MANUAL DE APOYO DE PRÁCTICAS DE ECOLOGÍA

(2º de Grado en Ciencias Ambientales)

DEPARTAMENTO DE CIENCIAS DE LA VIDA

UNIDAD DOCENTE DE ECOLOGÍA

UNIVERSIDAD DE ALCALÁ

Pilar Castro (coordinadora), Álvaro Alonso, Asunción Saldaña, Margarida Santos, Paloma Ruiz-Benito.

ÍNDICE

I. PRESENTACIÓN

Este manual contiene materiales de consulta como apoyo a las prácticas de Ecología. Se estructura en los siguientes apartados.

I. Presentación
II. Métodos de investigación en Ecología. Este documento es una aproximación al método científico; resume métodos para diseñar experimentos controlados y muestreos de campo.
III. Métodos de análisis de datos en Ecología. Sintetiza los tests de estadística bivariante que utilizaremos a lo largo del curso.
IV. Elaboración de un trabajo científico en Ecología

Os recomendamos que tengáis a mano este manual en todas las sesiones de prácticas, ya sea en versión escrita o digital.

II. MÉTODOS DE INVESTIGACIÓN EN ECOLOGÍA

1. EL MÉTODO CIENTÍFICO EN ECOLOGÍA

El método científico arranca con un problema o pregunta, que puede derivar de observar un fenómeno de la naturaleza, de la necesidad de tomar decisiones de gestión, de un razonamiento, etc. Si los antecedentes publicados hasta el momento no permiten encontrar una respuesta satisfactoria, el siguiente paso es plantear una explicación provisional a la pregunta (hipótesis) y comprobar de forma empírica si estamos en lo cierto o no. Para que los esfuerzos de distintas personas contribuyan al desarrollo progresivo del conocimiento, hemos de seguir un método común, riguroso, repetible y comprobable. Este método se denomina "Método Científico" y se utiliza en un gran número de disciplinas. En cada una de ellas adquiere particularidades propias dependiendo del objeto de estudio. En nuestro caso veremos cómo el método científico se aplica en Ecología. Las etapas del método científico se resumen en la Figura II.1:

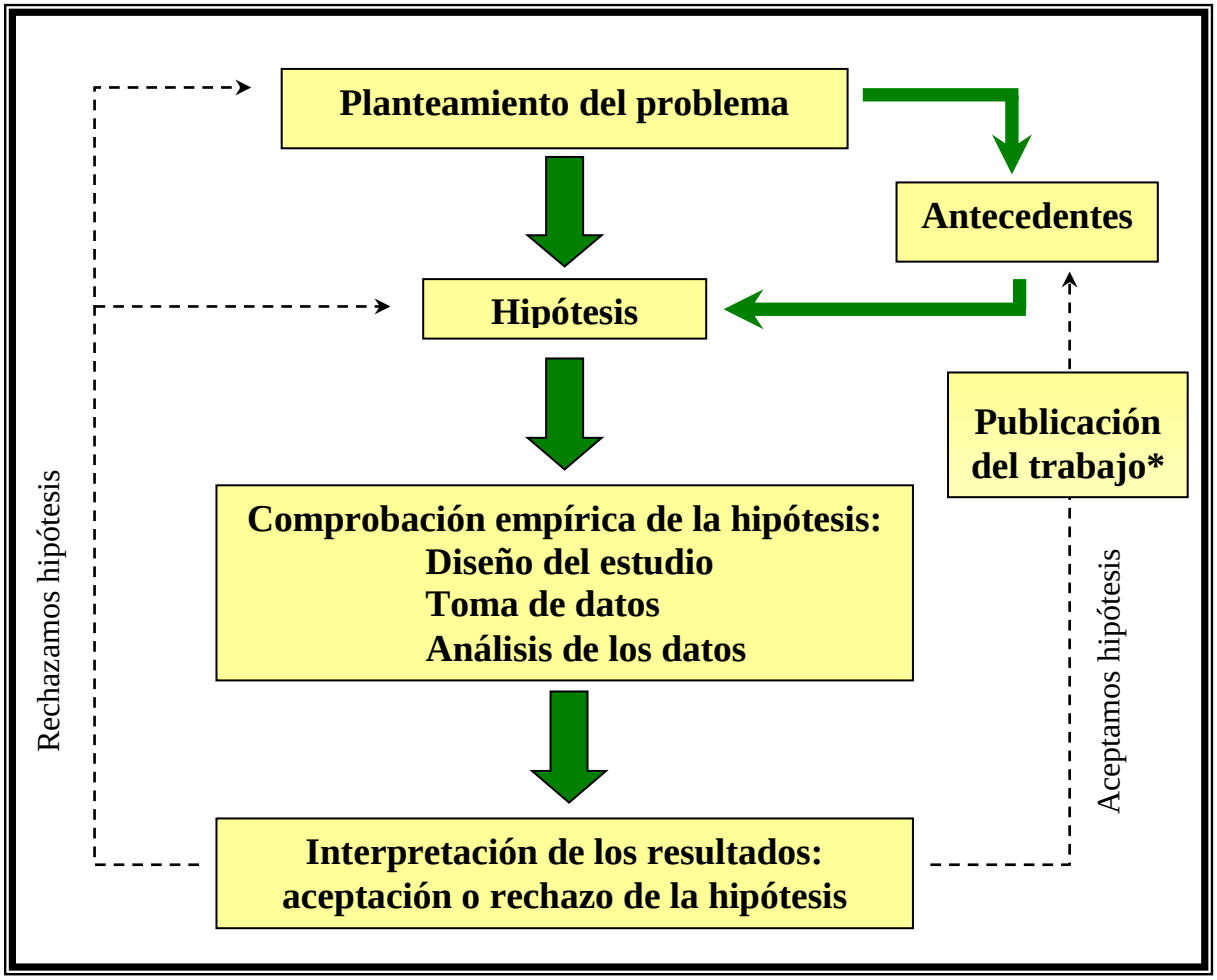

Figura II.1. *Etapas del método científico. *También se pueden publicar trabajos con las hipótesis rechazadas.*

Las hipótesis en Ecología suelen implicar una predicción sobre cómo una o varias variables independientes (o factores) afectan a una o más variables respuesta. La comprobación de hipótesis implica realizar muestreos de campo o experimentos donde se miden los factores y las respuestas para ver si coinciden con la predicción. Si finalmente aceptamos nuestra hipótesis, ésta dejará de ser una hipótesis para formar parte del cuerpo de conocimiento ya comprobado que otros científicos podrán consultar, previa publicación de nuestro estudio. El rechazo de la hipótesis implica que hay que plantear una hipótesis alternativa y comprobarla con un nuevo experimento o muestreo. Esto puede dar lugar al planteamiento de nuevas y más interesantes hipótesis. Este es el modo en que se construye la ciencia, por ensayo y error. Hay que subrayar que nunca se demuestra la veracidad de las hipótesis sino su falsedad, es decir, una interpretación o teoría se mantiene hasta que se demuestra que es falsa. Por último, para que se pueda construir un cuerpo de conocimiento cada investigador debe dar a conocer sus resultados mediante la publicación de un trabajo científico, donde se exponga la pregunta, la hipótesis, el proceso de comprobación de hipótesis y la interpretación de los resultados.

2. COMPROBACIÓN EMPÍRICA DE LA HIPÓTESIS

La comprobación empírica de la hipótesis es la fase clave en la respuesta a nuestra pregunta ecológica. La primera decisión a tomar se refiere a si el estudio que vamos a realizar es "experimental" u "observacional", es decir, si vamos a controlar o no los factores que afectan a la respuesta esperada (ver más abajo). En ambos casos habrá que determinar qué variables se van a medir, cómo se van a recoger los datos (es decir, qué tipo de muestreo o de experimento se va a llevar a cabo), y cuáles van a ser los análisis que se van a realizar para poder responder a nuestra pregunta. Una vez tomadas estas decisiones podremos hacer la recogida efectiva de los datos y su análisis (Figura II.2).

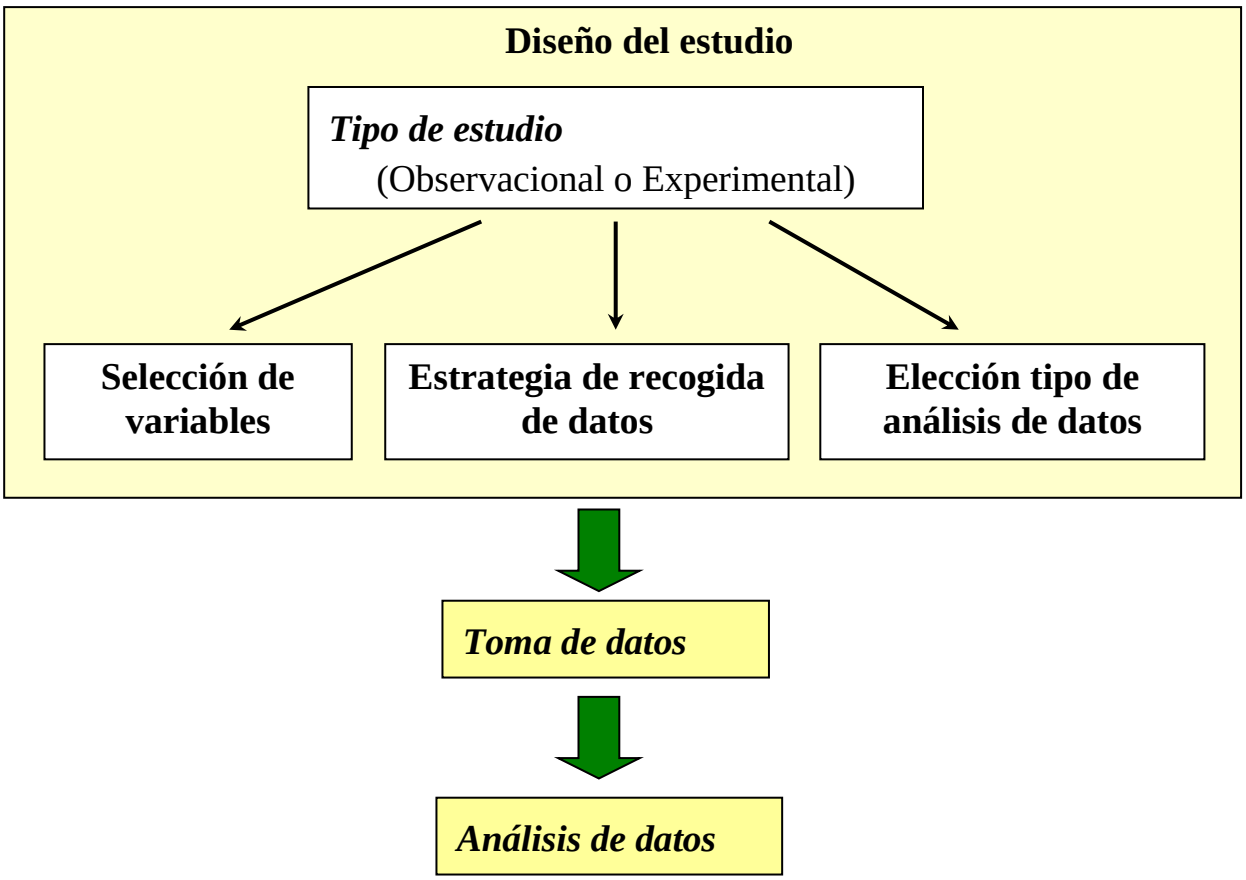

***Figura II.2**. Esquema detallado de la fase de valoración empírica de la hipótesis.*

2.1. Tipos de estudio

En Ecología solemos comprobar las hipótesis mediante estudios observacionales o experimentales. En realidad constituyen los dos extremos de un gradiente de control de los factores que esperamos que afecten a la variable respuesta. En el estudio observacional no hay control de estos factores, sino que el investigador se limita a registrar los valores de esos factores (como ocurre en muchos estudios de campo). En cambio en el estudio experimental el investigador controla los factores que espera que afecten a la respuesta, manteniendo constantes los que no le interesan (esto

último lo podemos hacer en condiciones de laboratorio). Cada uno presenta ventajas e inconvenientes, tal como se representa en la Figura 3. Según la pregunta que se pretenda responder será más conveniente uno u otro, o bien cualquiera de las diferentes posiciones a lo largo del gradiente entre ambos extremos (por ejemplo experimentos en campo).

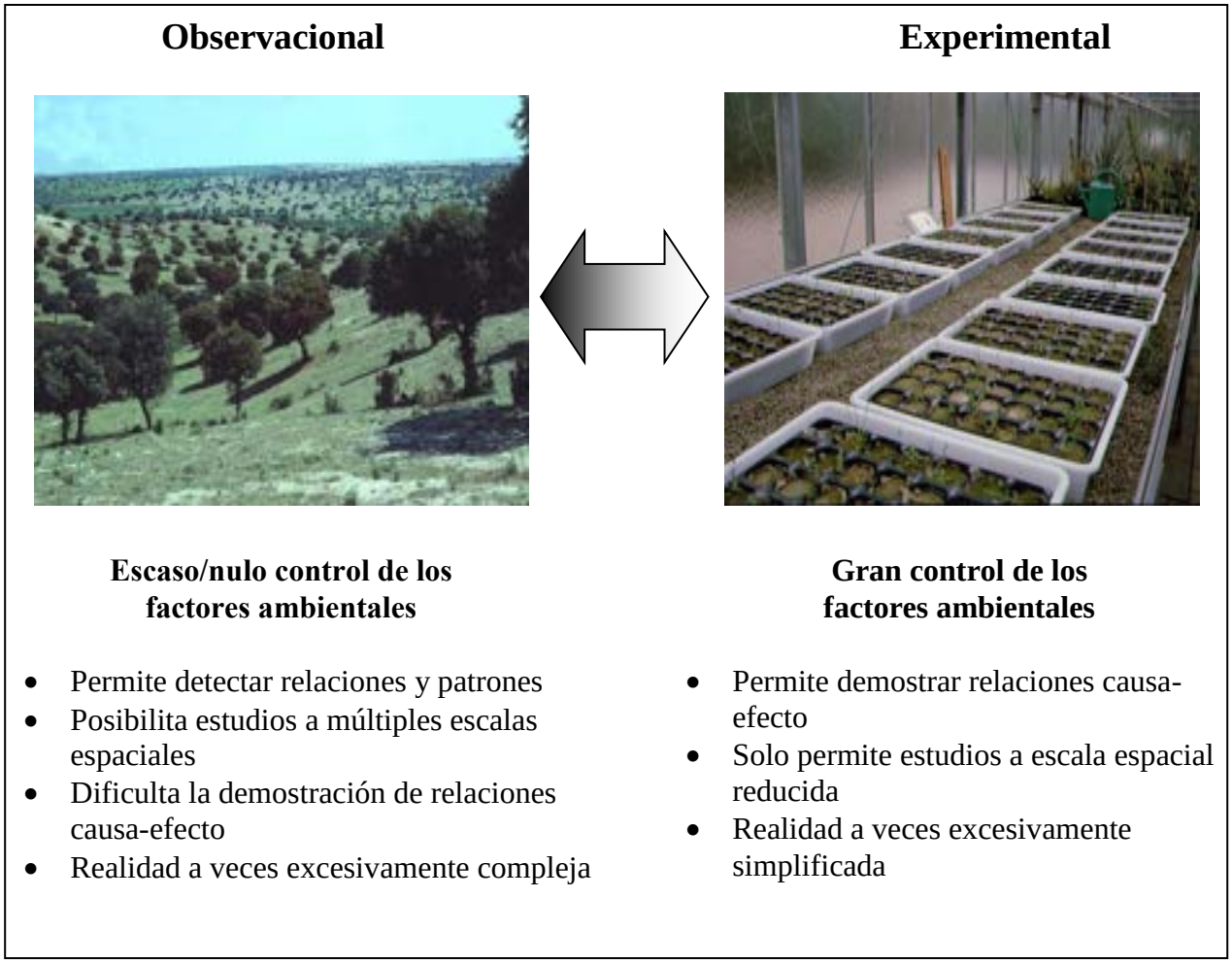

***Figura II.3**. Características de los estudios observacionales y experimentales*

3. SELECCIÓN DE VARIABLES

Una vez hemos decidido cuál es la mejor manera de abordar nuestro problema ecológico (experimento o muestreo de campo), tenemos que concretar qué variables vamos a considerar para responder más adecuadamente a nuestra pregunta. Las variables son características *observables* y *mensurables* que se desea estudiar (medir, controlar o manipular) y que toman diferentes valores. Se pueden clasificar principalmente de dos formas:

3.1 - Atendiendo al papel que cumplen en la hipótesis propuesta

INDEPENDIENTES o FACTORES: Son las que el investigador considera responsables de la respuesta esperada

DEPENDIENTES o RESPUESTA: Son las que el investigador mide para cuantificar la respuesta esperada

3.2- Atendiendo al tipo de medida que se les puede aplicar

Tabla III.1. *Tipos de variables que podemos encontrar en cualquier experimento o muestreo*

VARIABLES	DEFINICIÓN	SUBTIPOS	EJEMPLOS
CUALITATIVAS	Toman valores no numéricos	Dicotómicas	Sexo
		No dicotómicas	Raza, color de pelo
CUANTITATIVAS	Toman valores numéricos (al menos tres valores diferentes)	Discretas	Nº de descendientes
		Continuas	Peso

Ejemplos:

1. Hipótesis ecológica: la sombra que proyecta la vegetación reduce la evaporación del suelo y por tanto en zonas con vegetación el suelo estará más húmedo que en zonas sin vegetación.
   Variables: presencia/ausencia de vegetación (independiente, cualitativa con dos estados); humedad del suelo (dependiente, cuantitativa).
2. Hipótesis ecológica: el agua es el principal factor que limita el crecimiento de las plantas. Por lo tanto al aumentar la cantidad de agua que aportamos a las plantas aumentará su crecimiento.
   Variables: niveles de riego (alto, medio y bajo, independiente, cualitativa con tres estados); incremento en altura (dependiente, cuantitativa).

