 0.96.

3. Cuando p = q = 0.5 y n tiende a $\infty$, la distribución binomial**Error! Bookmark not defined.** es igual a la **Distribución Normal**.**Error! Bookmark not defined.**

Además de la distribución binomial**Error! Bookmark not defined.** existe un número infinito de distribuciones de probabilidad. Sin embargo, en Biología se trabaja generalmente con las tres distribuciones siguientes: la binomial, la de Poisson y la normal. Otras distribuciones como la uniforme, la multi-nomial, la circular, la 'log-normal', la geométrica y la hypergeométrica, también se pueden encontrar, pero con menos frecuencia.

# La Distribución de Poisson

A diferencia de la distribución binomial**Error! Bookmark not defined.**, que describe las probabilidades de un número fijo de 'éxitos', la distribución de Poisson**Error! Bookmark not defined.** describe las probabilidades de ocurrencias aleatorias. Estas ocurrencias están representadas por objetos que se encuentran en un espacio o eventos que ocurren en el tiempo. Para objetos distribuidos aleatoriamente (plantas, animales, rocas, etc.) se asumirá que cada porción de espacio (ej. un cuadrante) tiene una oportunidad igual de contener ese objeto y que la presencia o ausencia de un objeto en una porción de espacio ninguna manera influencia la presencia o ausencia de otro objeto en cualquier otra área. Una distribución aleatoria de eventos en el tiempo es una en la cual un periodo dado de tiempo tiene una probabilidad igual de presenciar un evento. De igual forma, la ocurrencia de un evento en un periodo de tiempo dado en ninguna manera influirá la ocurrencia de otro evento en otro periodo de tiempo.

En teoría se conoce que la distribución de Poisson**Error! Bookmark not defined.** tiene la siguiente forma:

$$P(X) = \frac{e^{-\mu}\mu^{x}}{X!} \qquad (4.7)$$

donde: e es el logaritmo natural
μ es el número promedio de 'éxitos' u ocurrencias
X es el número de 'éxitos'
X! es el factorial de X.

'P(X)=' debe leerse "la probabilidad de observar X 'éxitos' es igual a"

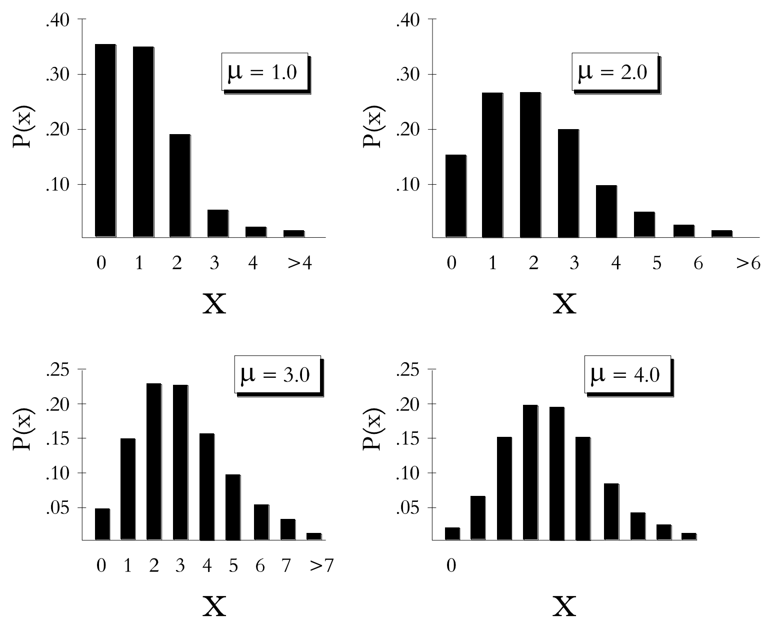
En realidad, esta ecuación no es tan complicada como parece a primera vista. El valor de μ es el número promedio de 'éxitos' que se esperan en cada prueba o muestra. Por ejemplo, imagínese que se han tomado muestras al azar de agua marina y se está contando el número de un tipo particular de organismo marino (donde cada individuo observado se considera como un 'éxito') en la muestra. La μ se puede calcular en base al número promedio de individuos observados luego de n muestras. Usando la ecuación 4.7 se puede calcular, el número de muestras que se espera que contenga X (0, 1, 2, 3, etc.) individuos **SI** los individuos están distribuidos aleatoriamente.