4. ESTRATEGIA DE RECOGIDA DE DATOS

Según hayamos decidido realizar un estudio observacional o experimental tendremos que decidir la estrategia de recogida de datos, es decir, tendremos que diseñar el experimento o el muestreo más adecuado para contestar a nuestra pregunta ecológica. Vamos a ver ahora en detalle los tipos de experimentos y de muestreos, y las decisiones a tomar en cada caso.

4.1. Diseños experimentales

Si hemos optado por un estudio experimental, en nuestra hipótesis consideraremos uno o varios factores (variables independientes) como causa del fenómeno que queremos estudiar y reproducir bajo condiciones controladas. El experimento más sencillo conlleva un solo factor, por ejemplo efectos del riego en el crecimiento de plantones de encina. Los pasos para establecer el diseño experimental son los siguientes:

1. Determinar los niveles del factor que controlamos. Por ej. podemos establecer en un cultivo tres intensidades de riego (alto, medio y bajo), dos niveles de fertilización (con/sin), cuatro niveles de temperaturas (10º, 15º, 20º y 25ºC), etc.

2. Determinar la unidad experimental básica, que es la unidad mínima sobre la que se aplica cada nivel de tratamiento/s. Puede ser un individuo, un grupo de individuos, una unidad de superficie, etc.
3. Asignar un número de unidades experimentales a cada tratamiento (número de réplicas). Este número ha de ser manejable pero representativo. Ha de ser proporcional a la variabilidad esperada entre unidades experimentales.
4. Establecer la distribución en el espacio de las unidades. Para ello hay tres maneras principales.
   - *Al azar*: Las unidades experimentales se distribuyen de forma totalmente aleatoria en el espacio disponible. Supongamos que queremos asignar a un grupo de 90 plantones de encina tres niveles de riego (los que aparecen en la figura con tres rellenos diferentes). La unidad experimental sería un plantón y asignaríamos 30 unidades a cada nivel de riego (Fig. III.4). Este diseño resulta adecuado cuando el espacio en el que se dispone el experimento es muy homogéneo.

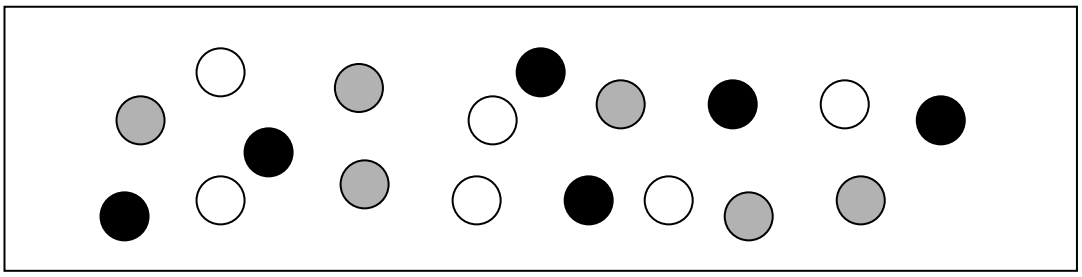

Figura II.4*. Distribución al azar de tres niveles de riego (blanco-alto, gris-medio, negro-bajo) entre unidades experimentales (círculos)*

   - *En bloques al azar*. Las unidades experimentales se agrupan en bloques (un bloque consta de una unidad de cada tratamiento distribuidas al azar) y los bloques se distribuyen en el espacio al azar o de forma regular. En el ejemplo anterior, cada bloque constaría de tres encinas, cada una de las cuales con un nivel de riego distinto, lo que daría un total de 30 bloques. Este diseño se utiliza cuando se sospecha que el área experimental no es homogénea (por ej. en un invernadero hay una pared que proyecta un gradiente de sombra, o puede que el sistema de riego no nos asegure un riego homogéneo en todo el invernadero) (Fig. III.5). Con este diseño la variabilidad ambiental no deseada se reparte por igual entre los tratamientos, para evitar artefactos en los resultados. El análisis estadístico que se aplica a este diseño permite cuantificar la varianza entre bloques y eliminarla.

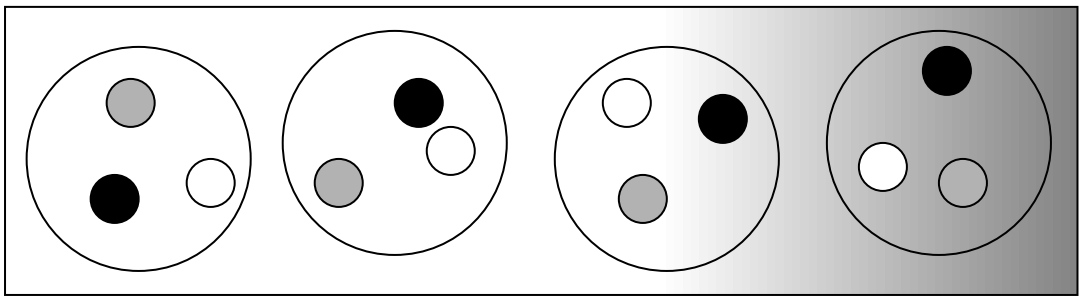

Figura II.5*. Distribución de unidades experimentales en bloques al azar para asegurar que el gradiente de sombra afecta por igual a todos los tratamientos (representados blanco, gris y negro)*

4.2. Diseño de muestreos observacionales

Cuando se ha optado por comprobar una hipótesis con un muestreo observacional, es necesario realizar un buen diseño del muestro antes de tomar ningún dato. Para ello hay que tomar una serie de decisiones, que componen la *estrategia de muestreo*.

La forma ideal de comprobar una hipótesis sería medir las variables implicadas en todos y cada uno de los individuos (o unidades) que componen la población a estudiar. Por ejemplo, queremos comprobar si existe relación entre la longitud del pico y el peso de los gorriones. La captura y medida de todos los gorriones de la población nos daría una certeza absoluta sobre nuestro problema, pero normalmente ésto resulta inviable. Lo que se hace en su lugar es tomar una **muestra** representativa dentro de un **universo de muestreo**, en este caso, la población. En otras palabras, realizamos un **muestreo**. Sobre esa muestra se toman las medidas y se comprueba la hipótesis con ayuda de los métodos estadísticos, que nos permiten cuantificar la probabilidad de cometer un error al extrapolar las conclusiones obtenidas sobre la muestra para el conjunto de la población.

En los estudios observacionales no existe un control de las variables, como ocurre en los estudios experimentales, por lo que cabe esperar una mayor variabilidad entre unidades. Es por ello por lo que es fundamental elegir una muestra suficientemente **representativa** de la población, compuesta por un número de **réplicas** adecuado. En el ejemplo anterior de los gorriones, necesitaremos tomar las medidas del peso corporal y de la longitud del pico en un número de individuos significativo, por ejemplo 100 individuos. En este caso la **unidad de muestreo** (sobre el que se toman las medidas) es el gorrión y el **número de réplicas** es 100.

En algunos casos el concepto de réplica no está tan claro. Por ejemplo, si queremos caracterizar el tamaño de las hojas de un bosque de encinar para compararlo con otro bosque, podemos realizar la replicación en dos niveles: por un lado el bosque está compuesto de árboles, pero cada árbol tiene un elevado número de hojas, siendo la hoja la unidad última donde tomamos la medida. En este caso es necesario diseñar el muestreo teniendo en cuenta esos dos niveles. Si elegimos 1000 hojas del mismo individuo y promediamos sus tamaños, no tendremos un valor representativo del bosque, ya que el individuo muestreado puede ser más grande o más pequeño de lo normal. Tampoco sería adecuado elegir una hoja en 1000 individuos distintos, ya que en este caso cada individuo quedaría pobremente representado con una única hoja. Sería más correcto elegir, por ejemplo, 100 árboles distribuidos por todo el bosque, y recoger de cada uno 10 hojas distribuidas por distintas partes de la copa. En este caso la **réplica** sería el individuo (100 réplicas), mientras que la hoja será una **pseudo-réplica** (1000 hojas). La forma correcta de analizar estos datos sería promediar las 10 hojas de cada individuo y utilizar las 100 réplicas en el análisis. Si utilizamos los 1000 valores como réplicas estaremos cometiendo un error de muestreo llamado **pseudo-replicación**.

En resumen, al diseñar un muestreo debemos seguir los siguientes pasos:

a) **Seleccionar las variables** a medir. Para realizar un adecuado diseño de muestreo, es fundamental tener claro desde el principio cuáles son las variables que se van a medir, cuáles son dependientes e independientes y si su naturaleza es cualitativa o cuantitativa. El tipo de variables condiciona el tamaño de la unidad de muestreo y el número de repeticiones que se pueden hacer. La siguiente sección se dedica al estudio de las variables bióticas más frecuentemente estudiadas en ecología.

b) Selección de la **unidad de muestreo**. Puede ser una superficie, un volumen, un individuo, etc. Además debemos saber si nuestro muestreo necesita de pseudo-réplicas.

c) **Número** de unidades de muestreo que se considera necesario (réplicas, y en su caso pseudo-réplicas). Se considera un número representativo de muestras cuando el valor del parámetro que se va a medir varía poco con la adición de nuevas muestras (Figura II.6).

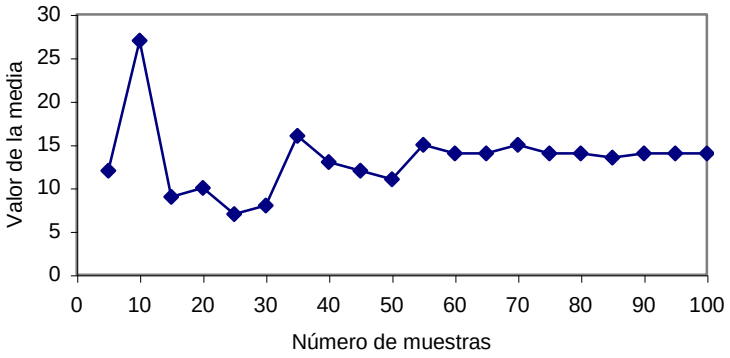

Figura II.6. *Valor de la media en función del número de réplicas. Se puede observar que cuando el número es bajo, la adicción de nuevas réplicas puede cambiar considerablemente la media, pero a partir de unas 60 réplicas la media ya no cambia al añadir más réplicas.*

4. Cómo se distribuyen las unidades en el espacio y en su caso en el tiempo. El objetivo de un muestreo es seleccionar una muestra donde todas las unidades sean igualmente independientes entre sí. Para ello hay distintos métodos. La selección de uno u otro dependerá de cómo es de homogéneo el universo de muestreo.

a) **Muestreo aleatorio simple o al azar**: cada elemento de la población tiene la misma probabilidad de ser elegido. Es apropiado en el caso de que el universo de muestreo sea homogéneo o no tengamos información que indique lo contrario. Para conseguir una verdadera selección al azar se pueden generar con el ordenador, o hacer un sorteo después de haber numerado todos los individuos de la población, etc. (Fig. III. 6).

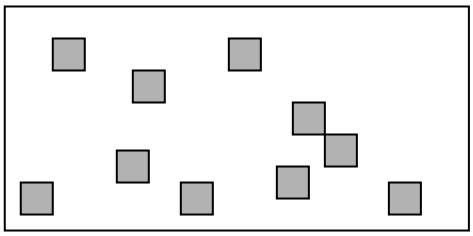

Figura II.6. *Distribución aleatoria de unidades de muestreo (por ej. cuadrículas UTM de un mapa).*

b) **Muestreo sistemático o regular**: las unidades de muestreo se distribuyen a intervalos regulares según un criterio preestablecido. Como en el caso anterior, se usa cuando el universo de muestreo es homogéneo, pero sospechamos que las unidades más próximas se parecen entre sí más que las más distantes. Por ejemplo, si en una plantación de pinos queremos conocer la altura media de los árboles, la probabilidad de que dos árboles muy próximos se vean afectados por condiciones microambientales similares (por ej. presencia de un arroyo cercano) es muy elevada, por tanto aplicamos un muestreo sistemático con una distancia mínima entre árboles para asegurarnos de

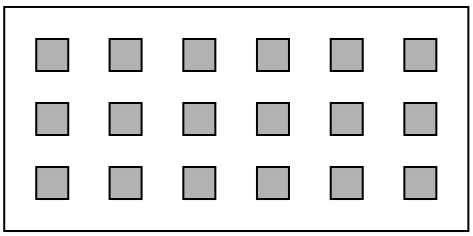

Figura II.7. *Distribución regular de unidades de muestreo (por ej. parcelas de 1 m^2 tomadas a distancias regulares).*

recoger condiciones microambientales diversas. (Fig. III.7).

c) Un caso especial del muestreo sistemático es el **transecto,** que se utiliza cuando la variable independiente (factor) varía gradualmente a lo largo del espacio. En este caso las muestras se han de distribuir a lo largo de ese gradiente, a intervalos regulares, con el fin de cubrir todo el rango de variación del factor. Por ej., para saber cómo afecta la altitud a la altura máxima de los pinos en la Sierra de Guadarrama, partiremos del límite inferior del bosque y seleccionaremos una unidad cada 100 metros de incremento de altitud, hasta llegar al límite superior del bosque (Fig. III.8).

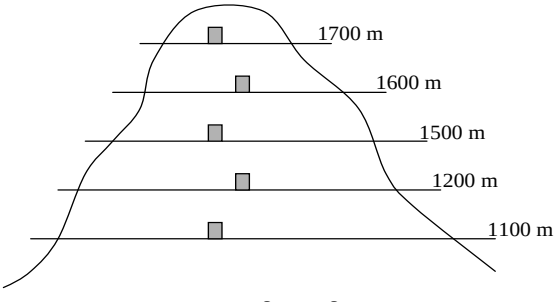

Figura II.8. *Distribución de unidades de muestreo a lo largo de un gradiente altitudinal*

d) Otro caso especial del muestreo sistemático es el de los **recorridos**, muy utilizados para muestrear la fauna. Se trata de establecer líneas que se recorren tomando notas a intervalos de tiempo o espacio previamente fijados (Figura II.9).

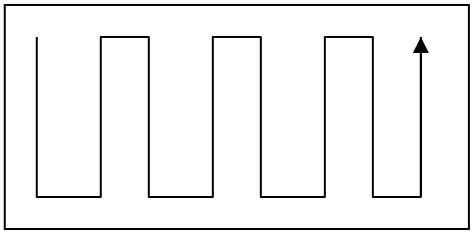

Figura II.9. *Recorrido para fauna marcado sobre un mapa*

e) **Muestreo sectorizado o estratificado**: se utiliza cuando la variable independiente, o factor, es de naturaleza cualitativa y permite dividir el universo de muestreo en estratos o sectores, que corresponden a cada uno de los niveles del factor. En cada estrato se toma un número similar de unidades (siguiendo alguno de los criterios anteriores) para asegurar que quedan igualmente representados en nuestra muestra. Por ejemplo, queremos saber cómo afecta la naturaleza del sustrato a la riqueza de especies de un pastizal. A partir de un mapa geológico, dividimos el área de estudio en tres sectores con distinta litología, y en cada uno disponemos 20 unidades de muestreo al azar (Fig. III.10).

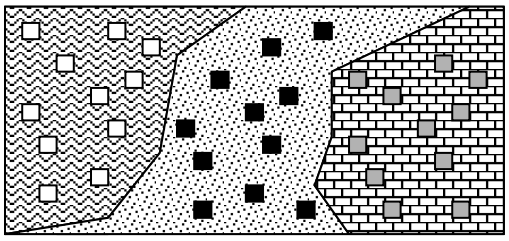

Figura II.10. *Ejemplo de muestreo sectorizado, donde el universo de muestreo se ha dividido en sectores litológicos y dentro de cada uno se han dispuesto las unidades de muestreo con un criterio regular.*

5. Organización de los datos. Antes de llevar a cabo el muestreo, es fundamental prever la organización de los datos que se van a tomar. Para ello hay que realizar una tabla o **estadillo**, que facilite la recogida eficaz de los datos. Algunos de los datos que conviene incluir en cualquier estadillo son: la fecha y la localidad de muestreo y el autor de las observaciones cuando hay varias

personas implicadas en el muestreo. Normalmente las variables se colocan en columnas y las unidades de muestreo en filas (Tabla II.2).

***Tabla II.2**. Ejemplo de estadillo diseñado para caracterizar una comunidad de invertebrados acuáticos*

Autor:		Fecha:		Localidad:		
	Tempera-tura	pH	Abundancia especie 1	Abundancia especie 2	...	...
Unidad 1	15	7.5	1	15	-	-
Unidad 2	12	7.3	3	48	-	-
Unidad 3	12	7.2	8	78	-	-
Unidad 4	13	8.2	4	23	-	-
Unidad 5	14	6.2	7	64	-	-
Unidad 6	17	6.8	2	85	-	-
Unidad 7	19	5.5	12	14	-	-
Unidad 8	15	7.1	8	15	-	-
Unidad 9	13	7.0	9	32	-	-

5. BIBLIOGRAFÍA RECOMENDADA

- Begon, M., Harper, J. L. y Townsend, C. R. 1996. Ecología: individuos, poblaciones y comunidades. 3ª Edición. Ediciones Omega, Barcelona.
- Mackenzie, A., Ball, A.S. and Virdee, S. R. 1998. Instant Notes in Ecology. Bios Scientific Publishers, UK.
- Sokal, R. R. and Rohlf, F. J. 1986. Introducción a la Bioestadística. Editorial Reverté.
- Zar, J.H., 1996. Biostatistical Analisys. 3rd Edition. Prentice-Hall International, London.

APÉNDICE II.I: FICHA PARA GUIAR EL DISEÑO DE UN ESTUDIO ECOLÓGICO

1. Planteamiento del **objetivo o pregunta.**

2. Planteamiento de **hipótesis y su justificación**

3. Diseño de estudio para comprobación de cada predicción:
   - *Tipo de estudio*: Observacional o experimental. Justificación.

   - *Variables* (en variables cualitativas, especificar las categorías que se incluyen)

VARIABLE(S) INDEPENDIENTE(S)	Cualit*. o cuantit	VARIABLE(S) DEPENDIENTES	Cualit*. o cuantit

*Si la variable es cualitativa, indica también las categorías de que consta

- *Protocolo de muestreo o experimento*

MUESTREO:
- ¿Cuál es la unidad de muestreo?
- ¿Cuántas réplicas se toman?
- ¿Cómo se distribuyen las réplicas en el espacio?

EXPERIMENTO:

- ¿En qué consiste la unidad experimental?
- ¿En qué consisten los tratamientos?
- ¿Cuántas réplicas hay en cada tratamiento?
- ¿Cómo se distribuyen los tratamientos en el espacio? (puedes ayudarte de un esquema).

- *Método estadístico* que se va utilizar:

4. Dibujo del **estadillo** apropiado para la toma de datos.

III. MÉTODOS DE ANÁLISIS DE DATOS EN ECOLOGÍA

1. INTRODUCCIÓN

La Estadística proporciona a la Ecología (y a otras ciencias experimentales) las herramientas necesarias para el análisis de los datos. Dado que no podemos hacer estudios en toda la población (no es posible contar todos los ácaros que hay en un suelo, ni medir el área foliar de todas las hojas de un bosque, ni medir la longitud de todas las carpas que tiene un lago), la estadística nos permite cuantificar la probabilidad de cometer un error al extrapolar los resultados obtenidos de una muestra al conjunto de la población.

La **estadística descriptiva** reúne un conjunto de técnicas que facilitan la organización, resumen y comunicación de datos; la **estadística inferencial** permite hacer pruebas de contraste de hipótesis.

1.1. Exploración de los datos

Cuando tenemos una colección de datos resultantes de un experimento o muestreo, conviene realizar una primera exploración de cómo son esos datos, antes de realizar ningún análisis complejo. La **estadística descriptiva** aporta parámetros que nos dan una idea inicial sobre cómo son esos datos. Concretamente, disponemos de "medidas de tendencia central" y de "medidas de dispersión" de los datos alrededor de ese valor central.

MEDIDAS DE TENDENCIA CENTRAL

Estas medidas indican alrededor de qué valor se agrupan los datos observados. Distinguimos:

1. Media aritmética (X): es el centro de gravedad de la serie de datos y se calcula como

$$X = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{(Ecuación 1)}$$

donde x_i representa cada uno de los valores de la variable y n el número de réplicas.

2. Mediana: es el punto medio de una serie ordenada de datos
3. Moda: es el valor más frecuente de la serie de datos.

MEDIDAS DE DISPERSIÓN

Estas medidas indican si los valores de la variable están muy dispersos o se concentran alrededor de la medida de centralización. Son:

1. Rango (R): Diferencia entre el valor máximo (x_{max}) y el mínimo (x_{min}) observado.

$$R = x_{max} - x_{min} \quad \text{(Ecuación 2)}$$

2. Varianza (s^2): Expresa la dispersión de valores entorno a la media (X)

$$s^2 = \frac{\sum(x_i - X)^2}{n-1} \quad \text{(Ecuación 3)}$$

3. Desviación estándar (s): Es la raíz cuadrada de la varianza.
4. Error estándar (*SE*): *s/raiz(n)*

Para tener una representación visual de estas medidas, es recomendable representar gráficamente la media junto con medidas de dispersión.

DISTRIBUCIÓN DE LOS DATOS

Tras calcular los parámetros de la estadística descriptiva, debemos explorar cómo se distribuyen los datos. Los histogramas de frecuencias (Fig. IV.1) son una herramienta de representación de datos que nos permiten observar cómo se distribuyen los mismos. Están formados por rectángulos adyacentes que tienen por base cada uno de los intervalos de la variable medida y por altura las frecuencias absolutas (nº de veces que aparecen datos dentro de ese intervalo). El número de intervalos a utilizar (k) se puede calcular según la regla de Sturges (1926): $k = 1 + 3.322 * \log(n)$, donde n es el tamaño de muestra.

De entre todas las distribuciones posibles que puedan seguir unos datos, la **distribución normal** es la más interesante desde el punto de vista estadístico, pues reúne unas propiedades que han hecho posible que a partir de ella se desarrollaran numerosos métodos de análisis de datos.

Propiedades de la distribución normal:

- Los valores cercanos a la media son los más abundantes, y a medida que nos alejamos de la media, los datos presentan una frecuencia cada vez menor.
- Es simétrica alrededor de la media. Por tanto, media, mediana y moda coinciden.
- Se caracteriza por dos medidas: media y desviación típica
- Tiene forma de campana, sin un pico excesivo.
- El 50% de las observaciones se encuentran por debajo de la media y el 50% por encima.
- El 68% de las observaciones se encuentran dentro del intervalo $x \pm s$
- El 95% de las observaciones se encuentran dentro del intervalo $x \pm 1{,}96 * s$
- El 99% de las observaciones se encuentra dentro del intervalo $x \pm 2{,}57 * s$.

Cómo dibujar un histograma de frecuencias

Para saber si una serie de datos sigue una distribución normal o no, podemos dibujar histogramas de frecuencia o un gráficos *qq*. Por último, podemos utilizar test estadísticos para saber si la distribución de nuestros datos se ajusta a algún modelo de distribución. A continuación mostraremos cómo hacer cada una de estas pruebas con R studio.

Análisis en R de la distribución de datos *

** Los análisis en R mostrarán en verde los comentarios al texto (al estar precedidos de "#" R los interpreta como comentarios, no como comandos), en azul el código de R a incluir en el programa y en negro los resultados al mismo.*

Histograma de frecuencias para explorar la normalidad de una serie de datos

```
#Generamos un conjunto de datos imaginarios de longitud de tarso (en mm) en una población
de aves. Nuestros datos deben seguir una distribución normal, una media 69.7 cm y una
desviación estándar de 14.65. A continuación representamos un histograma de distribución
de frecuencias de esos datos.
```

```
data<-rnorm(n = 1000, mean = 69.7, sd = 14.65) #genero una serie de datos aleatorios que cumplan los requisitos de media y desviación estándar indicados y que sigan una distribución normal.
hist(data, #histograma con los datos Figura III.1
     xlab = "Longitud de tarso (mm)", ylab = "Probabilidad", #añado los nombres de los ejes.
     main = "", prob = TRUE,ylim = c(0,0.04))
lines(density(data), col="blue", lwd=2) #mostramos la probabilidad en el eje Y y aumentamos el límite para que se vea bien la curva
abline(v=69.7, col="red")#Dibujamos una línea roja en la media
```

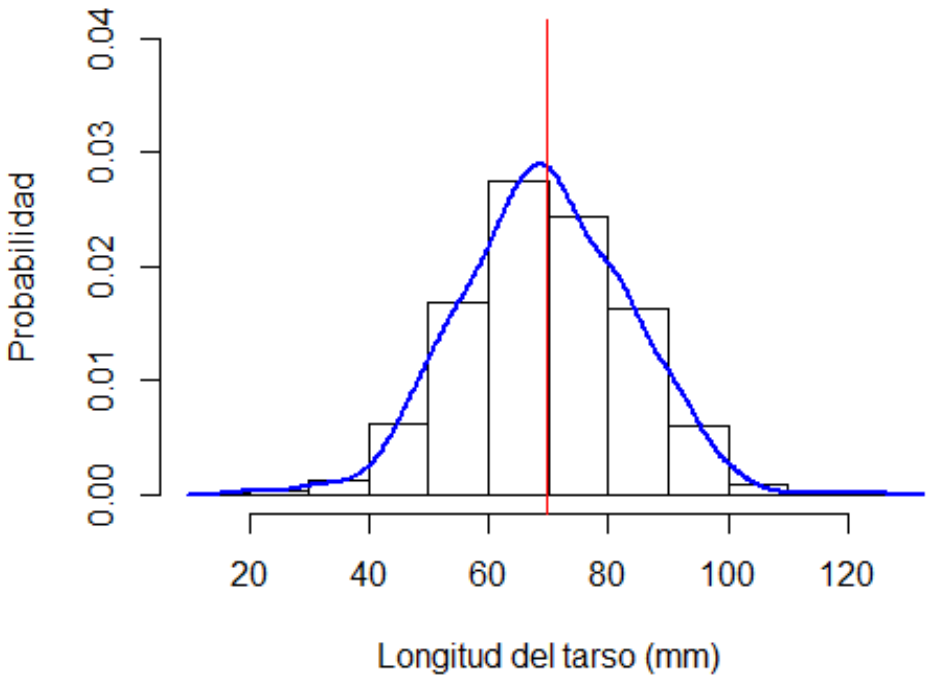

Figura III.1. *Histograma de probabilidad de la variable "Longitud del tarso (mm)" (eje X) de una población de aves. El eje Y muestra la probabilidad con la que aparecen valores en cada intervalo de X. En este caso se trata de una distribución normal.*

Cómo dibujar un gráfico qq para explorar la normalidad

Un gráfico qq-normal confronta los cuantiles teóricos en caso de que la distribución sea normal, con los cuantiles reales de los datos (Fig. IV.2). Si la distribución se ajusta a la normalidad los puntos se distribuyen a lo largo de una línea recta diagonal.

Podemos crear un gráfico qq con los datos creados en el ejemplo anterior con el siguiente código:

```
Gráfico qq para explorar la normalidad de una serie de datos

data<-rnorm(n = 1000, mean = 69.7, sd = 14.65) # genero una serie de datos aleatorios que cumplan los requisitos de media y desviación estándar indicados y que sigan una distribución normal.
```

```
qqnorm(data, col="springgreen4") # dibujo el gráfico qq,indicando que los puntos tengan color verde
qqline(data) # añado la línea diagonal, que indica la distribución teórica que seguirían los puntos en caso de ser normales.
```

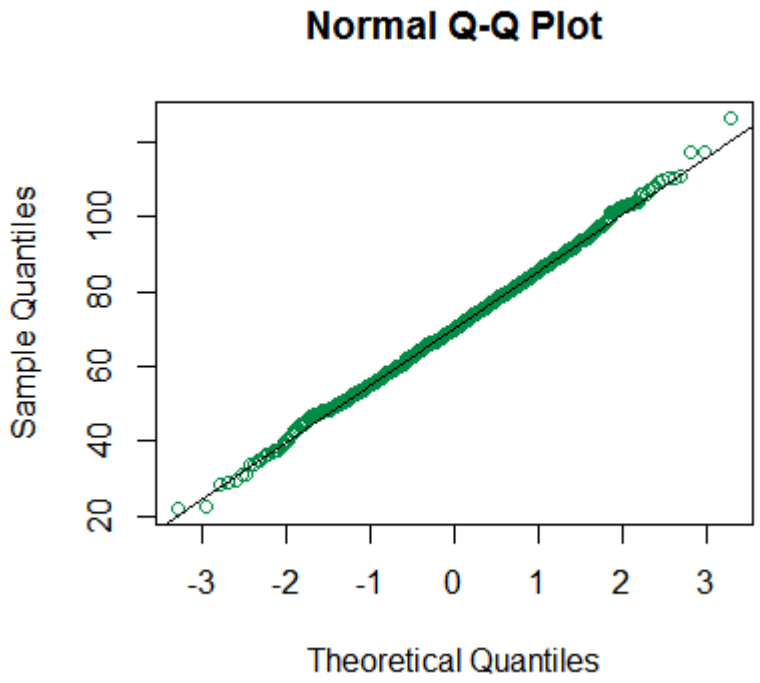

Figura III.2. *Gráfico qq para explorar la normalidad de la variable "Longitud del tarso (mm)". El eje X indica los cuantiles teóricos si la distribución es normal y el eje Y los cuantiles de la muestra. Como los datos (en verde) se ajustan a la línea diagonal teórica de una distribución normal, podemos considerar que la distribución de los datos es normal.*

Cómo hacer un test estadístico para comprobar la normalidad de los datos

Existen diversos test estadísticos que nos indican la probabilidad de que la distribución que sigue una serie de datos difiera de una distribución normal (hipótesis nula o H_o). El resultado del test nos devuelve un valor de probabilidad (*valor de p*). Si $p \leq 0.05$, rechazamos *Ho* y concluimos que nuestros datos no se ajustan a una distribución normal. Si por el contrario $p > 0.05$ aceptamos H_o) y concluimos que nuestros datos siguen una distribución normal (ver más detalles en el apartado 1.2).

En nuestro caso vamos a aplicar el test de Shapiro-Wilk para evaluar la normalidad a la serie de datos de los ejemplos anteriores.

```
Test de Shapiro-Wilk para explorar la normalidad de una serie de datos

data<-rnorm(n = 1000, mean = 69.7, sd = 14.65) # genero una serie de datos aleatorios que cumplan los requisitos de media y desviación estándar indicados y que sigan una distribución normal.

shapiro.test(data) # Test de normalidad
Shapiro-Wilk normality test
data:  data
W = 0.99865, p-value = 0.6524
```

Observamos que el valor de p que devuelve el test es muy superior a 0.05, por lo que aceptamos la hipótesis nula de que los datos siguen una distribución normal. Este resultado es concordante con la exploración gráfica mostrada en las figuras IV.1 y IV.2.

HOMOGENEIDAD DE VARIANZAS (HOMOCEDASTICIDAD)

Cuando queremos comparar dos o más series de datos, correspondientes a la misma variable (por ej. la humedad del suelo entre zonas con suelo desnudo y zonas recubiertas por tomillos), probablemente tendremos que saber si las varianzas difieren entre las series de datos. Esta información es necesaria para poder seleccionar el test estadístico adecuado (ver apartado 1.2). Para ello podemos aplicar manualmente el test de la F de Snedecor. Este test evalúa la hipótesis nula (H_0) de que las varianzas son iguales.

Cuadro IV.1. Prueba de comprobación de varianzas iguales: F de Snedecor

Se calculan las varianzas de cada una de las dos muestras: S^2_1 y S^2_2
Se calcula el estadístico F_{cal} a partir de la siguiente fórmula:

$$F_{cal} = \frac{S^2_{mayor}}{S^2_{menor}} \text{ (ecuación 4)}$$

Grados libertad: n_1-1, n_2-1 (n_1 tamaño de la muestra de varianza mayor)
H_o: varianzas iguales. Si $F_{cal} \geq F_{crítica}$ (La $F_{crítica}$ se busca en las tablas, ver sección dedicada al Anova), se rechaza la H_o, es decir, se concluye que las varianzas no son iguales.

Alternativamente, podemos realizar un test similar de homocedasticidad usando RStudio. La hipótesis nula que se evalúa, al igual que antes, es que las varianzas son iguales. El test nos devuelve un valor de *p*, que si es mayor de 0.05 nos llevará a aceptar la hipótesis nula. En caso contrario, rechazaremos H_0 y concluiremos que las varianzas no son iguales.

Test de Bartlett para explorar la homogeneidad de varianzas

```
# Genero una matriz (dataframe llamado "datos") con una variable que es "longitud del tarso" de una especie de ave y otra que es "habitat" con dos categorías. En cada tipo de hábitat (bosque y matorral) hay 50 medidas de longitud de tarso.

longitud<-rnorm(n = 100, mean = 69.7, sd = 14.65)
habitat<-c("bosque", "matorral")
habitat<-rep(habitat, each=50)
datos<-data.frame(habitat, data)

# Aplico el test de Bartlett para ver si la varianza de la longitud de tarso difiere entre ambos tipos de hábitat.
bartlett.test(longitud ~ habitat, data=datos) # Si los valores de la variable dependiente están en dos columnas tendría que usar esta forma: bartlett.test(list(var1, var2))
Bartlett test of homogeneity of variances
data:  longitud by habitat
Bartlett's K-squared = 0.0056683, df = 1, p-value = 0.94
```

El resultado ofrece un valor de p muy superior a 0.05, por lo que no rechazo la hipótesis nula y concluyo que la varianza de longitud de tarso es igual en ambos tipos de hábitat.