El primer paso en este proceso es calcular el valor de la ecuación 4.6 para varios valores de X.

$$P(X=0) = e^{-\mu} \qquad 4.8a$$
$$P(X=1) = e^{-\mu}\mu \qquad 4.8b$$
$$P(X=2) = (e^{-\mu}\mu^{2})/2 \qquad 4.8c$$
$$P(X=3) = (e^{-\mu}\mu^{3})/6 \qquad 4.8d$$
$$P(X=4) = (e^{-\mu}\mu^{4})/24 \qquad 4.8e$$

donde: P(X=0) es la probabilidad de no ocurrencias en la muestra, P(X=1) es la probabilidad de una ocurrencia, etc. Debería estar claro que cada valor de X depende solamente de μ y que hay una sola distribución asociada con cada valor de μ. Algunas de estas distribuciones se muestran a continuación en el Gráfico 4.5.

**Gráfico 4.4**




La distribución de Poisson**Error! Bookmark not defined.** también tiene algunas propiedades interesantes:

1. La varianza y la media son iguales: $\mu = \sigma^2$

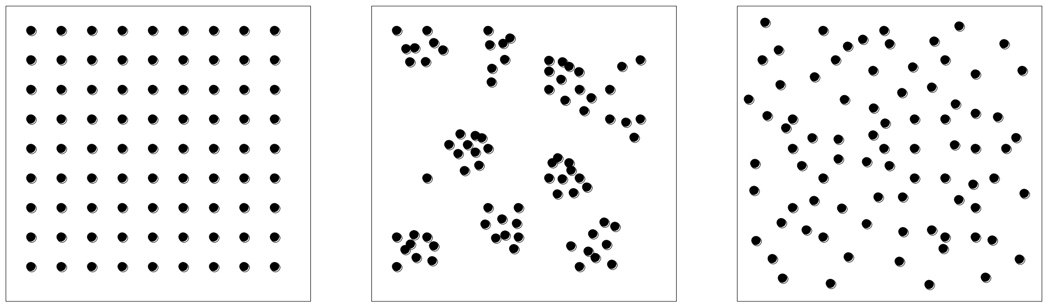
2. Llega a ser muy similar a la distribución normal cuando n es muy grande y p muy pequeño.

Mientras la distribución de Poisson**Error! Bookmark not defined.** tiene un rango entre 0 (no es posible observar menos de 0 ocurrencias de algo) a infinito, generalmente no es necesario hacer cálculos muy alejados del valor medio dado que las probabilidades caen fácilmente alrededor de este punto.

En bioestadística la distribución de Poisson**Error! Bookmark not defined.** se usa generalmente cuando se desea conocer si algún fenómeno se distribuye aleatoriamente o no. Hay dos formas diferentes de probar si hay desviaciones de una distribución al azar. La primera depende del hecho de que $\sigma^2=\mu$ para la distribución de Poisson. Si $\sigma^2<\mu$ entonces la distribución es más uniforme que aleatoria; Si $\sigma^2>\mu$ entonces la distribución es más agrupada, o contigua que aleatoria; y si $\sigma^2=\mu$ entonces la distribución es aleatoria. Por ejemplo, si se están muestreando plantas usando cuadrantes, pueden ocurrir las tres posibilidades que muestra el Gráfico 4.6.

La segunda forma de probar si hay desviaciones del azar utiliza la prueba del Chi-cuadrado**Error! Bookmark not defined.** (o la prueba G**Error! Bookmark not defined.**). En este caso los valores ‘esperados’ son aquellos calculados en base a la ecuación 4.7 y el valor de μ es estimado a partir de la muestra. Los valores ‘observados’ son los números de muestras que contienen 0, 1, 2, etc. ‘éxitos.’ Véase el Capítulo 9 para obtener mayores detalles sobre esta prueba.

**Gráfico 4.5**



$\sigma^2<\mu$ $\sigma^2>\mu$ $\sigma^2=\mu$

# La Distribución Normal

La última distribución que se tratará en este capítulo es la llamada **distribución Normal**Error! Bookmark not defined. o distribución **Gaussiana**. Esta distribución es sumamente importante en estadística ya que forma parte de la mayoría de pruebas estadísticas conocidas como **pruebas paramétricas** (véase el Capítulo 5).