1.2. Pruebas de contraste de hipótesis

Se han desarrollado numerosos tests estadísticos que permiten realizar pruebas de contraste de hipótesis a partir de la distribución normal y la existencia de homogeneidad de varianzas: son las pruebas **paramétricas**. Sin embargo, con frecuencia los datos que obtenemos en un trabajo no siguen los requisitos de normalidad y homoscedasticidad; en esos casos recurrimos a la **estadística no paramétrica**.

Para contrastar una hipótesis ecológica (H_{ec}, por ej. el factor A afecta a la respuesta B), hemos de plantear una hipótesis nula (H_o), que supone la negación de la hipótesis ecológica (el factor A no afecta a la respuesta B). Cuando realizamos cualquier test estadístico de contraste de hipótesis, lo que hacemos es calcular la probabilidad de equivocarnos al rechazar H_0 (y por tanto aceptar nuestra hipótesis). Esa probabilidad es el ***valor de p*** (o ***p-valor***) que acompaña a un resultado estadístico. Por tanto, para poder rechazar H_o el valor de p debe ser bajo. Para tomar una decisión respecto a cuál sea la hipótesis 'verdadera', el investigador fija el nivel máximo de error que se permite asumir al aceptar H_{ec} (que se suele denotar como α). Por convenio el umbral de significación se suele fijar en 0, es decir, nos permitimos un error máximo del 5% en nuestra afirmación de la hipótesis ecológica. **Por tanto, si $p \leq 0.05$, rechazamos *Ho* y aceptamos *Hec*.**

En función del número de variables implicadas en un análisis estadístico, distinguimos dos tipos de métodos de análisis de datos:

- **Métodos bivariantes**: Permiten evaluar la relación entre dos variables (normalmente un factor y una variable respuesta). Los tipos de pruebas bivariantes que se desarrollan en este manual dependen de la naturaleza de las variables implicadas y se resumen en la Tabla 1.
- **Métodos multivariantes**: El análisis implica manejar al mismo tiempo tres o más variables. Este tipo de pruebas no se incluyen en este manual.

Tabla III.1. *Resumen de los métodos estadísticos bivariantes en función de la naturaleza de las variables y del tipo de distribución (paramétrica o no)*

Variable 1 (dependiente)	**Variable 2 (independiente)**		**Los datos siguen distribución normal y/o tienen homogeneidad de varianzas**	
			SI	**NO**
Cualitativa	Cualitativa		-	Test de la χ^2 (tablas de contingencia)
Cuantitativa	Cualitativa	2 categorías	t-Student	U de Mann-Whitney/ Wilcox
		> 2 categorías	Análisis de la varianza (ANOVA)	Kruskal-Wallis
Cuantitativa	Cuantitativa	Se asume que una var. es causa de la otra	Regresión	-
		No se asume relación causa-efecto	Correlación de Pearson	Correlación de Spearman

2. ASOCIACIÓN ENTRE VARIABLES CUALITATIVAS: TEST DE LA χ^2

El test de la χ^2 se utiliza para analizar la asociación entre dos **variables cualitativas** (por ejemplo, la presencia/ausencia de una especie y el tipo de suelo, color de la flor y presencia/ausencia de polinizadores, etc.). Este test parte de una tabla de contingencia, donde las columnas indican las categorías de la variable A y las filas las categorías de la variable B. En cada celda se anota la frecuencia de observaciones correspondiente. El test compara las frecuencias observadas con las frecuencias esperadas en caso de que no existiera asociación (es decir, si las observaciones se distribuyen al azar entre las categorías de las variables).

2.1. Requisitos e hipótesis de trabajo

La aplicación de este test requiere que las muestras estén tomadas al azar y que las frecuencias esperadas sean superiores a 5. Como se trata de un test que relaciona variables cualitativas, no hay ningún requisito acerca de la distribución de las variables.

Las hipótesis de trabajo serán del tipo:

- H_{ecol}: Existe asociación entre las variables (por ej. esperamos mayor frecuencia de polinizadores en las flores amarillas que en las azules)
- H_0: Las dos variables son independientes (por ej. a los polinizadores no les importa el color de las flores y aparecerán con la misma frecuencia en las amarillas y en las azules)

2.2. Contraste de hipótesis

Se compara el valor obtenido de χ^2_{cal} con el valor χ^2_{crit} correspondiente al número de grados de libertad apropiados y al valor de α previamente seleccionado (normalmente, α=0.05 ó 0.01):

Si $\chi^2_{cal} \geq \chi^2_{crit}$, se rechaza la H_0 (hay asociación entre las variables)

Si $\chi^2_{cal} < \chi^2_{crit}$, se acepta la H_0 (no hay asociación entre las variables)

2.3 Procedimiento de cálculo de la χ2

Supongamos, por ejemplo, que queremos saber si existe asociación entre la presencia de la especie A (un invertebrado acuático) y el tramo del río (alto, medio y bajo) para el caso del río Henares. Nuestras hipótesis son:

- H_0: La presencia de la especie A es independiente del tramo del río
- H_{ec}: Existe relación entre la presencia de la especie A y el tramo del río

Para comprobar cuál se cumple hemos hecho un muestreo a lo largo del río y en cada tramo hemos registrado la presencia (+) o ausencia (-) de la especie en 15 muestras de agua tomadas al azar. Los resultados se muestran en la Tabla III.2. A partir de estos datos construiríamos una tabla de contingencia con los datos observados en campo (Tabla III.3)

Tabla III.2: *Presencia (+) o ausencia (-) en cada una de las 15 réplicas de agua tomadas en cada tramo del río Henares.*

Tramo Alto	Tramo Medio	Tramo Bajo
+	-	-
+	-	+
+	-	-
-	+	-
+	-	-
+	-	-
+	-	-
+	-	-
+	+	-
+	-	-
-	-	-
+	-	-
+	-	-
+	-	-
+	-	-

Tabla III.3. *Tabla de contingencia que muestra los valores observados de frecuencia de la especie A en cada tramo del río, según la tabla 2.*

		Tramo del río		
		Alto	**Medio**	**Bajo**
Especie A	+	13	2	1
	-	2	13	14

A continuación se calcula el estadístico χ^2_{cal} siguiendo la siguiente fórmula:

$$\chi^2_{(\alpha,gl.)} = \sum \frac{(o-e)^2}{e}$$

- o = frecuencias observadas en el inventario
- e = frecuencia esperada de una celda, suponiendo que no hubiese asociación

$$e = \frac{c_t * f_t}{N}$$

c_t = total de la columna donde está la celda
f_t = total de la fila donde está la celda
N = nº total de casos

- *gl.* (grados de libertad) = (nº columnas-1)*(nº filas-1)

Para calcular el estadístico χ^2_{cal} conviene añadir a la tabla de contingencia las frecuencias esperadas en cada celda (entre paréntesis), como se indica en la Tabla III.4.

Tabla III.4. *Tabla de contingencia que muestra la frecuencia de la especie A observada en cada tramo del río y la frecuencia esperada en caso de independencia entre variables (entre paréntesis)*

		Tramo del río			Total
		Alto	**Medio**	**Bajo**	**Total**
Especie A	+	13 (*5.3*)	2 (*5.3*)	1 (*5.3*)	16
	-	2 (*9.7*)	13 (*9.7*)	14 (*9.7*)	29
Total		15	15	15	45

$$\chi^2_{cal} = \frac{(13-5{,}3)^2}{5.3} + \frac{(2-5{,}3)^2}{5.3} + \frac{(1-5{,}3)^2}{5.3} + \frac{(2-9{,}7)^2}{9.7} + \frac{(13-9{,}7)^2}{9.7} + \frac{(14-9{,}7)^2}{9.7} = 25.8$$

$\chi^2_{\text{crít (2 g.l., } \alpha=0.05)} = 5.99$ [g.l. = (nº filas -1) x (nº columnas -1)]

$\chi^2_{\text{cal}} > \chi^2_{\text{crít}}$ ($p<0.05$)

Análisis en R del test de la χ^2)

```
#Primero creamos la tabla de contingencias, que corresponde a nuestros datos:
rio<-matrix(c(13,2,1,2,13,14),byrow=TRUE,ncol=3)
colnames(rio)=c("Alto","Medio","Bajo")
rownames(rio)=c("Presente","Ausente")
rio #Para ver si la tabla está bien
         Alto Medio Bajo
Presente   13     2    1
Ausente     2    13   14

#Aplicar el test de la chi-cuadrada
chisq.test(rio)
Pearson's Chi-squared test
data:  rio
X-squared = 25.797, df = 2, p-value = 2.501e-06
```

En consecuencia, se rechaza H_0 (la presencia de la especia A es independiente del tramo de río) con una probabilidad de equivocarnos = 2.5×10^{-6}. Observando la tabla de contingencia, concluimos que la especie A aparece preferentemente en los tramos altos del río, ya que es en éstos donde su frecuencia observada es mayor que la esperada.

Caso especial: En las tablas de contingencia de 2x2, como la de la Tabla III.5, el estadístico χ^2_{cal} se puede calcular con las fórmulas que aparecen debajo.

Tabla III.5. *Tabla de contingencia de 2 x.2*

		Variable 1		
		A	**B**	**Total filas**
Variable 2	+	(a)	(b)	(a+b)
	+	(c)	(d)	(c+d)
Total columnas		(a+c)	(b+d)	(a+b+c+d)

Si $N \geq 30$

$$\chi^2_{cal} = \frac{(a*d - b*c)^2 * N}{(a+b)*(c+d)*(a+c)*(b+d)}$$

Si $N < 30$ (Corrección de Yates)

$$\chi^2_{cal} = \frac{N*(|a*d - b*c| - N/2)^2}{(a+b)*(c+d)*(a+c)*(b+d)}$$

3. TESTS DE COMPARACIÓN DE DOS MEDIAS

Sirven para comparar la media o mediana de una variable respuesta cuantitativa entre dos grupos definidos por dos categorías de una variable independiente cualitativa. Por ejemplo, si queremos comparar el peso corporal de los conejos entre una población que vive en un retamar y otra que vive en una pradera sin retamas. En ese caso, la variable independiente cualitativa es la población (de retamar o de pradera) y la variable dependiente cuantitativa es el peso corporal.

3.1. Selección del test

Para seleccionar el test apropiado debemos saber si los valores de la variable respuesta cuantitativa siguen una distribución normal dentro de cada grupo. Esto se puede comprobar visualmente dibujando un histograma o un gráfico qq, o aplicando el test de Shapiro-Wilk (ver sección 2.1). Asimismo, hay que comprobar si las varianzas de ambos grupos son similares, con la F de Snedecor (Cuadro IV.1) o con el test de Barlett (ver sección 1.2).

Si la variable cuantitativa sigue la distribución normal y las varianzas de ambos grupos son iguales, se utilizará el test paramétrico: **t de Student.** En cualquier otro caso se realizará el test no paramétrico: **U de Mann-Whitney.**

3.2. Hipótesis de una o dos colas

Cuando la hipótesis ecológica establece que existen diferencias entre las medias (o medianas) de los dos grupos, sin presuponer cuál de las dos medias es mayor que la otra, se dice que la hipótesis es de "dos colas", ya que incluye dos posibilidades (que la media del grupo A sea mayor que la del B o viceversa).

H_{ec}: $\mu_1 \neq \mu_2$

H_0: $\mu_1 = \mu_2$

Por el contrario, si la hipótesis ecológica establece que una de las dos medias es mayor que la otra, la hipótesis es de una cola, porque solo incluye una posibilidad. En este caso la hipótesis nula es la que abarca dos posibilidades (que la diferencia de las medias vaya en sentido contrario al esperado o que las medias sean iguales).

H_{ec}: $\mu_1 > \mu_2$

H_0: $\mu_1 \leq \mu_2$

Es importante establecer esta diferencia porque el resultado del p-valor difiere.

3.3. Cálculo de la t de Student para muestras independientes

Este es el cálculo que tenemos que aplicar cuando las muestras tomadas en las dos situaciones definidas por la variable categórica son independientes, es decir, no hay unas muestras que a priori sean más similares a otras.

Si los datos cumplen los requisitos establecidos, se puede calcular el estadístico t_{cal} a partir de la siguiente fórmula:

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{Sc\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{donde:} \qquad Sc = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

n_1 y n_2= tamaños de las muestras 1 y 2 respectivamente

$\bar{x}_1$ y $\bar{x}_2$ = medias de las muestras 1 y 2 respectivamente

s_1^2 y s_2^2 = varianzas de las muestras 1 y 2 respectivamente

A continuación se mide la significación del estadístico t_{cal}, comparando ese valor con el valor de un estadístico t_{crit} que se obtiene mirando las tablas correspondientes. Para identificar el t_{crit} que nos corresponde hemos de fijarnos en el número de colas que tiene nuestra hipótesis (una cola: *one-tailed*; dos colas: *two-tailed*), en el nivel de significación (α) con el que pretendemos rechazar la hipótesis nula (normalmente α = 0.05) y en los grados de libertad del test ($n_1 + n_2 - 2$).

- Si $|t_{cal}| \geq t_{crit}$ (α=0.05 o inferior) $\Rightarrow$ se rechaza H_0 y se acepta H_{ecol} (las medias son diferentes)
- Si $|t_{cal}| < t_{crit}$ (α=0.05) $\Rightarrow$ se acepta H_0 y se rechaza H_{ecol} (las medias son iguales)

Cuadro IV.2: Ejemplo de cálculo de la t de Student

Queremos saber si la humedad del suelo en un determinado lugar varía en función de la cubierta vegetal del mismo (tomillar o suelo desnudo), pues suponemos que la cubierta vegetal contribuye a aumentar la humedad del suelo por disminución de la evaporación. Para ello se ha realizado un muestreo en el que se ha medido la humedad de suelo (en % del volumen) en seis muestras distribuidas al azar bajo tomillares y en 8 muestras también distribuidas al azar en la misma zona, pero en condiciones de suelo desnudo.

Variables:

- Tipo de cobertura de suelo, con dos categorías: tomillar y suelo desnudo (cualitativa, independiente)
- Humedad del suelo (cuantitativa, dependiente)

Hipótesis

- H_{ecol}: la humedad de suelo es mayor bajo el tomillar: $\mu_{tomillar} > \mu_{suelo\ desnudo}$ (una cola).
- H_0: $\mu_{tomillar} \leq \mu_{suelo\ desnudo}$

Tabla de datos:

Cobertura	Humedad de suelo (%)	n	Media	s^2
tomillar	73.0 74.2 75.0 75.3 75.5 75.8	6	74.8	1.04
suelo desnudo	71.0 71.5 72.0 72.4 73.5 74.0 74.3 75.2	8	72.9	2.20

Cálculos:

$$t_{cal} = \frac{74.8 - 72.9}{1.42\sqrt{\frac{1}{6} + \frac{1}{8}}} = 2.36$$

$t_{cal} = 2.36 > t_{crít\ (\alpha=0.05,\ 12\ gl,\ una\ cola)} = 1.782$

Interpretación:
Se rechaza la H_0, y se acepta la H_{ecol}, es decir, se concluye que existen diferencias significativas en la humedad del suelo en función de la cobertura vegetal, siendo mayor en condiciones de cubierta vegetal de tomillar que en condiciones de suelo desnudo.

Análisis en R del test paramétrico de comparación de dos medias (t de student)

```
#Creamos un vector con datos para cada una de las dos categorías de la variable
independiente que queremos comparar. Los valores son los de humedad del suelo.
tomillar<-c(73.0,74.2,75.0,75.3,75.5,75.8)
suelo<-c(71.0,71.5,72.0,72.4,73.5,74.0,74.3,75.2)
#Representamos gráficamente los datos para ver dónde parece que hay mayor humedad
boxplot(tomillar, suelo, ylab="humedad del suelo", col="tan", names=c("tomillar",
"suelo"))
#La figura resultante es la Fig. IV.3
#REQUISITOS PARA APLICAR LA T-STUDENT
#Comprobamos la normalidad de la variable en "tomillar" y en "suelo".
#Lo podemos hacer visualmente con histogramas y gráfico qq.
#También podemos aplicar el test de Shapiro-Wilk (ver scripts en sección 1.1)
#Analizamos si las varianzas de la variable en tomillar y en suelo son iguales.
#Aplicamos el test de Barlett (ver sección 1.1).
#Como se cumplen todos los requisitos, aplicamos la t de Student
t.test(tomillar, suelo) #Aquí hacemos un test de dos colas.
t = 2.6896, df = 11.977, p-value = 0.01971
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: 0.3438945 3.281105
```

```
sample estimates:  mean of x          mean of y
                         74.8000   72.9875
#Si quiero plantear una hipótesis de una cola donde la hipótesis alternativa es que la
humedad del suelo desnudo es mayor que la del tomillar haré lo siguiente:
t.test(tomillar, suelo, alternative="greater")
t = 2.6896, df = 11.977, p-value = 0.009857
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.6112328       Inf
sample estimates:
mean of x mean of y
  74.8000   72.9875
#En este caso el p-valor es la mitad que en el test de dos colas, porque solo estoy
considerando uno de los dos posibles resultados.
# Prueba a hacer el test de una cola pero con la alternativa contraria:
t.test(tomillar, suelo, alternative="less")
```

Los datos de humedad de suelo siguen una distribución normal con forma de campana de Gauss, tanto en el tomillar como en suelo desnudo (Figura III.3a,b). Además, no muestran desviaciones notable de la normalidad en el gráfico qq. Igualmente, los test de Shapiro, que testan normalidad, sugieren que en ninguna de las dos situaciones (cobertura de tomillar y suelo desnudo) los datos presentan desviaciones de la normalidad (p>0.05). Cuando hemos realizado el test paramétrico t de Student, la variable dependiente "humedad del suelo" difiere significativamente entre las dos categorías de la variable independiente (tomillar y suelo desnudo) (P < 0.001).

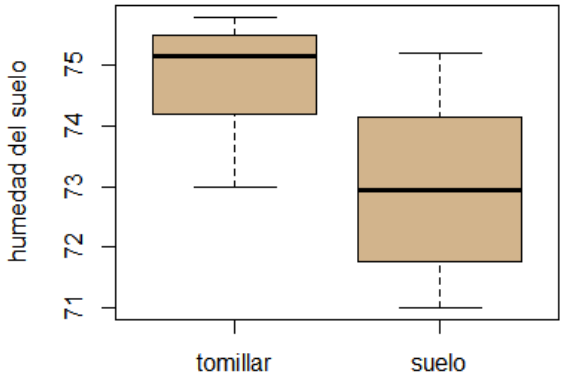

Figura III.3*. Valores medios y desviaciones de la variable dependiente "humedad del suelo" en cada categoría de la variable independiente "tipo de cobertura" (tomillar y suelo desnudo).*

3.4. Cálculo de la t de Student para muestras pareadas

Los test de comparación de medias pareadas son necesarios cuando las muestras tomadas no son igualmente independientes entre sí. Por ejemplo, se quiere comparar la variable "riqueza de especies" entre dos situaciones: lugares invadidos por una especie invasora (+I) y lugares no invadidos (-I). El muestreo se realiza a lo largo de un universo de muestreo muy heterogéneo (por ej. con distintos tipos de sustrato, o distinta altitud sobre el nivel del mar). Por tanto, cabe esperar que dos muestras lejanas en zonas invadidas sean más diferentes entre sí que dos muestras cercanas, una en +I y otra en –I. Para evitar que el efecto de la heterogeneidad ambiental diluya el efecto que la invasión tiene sobre la riqueza de especies, diseñamos un muestreo pareado, de forma que cada muestra en +I se compara con un control próximo en –I (Fig. IV.3). A continuación se estimará la diferencia de riqueza entre

cada muestra y su control y se realizará una t de Student para ver si la media de las diferencias difiere o no de cero.

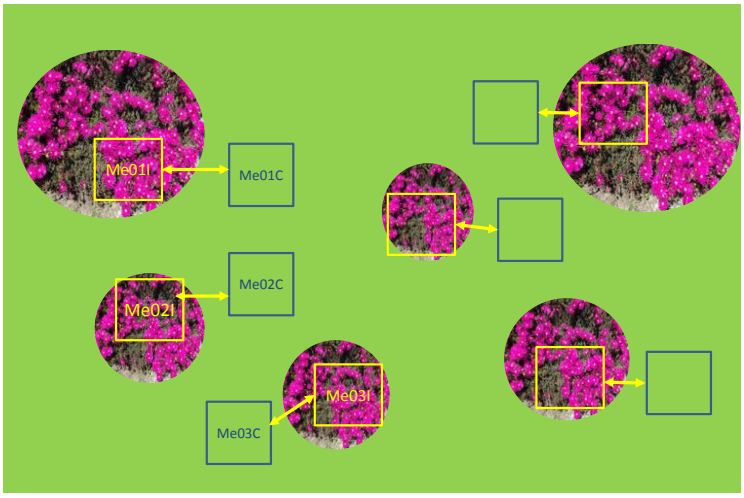

***Figura III.4**. Representación de un universo de muestreo en el que una comunidad ha sido invadida por una especie invasora (manchas de flores rosas). Para evitar la interferencia de factores ambientales no deseados, se ha diseñado un muestreo pareado, donde cada muestra en un lugar invadido (cuadros amarillos) será comparada con una muestra control en un lugar no invadido (cuadros azules).*

Cuadro IV.3: Ejemplo de cálculo de la t de Student con medias pareadas

Queremos saber si un tratamiento realizado durante un año en una población de pinares afectada por procesionaria ha tenido un impacto en la cantidad de cobertura arbórea. Por ello la cobertura arbórea se ha medido dos veces: antes y después del tratamiento en 10 sitios distintos. Esto nos proporciona 10 valores antes del tratamiento y 10 valores después del tratamiento, midiendo dos veces la cobertura arbórea del mismo sitio. Por ello, aquí debe usarse una t de Student de medias pareadas para comparar la media antes y después del tratamiento (las muestras correspondientes al mismo sitio están más relacionadas entre sí que muestras de sitios distintos).

Variables:

- Cobertura arbórea (cuantitativa, dependiente).
- Tratamiento (cualitativa, independiente con dos categorías: antes/después).

Hipótesis

- H_{ecol}: la cobertura arbórea es mayor después del tratamiento que antes del tratamiento: $cobertura_{después} > cobertura_{antes}$ (una cola).
- H_0: $cobertura_{después} \leq cobertura_{antes}$.

Tabla de datos:

Tratamiento	Cobertura arbórea (%)	n	Media	s^2
Antes	61.98777 63.72664 59.74309 62.70319 61.67165 61.40515 60.56058 61.16024 60.47696 61.69815	10	61.5	1.54
Después	96.94371 95.31954 99.16069 96.24361 96.30076 96.52271 97.83126 95.60667 94.57118 97.35336	10	95.9	1.20

Cálculos:

$$t = \frac{m}{s \times \sqrt{n}}$$

Donde, m es la media de las diferencias entre cada par de muestras, s es la desviación estándar de las diferencias y n es el tamaño de las diferencias.

$$t = \frac{-35.0720}{2.1099 \times \sqrt{10}} = -49.04632$$

t_{cal} = -52.56 > $t_{crít}$ (α=0.05, 9 gl, una cola)

Interpretación:
Se rechaza la H_0, y se acepta la H_{ecol}, es decir, se concluye que existen diferencias significativas en la cobertura arbórea antes y después de aplicar el tratamiento, siendo mayor después de aplicar el tratamiento.

Comparación de dos medias pareadas (t de Student) en R

```
#Creamos dos vectores con datos para cada una de las dos situaciones que queremos
#comparar
antes<-c(61.98777,63.72664,59.74309,62.70319,61.67165,
         61.40515,60.56058,61.16024,60.47696,61.69815)
despues<-c(96.94371,95.31954,99.16069,96.24361,96.30076,
           96.52271,97.83126,95.60667,94.57118,97.35336)

#Representamos gráficamente los valores medios y las desviaciones
boxplot(antes, despues, ylab="Cobertura de árboles", col= "tan",
        names=c("antes", "después")) #La figura resultante es la Fig. IV.5

#Comprobamos que se cumplen los requisitos de normalidad y homoscedasticidad (ver sección
1.1).
#Test de comparación de medias pareado
t.test(antes, despues, paired = TRUE)
data:  antes and despues
t = -52.564, df = 9, p-value = 1.64e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: -36.58138 -33.56263
sample estimates: mean of the differences -35.07201
```

La Figura III.5 sugiere que hay una gran diferencia de cobertura arbórea antes y después del tratamiento. Cuando hemos realizado el test paramétrico t de Student de medias pareadas se puede concluir que las medias muestran unas diferencias altamente significativas ($p << 0.001$).

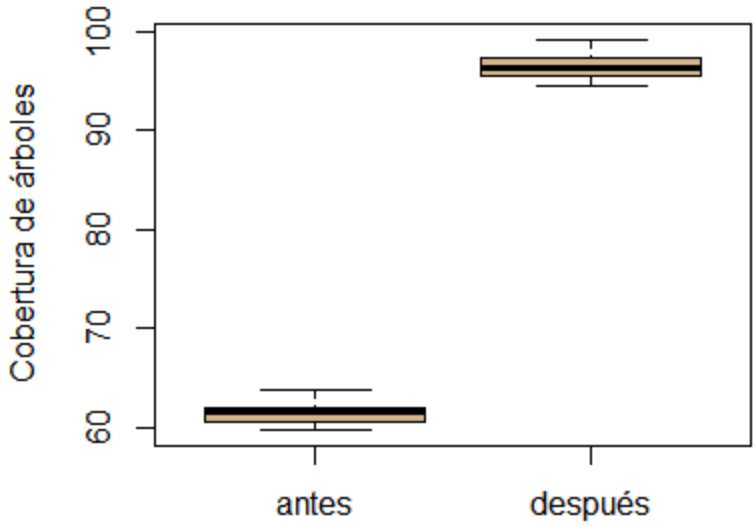

***Figura III.5**. Valores medios y desviación de cobertura arbórea antes y después del tratamiento aplicado para controlar la procesionaria del pino.*

3.5. Comparación no paramétrica para muestras independientes

Si nuestros datos no cumplen los requisitos necesarios para aplicar una t-Student, podemos aplicar una U de Mann-Whitney, que compara las diferencias entre dos medianas (en lugar de las medias). Este test se basa en rangos (número de orden de los datos en función de su magnitud) en

lugar de en los parámetros de la muestra (media, varianza), por ello se dice que es un test no paramétrico.

Los pasos a seguir para calcular la U de Mann-Whitney son:

1. Asignación de rangos a cada dato: se ordenan todos los valores (juntando los dos grupos) de menor a mayor. El rango de cada dato será el número de orden que le corresponde a cada dato. Cuando se repita el mismo valor numérico, el rango que se asigna a esos datos es la media aritmética de los rangos que les corresponderían en función del número de orden que ocupan.
2. Se suman los rangos de cada uno de los grupos que se comparan y se calcula la suma de los rangos de los datos de cada uno de los grupos (R_1 y R_2)
3. Se calculan los estadísticos U_1 y U_2 a partir de las siguientes fórmulas:

$$U_1 = n_1 \cdot n_2 + \frac{n_2(n_2+1)}{2} - R_2 \qquad U_2 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

   Se obtiene el estadístico U_{cal} escogiendo el valor más grande entre U_1 y U_2 .
4. Se comprueba la significación estadística del estadístico U_{cal} comparando este valor con el valor de un estadístico $U_{crít}$ obtenido a partir de las tablas correspondientes.
5. Si $U_{cal} \geq U_{crít}$ (α=0.05 o inferior) $\Rightarrow$ se rechaza H_0 y se acepta H_{ecol} (las medianas son diferentes)
6. Si $U_{cal} < U_{crít}$ (α=0.05) $\Rightarrow$ se acepta H_0 y se rechaza H_{ecol} (las medianas son iguales)

Cuadro IV.4: Ejemplo de cálculo de la U de Mann-Whitney

Se quiere estudiar si el número de especies de ácaros edáficos se ve influido por un incendio de baja intensidad. Para ello se simuló un incendio de baja intensidad en una parcela de un territorio homogéneo y se tomaron 6 muestras al azar de la zona incendiada y 7 muestras también al azar de la zona no incendiada, contándose el número de especies de ácaros edáficos en cada muestra.

Variables:
- Variable dependiente: número de especies de ácaros edáficos (cuantitativa)
- Variable independiente: ocurrencia de un incendio (cualitativa)

Hipótesis:
- H_{ecol}= La mediana del número de especies de ácaros edáficos varía dependiendo de que se haya producido un incendio: $M_{quemada} \neq M_{no\ quemada}$. (dos colas).
- H_0= La mediana del número de especies de ácaros edáficos es igual en la parcela quemada que en la no quemada: $M_{quemada} = M_{no\ quemada}$

Tabla de datos:

Parcela	Número de especies de ácaros edáficos	n
quemada	6 9 12 12 15 16	6
no quemada	10 13 16 16 17 19 20	7

Asignación de rangos a cada dato:

dato *	**6**	**9**	10	**12**	**12**	13	**15**	**16**	16	16	17	19	20
rango	1	2	3	4’5	4’5	6	7	9	9	9	11	12	13

* en negrita los valores correspondientes al inventario de la parcela quemada

Cálculo de U_{cal}
- - Se suman los rangos de cada grupo: R_1=28 R_2=63

- U_1=6x7+[(7x8)/2]-63=7
- U_2=6x7+[(6x7)/2]-28=35 ➔ U_{cal}
- $U_{cal} = 35 < U_{crít\ (\alpha=0.05)} = 36$

Interpretación:
No se rechaza la H_0, concluimos que el número de especies de ácaros edáficos no se ve influido significativamente por la ocurrencia de un incendio de baja intensidad.

Si queremos realizar el test en RStudio, podemos aplicar el test de Wilcox

Realización en R del test no paramétrico de comparación de dos muestras independientes

```
#Creamos dos vectores con datos para cada una de las dos situaciones que queremos
#comparar. (Ojo, los vectores no se corresponden con las variables, que son: tiempo
#(independiente) y riqueza (dependiente)
quemada<-c(0,0,12,12,15,15)
noquemada<-c(10,13,16,16,17,19,20)

#Antes de nada, conviene representar gráficamente los datos para ver qué tendencias
# muestran
boxplot(quemada, noquemada, ylab="riqueza de ácaros", names=c("quemada", "no quemada"))

#Test de normalidad
#Podemos comprobar visualmente la normalidad con histogramas y gráficos qq (sección 1.2
#y Figura III.6)
#También podemos aplicar el test de normalidad de Shapiro a los datos de cada categoría
shapiro.test(quemada)
Shapiro-Wilk normality test
data:  quemada
W = 0.76178, p-value = 0.02591
shapiro.test(noquemada)
Shapiro-Wilk normality test
data:  noquemada
W = 0.94671, p-value = 0.6997
#El resultado sugiere que la variable no es normal en la categoría "quemada"

#Test de comparación de medias
wilcox.test(quemada, noquemada)
Wilcoxon rank sum test with continuity correction
data:  quemada and noquemada
W = 6, p-value = 0.03726
alternative hypothesis: true location shift is not equal to 0
```

El número de especies de ácaros parece ser mayor en la madera no quemada (Figura III.5). Dado que la variable sigue una distribución no normal en la categoría "quemada" aplicamos el test de Wilcox. El resultado indica que las diferencias que muestra la figura sí son significativas (P > 0.05).

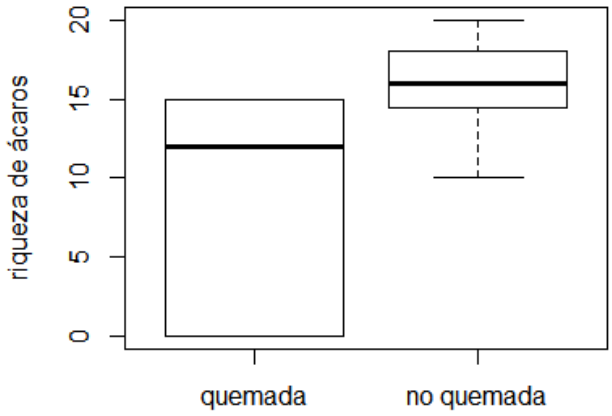

Figura III.6*. Valores medios y desviación de la riqueza de ácaros encontrada en la zona quemada y no quemada.*

3.6. Comparación no paramétrica para muestras pareadas

Vamos a suponer ahora que en el ejemplo anterior queremos comparar la riqueza de ácaros en los mismos puntos de muestreo antes y después de un incendio. En este caso las muestras están pareadas, es decir, quiero comparar cada valor de riqueza en el mismo sitio antes y después del incendio.

Realización en R del test no paramétrico de comparación entre dos muestras pareadas

```
#Creamos dos vectores con datos para cada una de las dos situaciones que queremos
#comparar
antes<-c(10,10,16,19,20,20,24)
despues<-c(0,1,9,12,15,15,15)
#Represento la media y la dispersión de los datos para ver qué tendencia siguen
boxplot(antes, después, ylab=”riqueza de ácaros”, names=c(“antes”, “después”)
#La figura resultante es la IV.7
#Test de normalidad
#Podemos comprobar visualmente la normalidad con histogramas y gráficos qq (sección 1.2
#y Figura III.5)
#También podemos aplicar el test de normalidad de Shapiro
shapiro.test(antes)
Shapiro-Wilk normality test
data:  antes
W = 0.89258, p-value = 0.2884
shapiro.test(despues)
Shapiro-Wilk normality test
data:  despues
W = 0.80237, p-value = 0.04324
#El resultado sugiere que la variable no es normal en la categoría “despues”

#Test de comparación de medias
wilcox.test(antes, despues, paired= TRUE)
Wilcoxon signed rank test with continuity correction
data:  antes and despues
V = 28, p-value = 0.02178
alternative hypothesis: true location shift is not equal to 0
```

Los resultados indican que la riqueza de ácaros en los mismos puntos disminuyó tras el incendio (Figura III.6), y que la diferencia es significativa, ya que el resultado del test estadístico nos da una p=0.02178, que se menor de 0.05.