Como ya se mencionó, una de las propiedades de la distribución binomial**Error! Bookmark not defined.** es que cuando p es igual a q y n es muy grande, la distribución binomial se vuelve muy similar a la distribución normal. Lo que se encuentra, es que para muchas situaciones biológicas la distribución de los datos toma frecuentemente forma de campana. Con suficientes muestras de gran tamaño, estas distribuciones se convierten fácilmente en distribuciones normales. La ecuación para una curva normal**Error! Bookmark not defined.** es dada por:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)_i^2/2\sigma^2} \qquad (4.9)$$

donde: $\mu$ es el valor de la media**Error! Bookmark not defined.**
$\sigma$ es la desviación estándar**Error! Bookmark not defined.**

La distribución normal se basa en variables **continuas**Error! Bookmark not defined., variables cuyo rango de valores (al menos en teoría) es $-\infty$ y $+\infty$. Las dos distribuciones anteriores, binomial y Poisson, están basadas en variables discretas. El rango de la distribución binomial**Error! Bookmark not defined.** está estrictamente determinado por el número de muestras, mientras la distribución de Poisson**Error! Bookmark not defined.** tiene un rango entre 0 y $+\infty$ que contiene solo números enteros.

Hay un número infinito de distribuciones normales, una por cada combinación de μ y σ. En general, la distribución normal se puede describir en términos de μ y σ (véase el Gráfico 4.6). La localización del pico central y la expansión de la distribución varían de acuerdo a los valores de μ y σ respectivamente (véase el Gráfico 4.7).

**Gráfico 4.6**

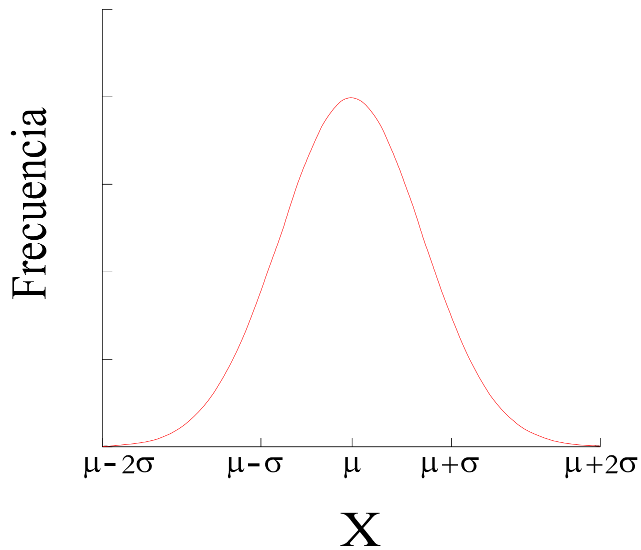
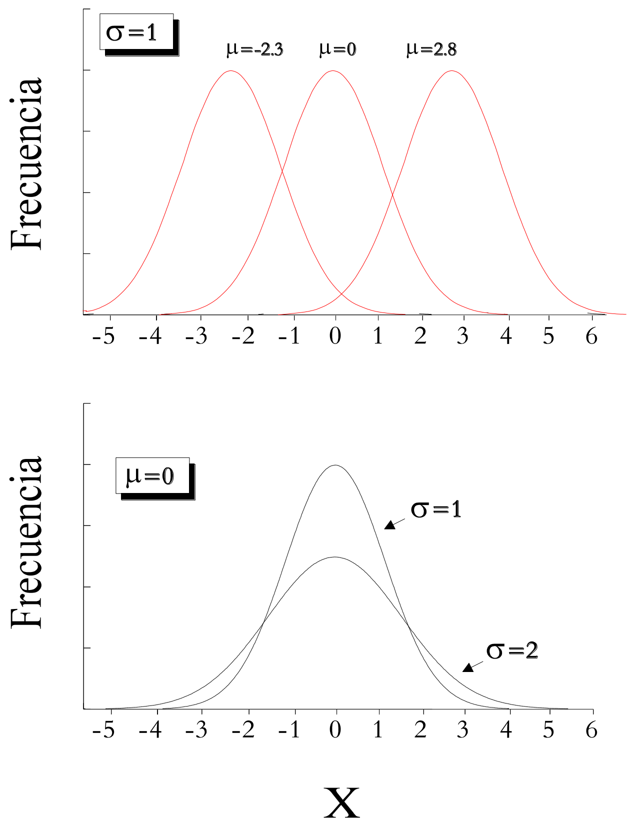