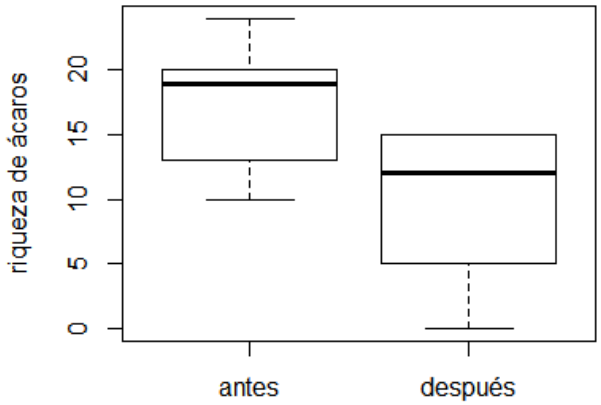

Figura III. 7. *Mediana y desviación de la riqueza de ácaros en varios puntos de muestras antes y después de un incendio.*

4. TESTS DE COMPARACIÓN DE MÁS DE DOS MEDIAS

Sirven para comparar las medidas de tendencia central (media o mediana) entre más de dos grupos de datos para determinar si existen o no diferencias entre ellos. Por tanto relacionan una variable cualitativa de más de dos estados (variable independiente) con otra cuantitativa (variable dependiente). Por ejemplo, habría que aplicar este test para determinar si existen diferencias significativas en la densidad de escarabajos (variable dependiente, cuantitativa) que encontramos en cuatro tipos de suelo (variable independiente, cualitativa con cuatro estados).

4.1. Selección del test

Al igual que en la comparación de dos medias, para elegir el test hay que comprobar si los datos siguen una distribución normal dentro de cada grupo, y si las varianzas entre grupos son homogéneas (ver sección 1.1). Si ambos requisitos se cumplen se utilizará el test paramétrico: **ANOVA.** En cualquier otro caso se realizará el test no paramétrico: **Kruskal-Wallis**

4.2. Hipótesis

La hipótesis ecológica establece que existen diferencias entre las medias (o medianas, en el caso del test no paramétrico) de los grupos considerados, es decir, que **al menos** dos de las medias serán distintas. La hipótesis nula establece que no existen diferencias entre dichas medias.

H_{ecol}: No todas las medias/medianas son iguales

H_0: $\mu_1 = \mu_2 = ... = \mu_k$

Si rechazamos la hipótesis nula, significa que al menos dos de los grupos difieren entre sí. Sin embargo, este test no indica qué grupos difieren o son iguales de qué grupos. Si quisiéramos conocer las diferencias entre todos los pares de grupos posibles, tendríamos que aplicar algún test "post-hoc", que no vemos en este manual. En cualquier caso, es incorrecto aplicar varios test de Student con este fin.

4.3. Procedimiento de cálculo del ANOVA

La valoración de las diferencias entre las medias de los distintos grupos se basa en la descomposición de la variabilidad total del conjunto de datos en dos términos: variabilidad debida a las diferencias entre los grupos (variabilidad entre grupos), y variabilidad debida al azar (variabilidad dentro de grupos).

$\text{Variabilidad}_{\text{total}} = \text{Variabilidad}_{\text{entre grupos}} + \text{Variabilidad}_{\text{dentro grupos}}$

La variabilidad entre datos se puede estimar con la varianza (s^2), y con Suma de Cuadrados (SS), que es el cociente entre la varianza y los grados de libertad (g-l.). Por tanto:

$\text{SS}_{\text{total}} = \text{SS}_{\text{entre grupos}} + \text{SS}_{\text{dentro grupos}}$

Las sumas de cuadrados se obtienen a partir de las siguientes fórmulas:

$$SS_{total} = \sum x^2 - \frac{\left(\sum x\right)^2}{N}$$

$$SS_{entre\ grupos} = \left[\frac{\left(\sum x_1\right)^2}{n_1} + \frac{\left(\sum x_2\right)^2}{n_2} + \ldots + \frac{\left(\sum x_k\right)^2}{n_k} \right] - \frac{\left(\sum x\right)^2}{N}$$

Donde

k = número de grupos

N= número total de datos

n_1, n_2, ... n_k = número de datos en cada grupo.

x = cada uno de los datos de cada grupo

El cálculo de la suma de cuadrados se obtiene despejando de la ecuación:

$\text{SS}_{\text{dentro grupos}} = \text{SS}_{\text{total}} - \text{SS}_{\text{entre grupos}}$

- Cálculo de los grados de libertad de las sumas de cuadrados:

$$g.l.\ SS_{total} = N - 1 \qquad g.l.\ SS_{entregupos} = k - 1 \qquad g.l.\ SS_{dentrogrupos} = N - k$$

- Conversión de las sumas de cuadrados (SS) en varianzas:

$$s^2_{entregrupos} = \frac{SS_{entre\,grupos}}{g.l._{entregrupos}} = \frac{SS_{entre\,grupos}}{k-1} \qquad s^2_{dentrogrupos} = \frac{SS_{dentro\,grupos}}{g.l._{dentrogrupos}} = \frac{SS_{dentro\,grupos}}{N-k}$$

- Cálculo del estadístico F:

$$F = \frac{s^2_{entregrupos}}{s^2_{dentrogrupos}}$$

Si en la población de la que proceden las muestras no hay diferencias reales entre los grupos definidos por la variable cualitativa, la varianza entre grupos será similar a la varianza dentro de grupos (por tanto, el cociente entre ambas estará cerca de 1). En el caso de que existan diferencias reales entre los grupos (lo que presupone la hipótesis ecológica) la varianza entre grupos será mayor

que la varianza dentro de los grupos (el cociente entre ambas será mayor de 1). El estadístico que nos dice si las desviaciones respecto a ese valor de 1 son significativas es *F*.

El contraste de hipótesis se realiza comparando el valor de la F_{cal} con el valor $F_{crít}$ obtenido a partir de la tabla para el valor de α previamente establecido (normalmente α=0.05 o inferior). La búsqueda de dicha $F_{crít}$ requiere del número de grados de libertad del numerador y del denominador. La forma habitual de notación que se usa en las tablas lleva el valor de α entre paréntesis, y los grados de libertad del numerador y del denominador a continuación, en orden consecutivo y separados por comas. Por ejemplo, $F_{crít\ (0.05)\ 3,\ 22}$. significa el valor del estadístico F de las tablas para un α=0.05, con 3 grados de libertad en el numerador y 22 en el denominador.

- Si $F_{cal} \geq F_{crít} \Rightarrow$ se rechaza H_0 y se acepta H_{ecol} (alguna de las medias es diferente)
- Si $F_{cal} < F_{crít} \Rightarrow$ se acepta H_0 y se rechaza H_{ecol} (las medias son iguales)

Cuadro IV.5: Ejemplo de cálculo de ANOVA

Se quiere saber si el tipo de cobertura de suelo (suelo desnudo, piedras, hojarasca y pastizal) influye sobre la densidad de hormigueros. Para ello se ha realizado un muestreo en el que se ha medido el número de hormigueros en diez muestras distribuidas al azar dentro de cada una de las zonas con diferente cobertura.

Variables:
- cobertura de suelo (cualitativa, independiente)
- densidad de hormigueros (cuantitativa, dependiente)

Hipótesis:
- H_{ecol}: Alguna de las medias es diferente (la cobertura de suelo influye sobre la densidad de hormigueros)
- H_0: $\mu_{suelo\ desnudo} = \mu_{piedras} = \mu_{hojarasca} = \mu_{pastizal}$

Tabla de datos:

Cobertura	Densidad de hormigueros	n	Media	Σx	$(\Sigma x)^2$	Σx^2
suelo desnudo	78 88 87 88 83 82 81 80 80 89	10	83.6	836	698896	70036
piedras	78 78 83 81 78 81 81 82 76 76	10	79.4	794	630436	63100
hojarasca	79 73 79 75 77 78 80 78 83 84	10	78.6	786	617796	61878
pastizal	77 69 75 70 74 83 80 75 76 75	10	75.4	754	568516	57006
Total		40		3170		252020

Cálculo de la suma de cuadrados total:
$SS_T = 252020 - (3170)^2/40 = 797.5$

Cálculo de la variabilidad entre grupos ($SS_{entre\ grupos}$):
$SS_{entre} = 698896/10 + 630436/10 + 617796/10 + 568516/10 - 3170^2/40 = 341.9$

Cálculo de la variabilidad dentro de los grupos ($SS_{dentro\ grupos}$):
$SS_T = SS_{entre} + SS_{dentro} \Rightarrow SS_{dentro} = SS_{total} - SS_{entre} = 797.5 - 341.9 = 455.6$

Determinar los grados de libertad de cada una de las sumas de cuadrados estimadas:
$SS_T = N - 1 = 40 - 1 = 39$ $SS_{entre\ grupos} = k - 1 = 4 - 1 = 3$ $SS_{dentro\ grupos} = N - k = 40 - 4 = 36$

Estimación de las varianzas dividiendo las SS por los grados de libertad:
$s^2_{entre\ grupos} = 341.9/3 = 113.97$ $s^2_{dentro\ grupos} = 455.6/36\ 12.66$

Cálculo del estadístico F_{cal} y comparación con el estadístico $F_{crít}$:
$F_{cal} = s^2_{entre\ grupos} / s^2_{dentro\ grupos} = 113.97/12.66 = 9.002$
$F_{crít\ (0.05)\ 3,\ 36} < 2.92$

Interpretación: $F_{cal} > F_{crít} \Rightarrow$ Rechazamos H_o La abundancia de hormigueros no es la misma en todas las zonas

Realización en R del test ANOVA

```
#Creamos la matriz con los datos con dos columnas, en una tenemos la variable dependiente
“densidad” y en otra las categorías de la variable independiente “cobertura”
cobertura<-c("desnudo","piedras","hojarasca","pastizal")
cobertura<-rep(cobertura,each=10)
hormigueros<-data.frame(cobertura,
"densidad"=c(78,88,87,88,83,82,81,80,80,89,78,78,83,81,78,81,81,82,76,76,79,73,79,75,77
,78,80,78,83,84,77,69,75,70,74,83,80,75,76,75))

#Analizamos la estructura de los datos
str(hormigueros)
'data.frame':       40 obs. of  2 variables:
 $ cobertura: Factor w/ 4 levels "desnudo","hojarasca",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ densidad : num  78 88 87 88 83 82 81 80 80 89 ...

#Representamos gráficamente los valores medios de densidad de hormigueros en cada tipo
de cobertura de suelo (Fig. IV.8)
boxplot(densidad ~ cobertura,   data=hormigueros, col="tan", cex.axis=0.7, las = 2,
ylab="Densidad de hormigueros", cex.lab=0.75)
#Añadimos al boxplots los datos individuales para ver cómo se distribuyen
stripchart(densidad ~ cobertura,  data=hormigueros, col="red",
           vertical = TRUE, method = "jitter", cex=0.5,
           add=TRUE, pch=19)

#ANOVA
modelo <- lm(densidad ~ cobertura,  data=hormigueros)
anovaModelo <- anova(modelo)
anovaModelo
Analysis of Variance Table
Response: densidad
          Df Sum Sq Mean Sq F value   Pr(>F)
cobertura  3  341.9 113.967  9.0053 0.000139 ***
Residuals 36  455.6  12.656
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

#Comprobamos las asunciones del ANOVA:
# 1. Normalidad de la variable dependiente en cada categoría de la variable
    independiente. Puedo hacerlo visualmente con histogramas o gráficos qq.
    (ver sección 1.1).
    También puedo aplicar el test de normalidad de Shapiro-Wilk a los residuos del
    modelo
shapiro.test(modelo$residuals)
Shapiro-Wilk normality test
data:  modelo$residuals
W = 0.97657, p-value = 0.5641
# Aceptamos H0: los residuos siguen una distribución normal.

#2. Homocedasticidad (varianzas iguales entre las categorías de la variable indep.)
bartlett.test(densidad ~ cobertura,  data=hormigueros)
Bartlett test of homogeneity of variances
data:  densidad by cobertura
Bartlett's K-squared = 2.5279, df = 3, p-value = 0.4703
# Aceptamos H0: las varianzas son iguales entre categorías de cobertura.
```

La fig. IV.8 sugiere que la densidad de hormigueros difiere entre sitios con distinta cobertura de suelo, y que es mayor en suelo desnudo y mínima en pastizal. Los datos cumplen con los presupuestos de la ANOVA (normalidad de los residuos – test de Shapiro-Wilk; y homocedasticidad – Test de Bartlett), y como tal se puede aplicar un modelo de ANOVA. El modelo de ANOVA confirma que las diferencias que sugiere la figura son significativas (p<0.001).

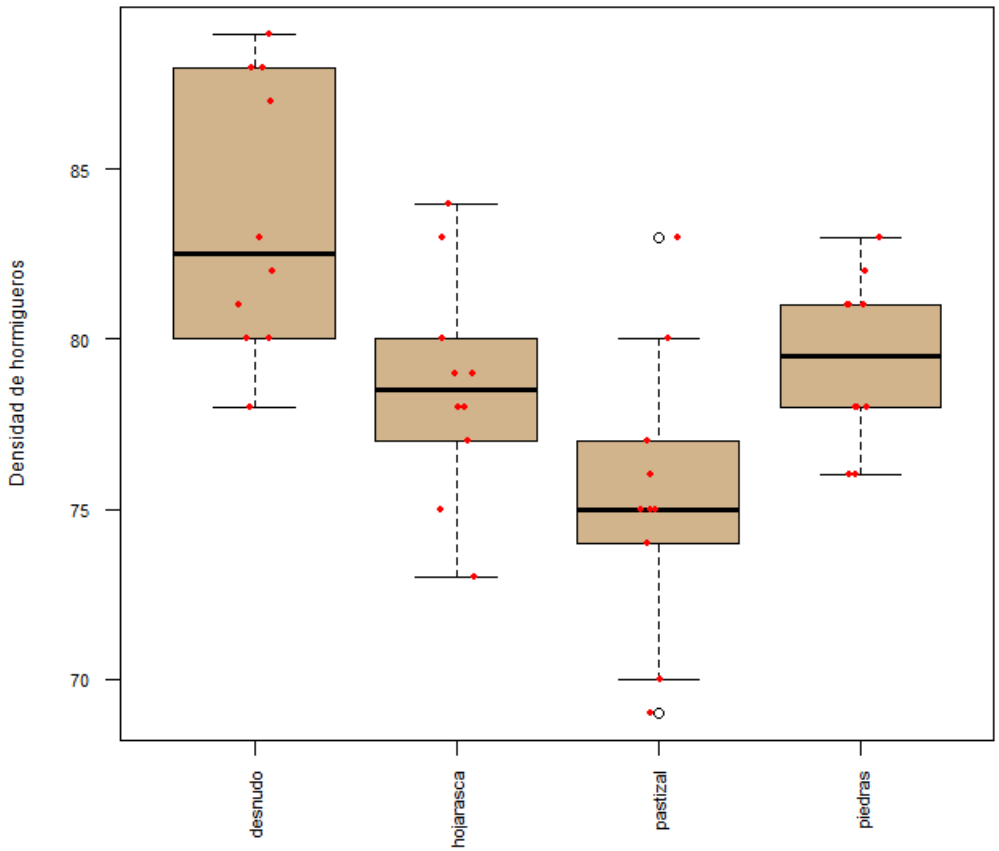

Figura III.8. *Medias y desviaciones de la densidad de hormigueros en los distintos tipos de cobertura de suelo. Los puntos rojos representan los valores de los datos individuales.*

4.4. Procedimiento de cálculo del test no paramétrico: Kruskal-Wallis

Se emplea cuando los datos no siguen la distribución normal y/o tienen varianzas distintas, en sustitución del ANOVA paramétrico. Al igual que la U de Mann-Whitney se basa en rangos en lugar de los parámetros de la muestra (media, varianza) y compara medianas en lugar de medias.

Los pasos a seguir son los siguientes:

- Asignación de rangos: se realiza exactamente igual que para la U de Mann-Whitney.

- Cálculo del estadístico *H*:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

k = número de grupos
N = número total de datos
n_i = número de datos en el grupo i

Cuando existen rangos ligados (dos o más números con el mismo rango) se aplica un factor de corrección, siendo H_c el estadístico que se utiliza en lugar de H, calculado según la siguiente expresión:

$$H_c = \frac{H}{C} \qquad C = 1 - \frac{\sum_{i=1}^{m} (t_i^3 - t_i)}{N^3 - N}$$

t_i = número de rangos ligados en cada grupo
m = número de grupos de rangos ligados

El valor crítico del estadístico calculado (H o H_c) se consulta en la tabla de la χ^2 si N≥15, o si $k > 5$, para (k-1) grados de libertad. Si N<15 y k<5 se consulta en la tabla específica para H.

- Si $H_{cal} \geq H_{crít}$ ($\chi^2_{crít}$) $\Rightarrow$ se rechaza H_0 y se acepta H_{ecol} (medianas diferentes)

Si $H_{cal} < H_{crít}$ ($\chi^2_{crít}$) $\Rightarrow$ se acepta H_0 y se rechaza H_{ecol} (medianas son iguales)

Cuadro IV.6: ejemplo de cálculo de Kruskal-Wallis

Se quiere estudiar si el pH de cuatro charcas situadas sobre sustratos diferentes es distinto. Para ello se obtuvieron 8 muestras de agua procedentes de cada una de las charcas, midiéndose el pH en cada una de ellas. Los datos de pH se ordenaron de forma ascendente para cada charca. (Una muestra de agua de la charca nº 3 se perdió, de forma que n3=7; pero el test no requiere igualdad en el número de datos de cada grupo). Los rangos (número de orden de menor a mayor) se muestran entre paréntesis.

Variables:

- Variable dependiente: pH (cuantitativa)
- Variable independiente: tipo de sustrato sobre el que se encuentra cada charca (cualitativa)

Hipótesis:

- Hecol= el pH no es el mismo en las cuatro charcas
- H0= el pH es el mismo en las cuatro charcas

Tabla de datos:

Charca 1	Charca 2	Charca 3	Charca 4
7.68 (1)	7.71 (6*)	7.74 (13.5*)	7.71 (6*)
7.69 (2)	7.73 (10*)	7.75 (16)	7.71 (6*)
7.70 (3.5*)	7.74 (13.5*)	7.77 (18)	7.74 (13.5*)
7.70 (3.5*)	7.74 (13.5*)	7.78 (20*)	7.79 (22)
7.72 (8)	7.78 (20*)	7.80 (23.5*)	7.81 (26*)
7.73 (10*)	7.78 (20*)	7.81 (26*)	7.85 (29)
7.73 (10*)	7.80 (23.5*)	7.84 (28)	7.87 (30)
7.76 (17)	7.81 (26*)		7.91 (31)
n1=8	n2=8	n3=7	n4=8
R1=55	R2=132.5	R3=145	R4=163.5

* Rangos ligados

$$N = 8+8+7+8 = 31$$

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1) = \frac{12}{31(32)}\left[\frac{55^2}{8}+\frac{132.5^2}{8}+\frac{145^2}{7}+\frac{163.5^2}{8}\right] - 3(32) = 11.876$$

Número de grupos de rangos ligados $= m = 7$

$$\sum_{i=1}^{m}(t_i^3 - t_i) = (2^3-2)+(3^3-3)+(3^3-3)+(4^3-4)+(3^3-3)+(2^3-2)+(3^3-3) = 168$$

$$C = 1 - \frac{\sum_{i=1}^{m}(t_i^3 - t_i)}{N^3 - N} = 1 - \frac{168}{31^3 - 31} = 1 - \frac{168}{29760} = 0.9944$$

$$H_c = \frac{H}{C} = \frac{11.876}{09944} = 11.943 \qquad \nu = k - 1 = 3 \qquad \chi^2_{0.05,\,3} = 7.815$$

$H_c cal > \chi^2_{crít}$ $\Rightarrow$ Se rechaza H_0

El pH no es el mismo en todas las charcas

Realización en R del test Kruskall-Wallis

```
## introducimos los datos
charca1 = c(7.68,7.69,7.70,7.70,7.72,7.73,7.73,7.76)
charca2 = c(7.71,7.73,7.74,7.74,7.78,7.78,7.80,7.81)
charca3 = c(7.74,7.75,7.77,7.78,7.80,7.81,7.84)
charca4 = c(7.71,7.71,7.74,7.79,7.81,7.85,7.87,7.91)
charcaID = c(rep("charca1",8),rep("charca2",8),
             rep("charca3",7),rep("charca4",8))
charcanum = c(rep(1,8),rep(2,8),
              rep(3,7),rep(4,8))
charcas = data.frame(charca = charcaID,
                     pH = c(charca1,charca2,charca3,charca4),
                     charca.num = charcanum)
charcas
   charca   pH charca.num
1  charca1 7.68          1
2  charca1 7.69          1
3  charca1 7.70          1
4  charca1 7.70          1
5  charca1 7.72          1
6  charca1 7.73          1
7  charca1 7.73          1
8  charca1 7.76          1
9  charca2 7.71          2
10 charca2 7.73          2
11 charca2 7.74          2
12 charca2 7.74          2
13 charca2 7.78          2
14 charca2 7.78          2
15 charca2 7.80          2
16 charca2 7.81          2
17 charca3 7.74          3...
#Examinamos la distribución de los datos, realizando un histograma para cada charca
par(mfrow=c(2,2)) # Esta instrucción es para que dibuje las cuatro figuras en
                  dos filas y dos columnas.
hist(charca1)
hist(charca1)
hist(charca1)
hist(charca1)
```

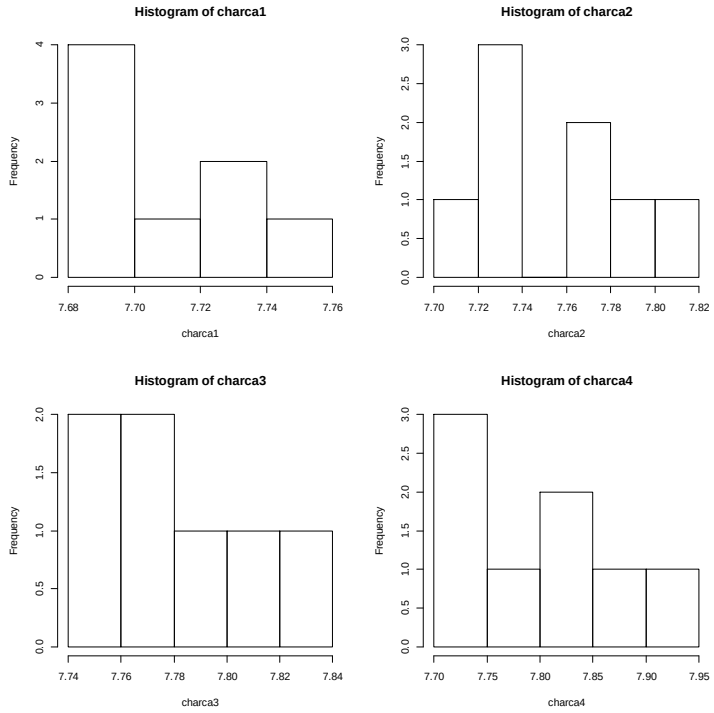

Figura III.9. *Histogramas mostrando la distribución de frecuencias del pH de cada una de las cuatro charcas del ejemplo.*

```
# examinamos la distribución del pH en cada charca con boxplot
dev.off()# Cierra la ventana del gráfico anterior y anula los parámetros gráficos
         previos
boxplot(pH ~ charca, data=charcas, col="tan", cex.axis=0.7, las = 2,
            ylab="pH de las charcas", cex.lab=0.75)
stripchart(pH ~ charca, data=charcas, col="red",
           vertical = TRUE, method = "jitter", cex=0.5,
           add=TRUE, pch=19)
```

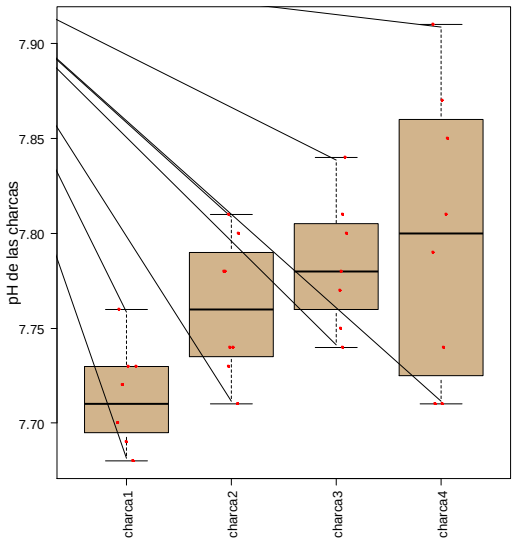

Figura III.10. *Diagrama de cajas y bigotes comparando las distribuciones de las medidas de pH en cada una de las cuatro charcas del ejemplo. Los puntos rojos muestran los valores correspondientes a cada observación.*

```
#Test de homocedasticidad (varianzas iguales)
# H0: las varianzas no difieren entre charcas. Test de Bartlett
bartlett.test(pH ~ charca, data=charcas)
Bartlett test of homogeneity of variances
data:  pH by charca
Bartlett's K-squared = 8.8272, df = 3, p-value = 0.03168
# las varianzas no son iguales, es decir, no se cumple la H0 de homocedasticidad ya que
p-value < 0.05, por tanto no podemos usar ANOVA, y procedemos a comparar las medias con
el test de Kruskal-Wallis

# Aplicamos el test Kruskal-Wallis
kruskal.charcas = kruskal.test(pH ~ charca.num, data = charcas)
kruskal.charcas
Kruskal-Wallis rank sum test
data:  pH by charca.num
Kruskal-Wallis chi-squared = 11.944, df = 3, p-value = 0.007579
# El valor de p-value < 0.05 nos informa de que no se cumple la H0 de que las medias de
pH de cada charca sean iguales.
```

La Figura III.10 sugiere que hay diferencias de pH entre las cuatro charcas comparadas, pero también indica que la dispersión de valores (medida con la varianza) es mayor en la charca 4. El test de Barlett confirma que las varianzas son heterogéneas. El resultado del Kruskal-Wallis confirma que las diferencias que muestra la figura son significativas.

5. ASOCIACIÓN ENTRE VARIABLES CUANTITATIVAS: COEFICIENTES DE CORRELACIÓN

El coeficiente de correlación cuantifica el grado de asociación entre dos variables cuantitativas. Se utiliza cuando no se asume que una variable es causa y la otra consecuencia. Por ejemplo, si queremos saber si el peso y la longitud del pico covarían dentro de una población de aves (no se asume una relación de causalidad).

ρ es el coeficiente de correlación real que existe entre dos variables en el conjunto de la población.

r y r_s son los coeficientes medidos sobre la muestra.

Los coeficientes de correlación varían entre –1 y 1 del siguiente modo (Fig. IV.3):

a) $1 \geq \rho > 0$: correlación positiva.

b) $-1 \leq \rho < 0$: correlación negativa.

c) $\rho \approx 0$: no hay correlación, los valores de x e y varían de forma independiente.

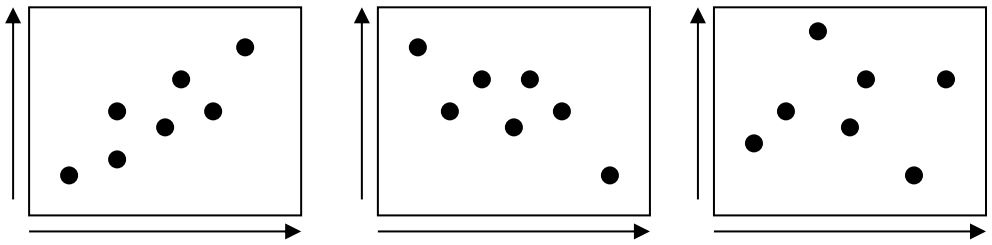

Figura III.11. *Tres posibles tipos de asociación entre variables: positiva (izqda.) negativa (centro) y ausencia de asociación (derecha)*

Cuanto más cerca esté el coeficiente de 1 ó –1, más fuerte es la correlación

5.1. Hipótesis de una cola y de dos colas

Cuando la hipótesis ecológica indica que existe correlación (sin precisar el signo) se trata de una hipótesis de dos colas, ya que implica dos posibilidades (que la relación sea positiva o negativa). La hipótesis nula solo implica una posibilidad: que no exista correlación entre las variables.

H_{ec}: $\rho \neq 0$ ($\rho < 0$ ó $\rho > 0$)

H_0: $\rho = 0$

Por el contrario, si la hipótesis ecológica precisa el signo de la correlación, entonces se trata de una hipótesis de una cola. La hipótesis nula implica dos posibilidades: que no haya correlación o que ésta sea del signo contrario al esperado en la hipótesis ecológica.

H_{ecol}: $\rho > 0 \Rightarrow H_0$: $\rho \leq 0$

H_{ecol}: $\rho < 0 \Rightarrow H_0$: $\rho \geq 0$

Es importante saber de cuántas colas es la hipótesis a la hora de evaluar la significación del test.

5.2. Correlación paramétrica: r de Pearson

Para poder aplicar este test las dos variables (dependiente e independiente) deben seguir una distribución normal.

El cálculo del índice de correlación de Pearson se hace a partir de la siguiente fórmula:

$$r = \frac{n\sum_{i=1}^{i=n} x_i y_i - \sum_{i=1}^{i=n} x_i \times \sum_{i=1}^{i=n} y_i}{\sqrt{\left(n\sum_{n=1}^{n=i} x_i^2 - \left(\sum_{n=1}^{n=i} x_i\right)^2\right) \times \left(n\sum_{n=1}^{n=i} y_i^2 - \left(\sum_{n=1}^{n=i} y_i\right)^2\right)}}$$

n- nº de pares de muestras

x_i- valores de la variable x

y_i- valores de la variable y

A continuación, se comprueba la significación del índice de correlación calculado comparándolo con el valor de un estadístico r_{crit} obtenido a partir de la tabla correspondiente, para una $\alpha = 0.05$ o inferior y las colas que establezca la hipótesis.

Si $|r_{cal}| \geq r_{crit}$ (α=0.05 o inferior) → Se rechaza la hipótesis nula. → Existe correlación.

Cuadro IV.7: Ejemplo de cálculo de r de Pearson
Un ornitólogo está interesado en conocer la longitud del pico de una población de aves que estudia. Sin embargo, esa medida resulta más costosa de tomar que el peso corporal. Por ello quiere saber si ambas variables se correlacionan para estimar la primera a partir de la segunda.

Variables (Ambas son cuantitativas y normales):
- x: longitud del pico.
- y: peso corporal.

Hipótesis:

- H_{ecol}: $\rho \neq 0$ ($\rho < 0$ ó $\rho > 0$) (dos colas)
- H_0: $\rho = 0$

Tabla de datos:

Obs.	Longitud del pico (mm)	Peso corporal (g)	x^2	y^2	xy
1	33.5	51	1122	2601	1708
2	38.0	59	14444	3481	2242
3	32.0	49	1024	2401	1568
4	37.5	54	1406	2916	2025
5	31.5	50	992	2500	1575
6	33.0	55	1089	3025	1815
7	31.0	48	961	2304	1488
8	36.5	53	1332	2809	1935
9	34.0	52	1156	2704	1768
10	35.0	57	1225	2349	1995
SUMA	342	528	11752	27990	18119

Cálculos
n = 10; r = 0.779, $r_{cal} = 0.779 > r_{crit\ (0.01)\ n=10} = 0.765$. Se rechaza H_0 y se acepta H_{ecol}

Interpretación:

Se puede concluir que existe una correlación positiva entre el peso corporal y la longitud del pico de esa población de aves. Esto significa que los cambios en peso corporal de esas aves son un fiel reflejo de los cambios en la longitud del pico.