**Gráfico 4.7**




La distribución normal en donde μ=0 y σ=1 se dice que está **estandarizada** o se llama **distribución Normal Estándar**Error! Bookmark not defined.. Sustituyendo μ=0 y σ=1 en la ecuación 4.9 se tiene:

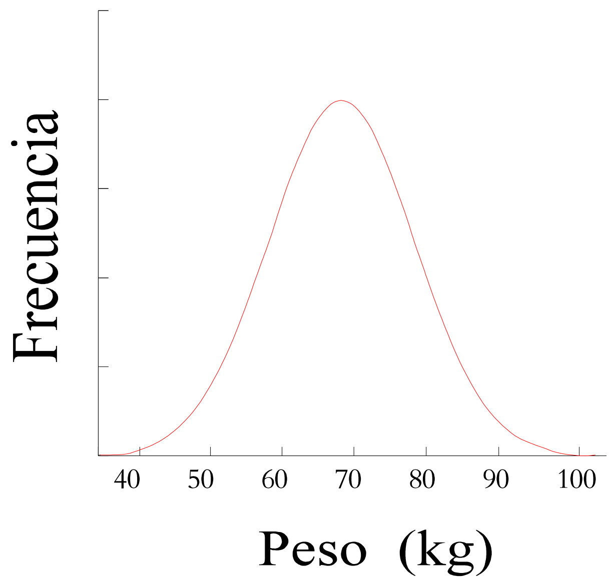
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad (4.10)$$

# Proporciones de la Distribución Normal

Al igual que con la distribución binomial**Error! Bookmark not defined.**, la distribución normal se puede usar para formular algunas preguntas interesantes sobre la probabilidad de obtener ciertos tipos de resultados. Cuando se trató la distribución binomial se dijo que un 'éxito' era algún valor o grupo de valores discretos (ej. >3 en el lanzamiento de un dado). Con las distribuciones normales se pueden hacer preguntas similares basadas en variables continuas.

Se considerará un ejemplo hipotético de una población de tortugas gigantes cuyos pesos corporales totales están distribuidos normalmente con una media (μ) de 70 kg. y una desviación estándar (σ) de 10 kg. Esta distribución se muestra a continuación en el Gráfico 4.8.

**<u>Gráfico 4.8</u>**




¿Qué se puede decir acerca de esta población? Dado a que las distribuciones normales son simétricas alrededor del valor de la media, se puede decir que tantas tortugas pesan más de 70 kg. como menos de 70 kg. Se puede decir, por ejemplo, que todas las tortugas que pesan más de 80 kg. pesan más de una **desviación estándar**Error! Bookmark not defined. más que el peso promedio. De igual forma, aquellas que pesan menos de 50 kg. pesan menos de dos desviaciones estándar menos que el valor promedio. Sin embargo, no se puede decir que <u>porcentaje</u> de individuos tienen un peso mayor o menor a uno determinado. Para poder hacerlo se necesita hacer uso de la siguiente ecuación:

$$Z = \frac{X - \mu}{\sigma} \qquad (4.11)$$

La ecuación 4.11 permitirá conocer a cuantas desviaciones estándar del promedio está un valor X. Esto se conoce como **normalización** o **estandarización** de X; y Z se conoce como la **desviación normal**Error! Bookmark not defined., ***standard score***, o el **valor z**.

Al usar los valores anotados en la tabla <u>Proporciones de la Curva Normal</u> (p. 131) y la ecuación 4.11, se puede calcular la proporción de una distribución normal**Error! Bookmark not defined.** que cae más allá de un valor dado de X. Los valores de esta tabla son las proporciones de la distribución mayores o iguales a un valor dado de Z.

Como ejemplo, si μ=70 kg. y σ=10 kg., ¿Qué proporción de tortugas tienen un peso mayor o igual a 80 kg.?

$$Z = (80 - 70)/10$$

= 1.00

El valor de Z = 1.0 en la tabla se busca usando primero la columna de la derecha para encontrar la línea correspondiente al primer decimal de Z, y usar entonces la fila a lo largo de la parte de arriba para localizar la columna correspondiente al segundo decimal de Z. Se encuentra que para Z = 1.00 la proporción es 0.1562. Esto quiere decir que 15.62% de la población tiene un peso mayor o igual a 80 kg. (Véase el Gráfico 4.9).

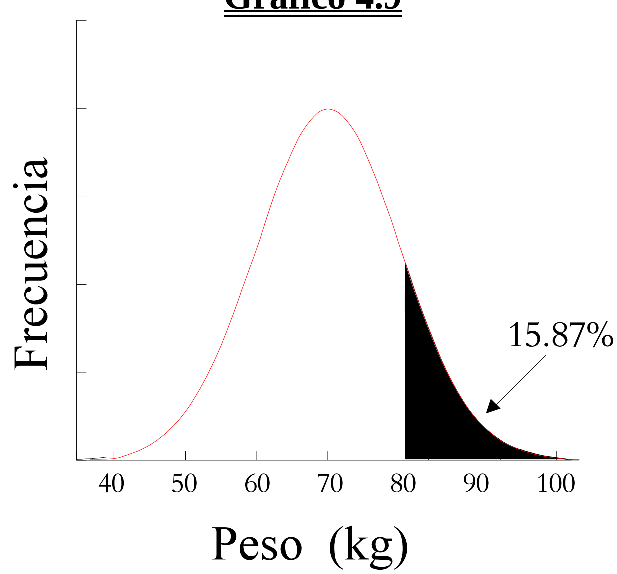
Si la diferencia X-µ es negativa (ej. ¿Qué proporción tiene un peso menor o igual a 50 kg? Z = (50-70)/10 = -2.00) se debe buscar la proporción correspondiente para Z = 2.00, recordando que ahora esto corresponde a una proporción menor o igual al valor de X. A partir de la tabla se encuentra que 2.28% de la población pesa menos de 50 kg.

También se puede preguntar qué proporción de la población pesa más que 50 kg. Ya se sabe que 2.28% pesa <50 kg. Para responder a esta pregunta solo se necesita recordar el uso de los términos ‘éxito’ y ‘fracaso.’ Si se llama a la proporción de los que pesa ≤ 50 kg. ‘fracaso’ y a aquellos que pesan > 50 kg. ‘éxito’, se puede aplicar la ecuación 4.4. Dado q = 0.0228, p = 1-0.0228 = 0.9772. En otras palabras, 97.72% de la población pesa > 50 kg.

El valor Z = 1.96 tiene un significado especial en estadística y se encuentra muy a menudo. Este valor corresponde a una proporción de 2.5% bajo la curva normal (véase Gráfico 4.10). ¿Por qué es este valor especial? Porque este es el valor que corresponde al famoso nivel de significancia**Error! Bookmark not defined.** de 5%.

En estadística frecuentemente se está interesado en saber si una observación específica es o no ‘rara’. Por ejemplo, se puede saber que el peso promedio de las tortugas es 70 kg. Supóngase que se captura una tortuga que pesa 100 kg., la pregunta sería: ¿es posible que esta tortuga pertenezca a la misma población en la que el peso promedio es 70 kg. (Desviación estándar = 10 kg.)?

**Gráfico 4.9**



Para responder a esta pregunta existe el problema de que los extremos de una distribución normal van de $+\infty$ hasta $-\infty$ y por lo tanto, al menos en teoría, cualquier valor puede ocurrir. Sin embargo, ¿es posible que una tortuga de 1000 kg. pertenezca a la población de la que estamos hablando? Probablemente no. Pero, ¿qué pasa con una tortuga que pese 200 kg.? ¿100 kg.? ¿90 kg.? ¿etc.? De alguna forma hay que decidir si una observación tiene suficiente probabilidad de pertenecer o no a una población. Esto se hace calculando el nivel de probabilidad mediante la ecuación 4.9. En Biología se usa generalmente el nivel de probabilidad del 5% como el nivel de **significancia**Error! Bookmark not defined.. Esto quiere decir que si hay una probabilidad del 5% o menos de que la observación pertenezca a la población de interés, se concluye que <u>probablemente no pertenece</u> a esa población. El valor de Z que indica el nivel de significancia es llamado el **valor crítico**Error! Bookmark not defined..

En lugar de usar el valor de Z que ofrece la proporción del 5% (1.645) se usa el valor que da el 2.5% (1.96). Esto se hace porque en pruebas en las que son importantes tanto los valores mayores como los menores $\mu$ se deben incluir los dos lados de la curva normal. Así, se incluye el 2.5% de las dos ‘colas’ una a cada lado de la distribución. Al sumar los valores de cada cola: 2.5 + 2.5 obtenemos el 5% para la distribución total (véase el Gráfico 10).

**Gráfico 4.10**