Realización en R de la correlación de Pearson

```
## Generamos una tabla de datos con los datos del ejemplo
datos.pico <- data.frame(
  Obs. = seq(1,10,1),
  longitud.pico = c(33.5,38.0,32.0,37.5,31.5,33.0,31.0,36.5,34.0,35.0),
  peso.corporal = c(51,59,49,54,50,55,48,53,52,57)
)
```

```
     Obs. longitud.pico peso.corporal
1     1          33.5            51
2     2          38.0            59
3     3          32.0            49
4     4          37.5            54
5     5          31.5            50
6     6          33.0            55
7     7          31.0            48
8     8          36.5            53
9     9          34.0            52
10   10          35.0            57
```

```
#Represento las variables gráficamente para ver si hay relación aparente entre ellas
plot(longitud.pico ~ peso.corporal, data=datos.pico, pch=16)
abline(lm(longitud.pico ~ peso.corporal,
          data=datos.pico), col="red")

#Compruebo que ambas variables cumplen el requisito de normalidad (sección 1.1).
#Test de correlación para comprobar la hipótesis nula (H0) de rho = 0
cor.test( ~ longitud.pico + peso.corporal, data=datos.pico,
          method = "pearson",  continuity = FALSE,
          conf.level = 0.95)
```

```
Pearson's product-moment correlation
data:  longitud.pico and peso.corporal
t = 3.5194, df = 8, p-value = 0.007853 #El p-value indica asociación significativa en-
tre ambas variables.
alternative hypothesis: true correlation is not equal to 0 #La hipótesis es de 2 colas
95 percent confidence interval:
 0.2942560 0.9452104
sample estimates:
      cor
0.7794691 #Este es el valor de la r de Pearson, que es positivo y próximo a 1
```

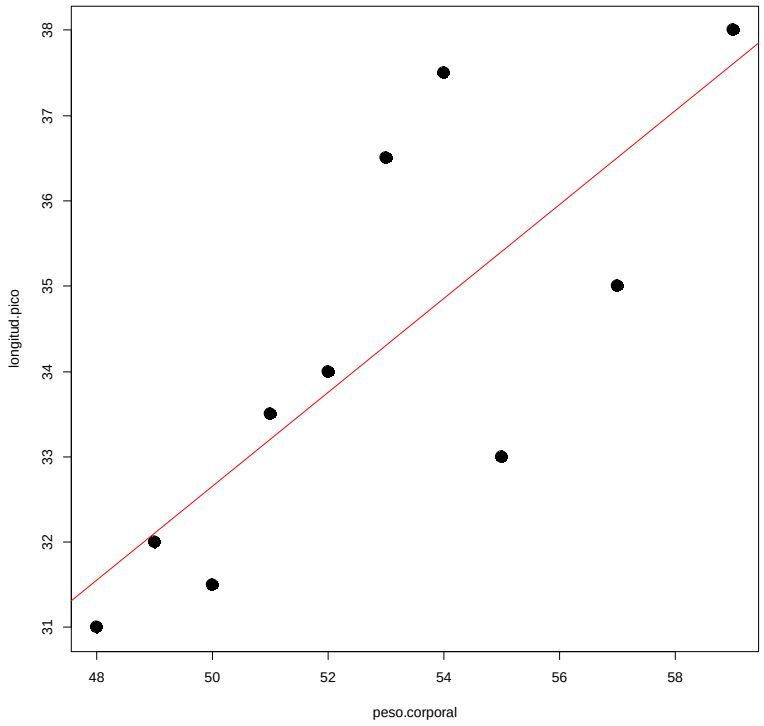

***Figura III.12**. Gráfico de dispersión entre la longitud del pico y el peso corporal y ajuste del modelo lineal (ver línea roja).*

La Figura III.12 sugiere que existe una relación entre ambas variables y el resultado del test confirma que esa relación es significativa.

5.3. Correlación no paramétrica: r *de Spearman*

Se aplica este test cuando una o ninguna de las dos variables implicadas sigue una distribución normal. Para calcular la *r* de Spearman hay que realizar los siguientes pasos:

- Ordenar los pares de datos en función del valor de *x* y asignar rangos a *x*.
- Repetir la ordenación en función de *y* y asignar rangos a *y*.
- Calcular el coeficiente:

$$r_s = 1 - \frac{6\sum_{i=1}^{i=n} d_i^2}{n^3 - n}$$

n = nº de pares de datos

d_i = diferencia de rangos en las variables del par *i*

Para comprobar la significación estadística del índice de correlación se consulta en la tabla correspondiente el valor crítico de r_s para *n* pares de datos, para *p*=0.05 o inferior y para el número de colas acorde con la hipótesis. Si $r_{s\ cal} \geq r_{s\ crít}$, se rechaza H_0.

Cuadro IV.8: Ejemplo de cálculo de r de Spearman

Se sospecha que la abundancia de la especie de gramínea *Poa bulbosa* en los pastizales mediterráneos depende en gran medida de la humedad que hay en el suelo. Para comprobar la hipótesis se realiza un muestreo con una cuadrícula de 20 cm de lado, que se dispone 12 veces al azar sobre la comunidad de pasto. En cada cuadrícula se mide la cobertura de la especie y la humedad del suelo mediante un TDR.

Variables: Ambas son cuantitativas y no siguen una distribución normal
- Cobertura de la especie (variable dependiente)
- Humedad del suelo (variable independiente)

Hipótesis
- H_{ec}: existe una correlación positiva entre la cobertura de *Poa* y la humedad $\rho > 0$ (de una cola)
- H_0: $\rho \leq 0$

Tabla de datos:

Obs.	Cobertura	Humedad	Rango cob.	Rango hum.	d	d^2
1	82	42	2	3	-1	1
2	98	46	6	4	2	4
3	87	39	5	2	3	9
4	40	37	1	1	0	0
5	116	65	10	8	2	4
6	113	88	9	11	-2	4
7	111	86	8	10	-2	4
8	83	56	3	6	-3	9
9	85	62	4	7	-3	9
10	126	92	12	12	0	0
11	106	54	7	5	2	4
12	117	81	11	9	2	4
					Suma	52

Cálculos

$$r_s = 1 - \frac{6 \times 52}{12^3 - 12} = 0.82 > r_s \text{ crit } (0.05) = 0.503$$

Interpretación:
Se rechaza H_0, hay correlación positiva entre la cobertura de Poa bulbosa y la humedad del suelo. Es importante destacar que este muestreo no es una demostración de una relación causa-efecto entre las variables, es decir, que con este muestreo no podemos concluir que la mayor humedad de suelo es la causa de la mayor abundancia de Poa bulbosa. Para determinar relaciones de causa-efecto se necesita realizar experimentos controlados y otros tests estadísticos que verifiquen ese tipo de relación.

Cálculo en R de la correlación de Spearman

```
#Genero una matriz con los datos que quiero analizar)
datos <- data.frame("cobertura" = c(82,98,87,40,116,113,111,83,85,126,106,117),
                    "humedad" = c(42,46,39,37,65,88,86,56,62,92,54,81))
str(datos)

#Dibujo el gráfico de dispersión para ver la relación entre las variables (Fig. IV.13)
```

```
plot(cobertura ~ humedad, data=datos, pch=16) #¡Ojo! La variable dependiente va delante
del símbolo "~" y la independiente detás.

#Correlación entre los datos

#Compruebo si ambas variables cumplen el requisito de normalidad (sección 1.1).
#Como no es así, aplico el test de correlación de Spearman.

cor.test( ~ cobertura + humedad,
          data=datos,
          method = "spearman",
          continuity = FALSE,
          conf.level = 0.95)
Spearman's rank correlation rho
data:  cobertura and humedad
S = 52, p-value = 0.002027
alternative hypothesis: true rho is not equal to 0
sample estimates:   rho = 0.8181818
```

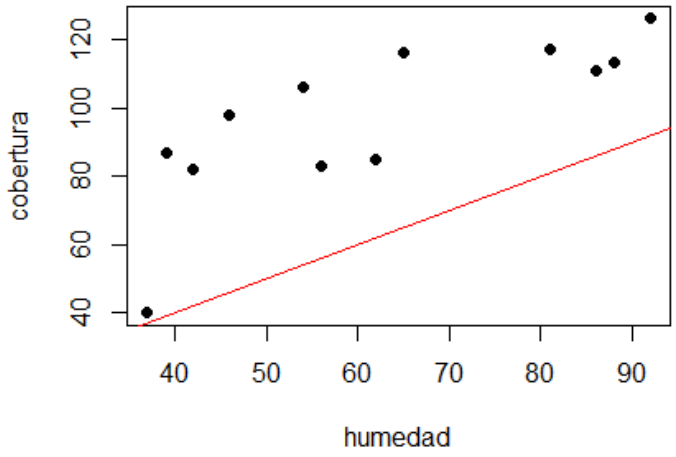

***Figura III.13**. Gráfico de dispersión entre la cobertura y la humedad mostrando una línea 1:1 en rojo.*

La Fig. IV.13 sugiere que hay relación entra las variables, ya que a medida que aumenta la humedad también lo hace la cobertura. El resultado del test da un valor de $p<0.05$, que confirma que efectivamente ambas variables están correlacionadas significativamente.

6. REGRESIÓN

La regresión se aplica cuando tenemos dos variables y asumimos (en nuestra hipótesis) que una depende de la otra. Por ejemplo, esperamos que la dosis de fertilizante que se aplica a una serie de plantas cultivadas en macetas cause diferencias en la altura de las plantas. La variable independiente es la dosis de fertilizante y la dependiente la altura.

Para poder aplicar este test, los residuos del modelo deben seguir una distribución normal.

El parámetro que se calcula en regresión es el coeficiente de regresión o R^2, que equivale al cuadrado del coeficiente de correlación de Pearson y se interpreta como el porcentaje de la varianza de la variable dependiente que explica la variable independiente. La significación de este coeficiente se calcula igual que el de la *r* de Pearson.

Cuadro IV.9: Ejemplo de cálculo de regresión

Las llanuras aluviales del tramo bajo del río Henares han sido parcialmente invadidas por *Ailanthus altissima*, un árbol exótico invasor procedente de China. Se sospecha que la invasión de esta especie tiene efectos negativos sobre la riqueza de especies del sotobosque. Para comprobar si es así, se realizó un muestreo en 17 parcelas de 20 x 20 m, a lo largo de un gradiente de abundancia relativa de *A. altissima*, contándose en cada parcela el número de especies leñosas que aparecían en el sotobosque.

Variables (Ambas son cuantitativas y normales):
- x: Abundancia relativa de *A. altissima* (valores entre 0 y 1).
- y: Nº de especies leñosas.

Hipótesis
- H_{ecol}: A mayor abundancia relativa de *A. altissima*, menor riqueza de especies leñosas.

Tabla de datos:

Nº de inventario	Abundancia relativa de *A. altissima*	Riqueza de especies leñosas
1	0	13
2	0.5	6
3	0.4	10
4	0.25	9
5	0.15	14
6	0.5	4
7	0.2	14
8	0.3	8
9	0.1	10
10	0.8	4
11	0.55	5
12	0.7	3
13	0	16
14	0.75	3
15	0.35	9
16	0.9	2
17	1	1

Hacemos una recta de regresión del tipo:

Riqueza = a + (b × Abundancia)

Donde a es la constante de la regresión y b la pendiente de la recta de regresión. Obtenemos:

Riqueza = 13.82 – 13.96 × Abundancia

Interpretación:

El parámetro b tiene un valor negativo, por lo que podemos aceptar nuestra hipótesis de un efecto negativo de la abundancia de *A. altissima* sobre la riqueza de especies leñosas.

Cálculo en R de una regresión

```
# Introducimos los datos y creamos las dos variables a correlacionar.

abu.aa<-c(0, 0.5,0.4,0.25,0.15,0.5,0.2,0.3,0.1,0.8,0.55,0.7,0,0.75,0.35,0.9,1)
riq<-c(13,6,10,9,14,4,14,8,10,4,5,3,16,3,9,2,1)

# Creamos el modelo de regresión lineal
modeloL = lm(riq ~ abu.aa) # Ojo, la variable dependiente (riqueza) se pone delante del
símbolo "~", que significa "en función de".

# Coeficientes de la recta
modeloL$coefficients
(Intercept)      abu.aa
  13.82469   -13.96237  # El primer valor es el intercepto (a) y el segundo la
pendiente (b)
#Para ver la significación puedes hacer un test de correlación con "cor.test".
#Representamos gráficamente la nube de puntos: Figura III.14
plot(riq ~ abu.aa, xlab="abundancia relativa A. altissima", ylab="riqueza de especies
leñosas")
#Añadimos la recta de regresión
abline(modeloL, col = "red")

# COMPRUEBA SI SE CUMPLEN LOS REQUISITOS DE LA REGRESIÓN
# H0: los residuos son normales.
shapiro.test(modeloL$residuals)
Shapiro-Wilk normality test
data:  modeloL$residuals
W = 0.96041, p-value = 0.6391 # Como los residuos son normales damos por buena la
regresión.
```

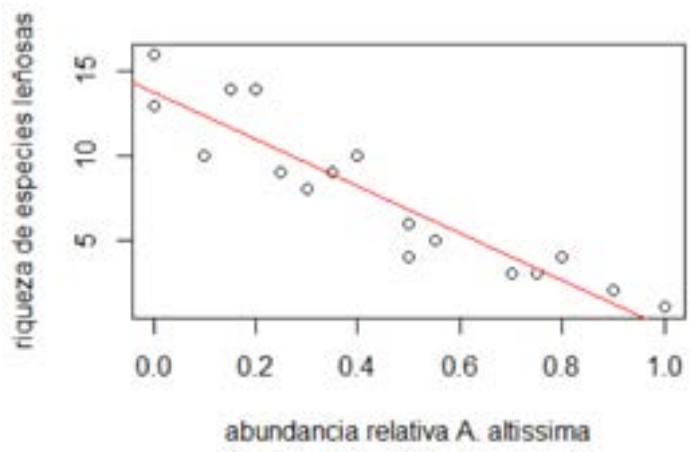

Figura III.14. *Gráfico de dispersión entre la abundancia de* A. altissima *y la riqueza de especies leñosas y ajuste del modelo lineal (línea roja)*

Se puede observar que la abundancia de *A. altissima* tiene un efecto negativo en la riqueza de especies leñosas. El intercepto de la regresión tiene un valor de 13.82 y la pendiente de -13.96 (ver Figura III.14). Al comprobar la normalidad de los residuos podemos afirmar que nuestros residuos se distribuyen normalmente ya que el p-valor es mayor de 0.05 (p = 0.639). En definitiva, podemos aceptar nuestra hipótesis de partida.

IV. ELABORACIÓN DE UN TRABAJO CIENTÍFICO EN ECOLOGÍA

Un trabajo científico se escribe para comunicar unas ideas que resultan de un proceso de investigación desarrollado mediante el método científico. A continuación se muestran los apartados que debe tener un trabajo científico estándar, así como algunas indicaciones sobre cómo elaborar cada uno de ellos. Os recomendamos leer algún artículo contrastando lo que leéis con esta estructura. Podéis descargar artículos en castellano de algunas revistas on-line, como *Ecosistemas* (http://www.revistaecosistemas.net/index.php/ecosistemas), *Pirineos* (http://pirineos.revistas.csic.es/index.php/pirineos) o *Revista Chilena de Historia Natural* (https://revchilhistnat.biomedcentral.com/)

El estilo de un artículo científico es muy distinto del que sigue un artículo periodístico o un texto literario. Se recomienda seguir las siguientes indicaciones.

- El lenguaje ha de ser sencillo, claro y conciso. Para ello se recomienda utilizar frases cortas y directas.
- Se escribe en forma impersonal, (por ej. no se dice "realicé un muestreo", sino "se realizó un muestreo".
- Hay que evitar imprecisiones del lenguaje. Las frases han de contener un mensaje claro. Antes de escribir cada frase, hay que tener clara la idea que se quiere reflejar (si la idea no está clara, la frase tampoco lo estará).
- Los resultados se escriben en pasado.
- El vocabulario ha de ser adecuado al contexto, no se pueden utilizar expresiones coloquiales y hay que evitar palabras de significado impreciso. No hacerlo da una impresión de descuido que el lector no puede aceptar. Por supuesto, no se admiten faltas de ortografía.
- Los nombres científicos se escriben en cursiva. El nombre del género va en mayúsculas y el de la especie en minúsculas (por ej. *Quercus robur*). Además los nombres científicos no llevan tildes porque están en latín. No hay que olvidar que se trata de nombres propios, luego no se puede poner un artículo delante (por ej. no es correcto decir “el *Quercus robur*”)
- El número de cifras decimales ha de ser homogéneo y razonable (por ej. 2-3 decimales para los valores de *p*).
- El texto debe ser coherente, tanto en estilo como en formato, y ha de seguir un hilo conductor. No es aceptable copiar párrafos de distintas fuentes y ponerlos juntos. Tampoco es aceptable que cada miembro del equipo escriba una parte por su cuenta para luego juntar esas partes sin más revisión. Para conseguir una redacción coherente y fluida, es fundamental que **todos los autores revisen el texto** con visión crítica. Se trata de un **trabajo cooperativo** en el que todos los miembros son igualmente responsables de todo lo que se dice en el trabajo.
- Hay que hacer un buen uso de la puntuación. Los puntos y aparte deben separar apartados con distinto contenido (no párrafos de igual tamaño).

1. TÍTULO

Ha de ser breve, pero informativo. Ha de proporcionar al lector una idea del contenido real del trabajo, evitando dar excesivos detalles, pero también evitando ser muy genérico. Ejemplos de títulos:

- *Impacto de la introducción de la abeja doméstica (Apis mellifera, Apidae) en el Parque Nacional del Teide (Tenerife, Islas Canarias).*
- *¿Se puede cartografiar la desertificación? Luces y sombras de una tarea desafiante.*
- *La ecología reproductiva de las plantas: estrategias reproductivas, fuerzas ecológicas y evolutivas.*
- *Las invasiones biológicas y su impacto en los ecosistemas.*

2. RESUMEN

Ha de contener una síntesis de cada una de las partes del trabajo (problema, objetivos, metodología, resultados y discusión). Ha de ser muy conciso (no más de 200 palabras). En el resumen no hay bibliografía. Ejemplo de un resumen, tomado de la revista *Ecosistemas*:

Cuadro IV.1: Ejemplo de un resumen, tomado de *Castro, A., Espinosa, C.I. 2016. Dinámica estacional de invertebrados en un matorral seco tropical a lo largo de un gradiente altitudinal. Ecosistemas 25(2): 35-45. Doi.: 10.7818/ECOS.2016.25-2.05*

Determinar la dinámica estacional de los seres vivos y su relación con variables climáticas a lo largo de gradientes ambientales resulta necesario para entender los posibles efectos del cambio climático y contribuir a conservar los ecosistemas tropicales estacionalmente secos. En un matorral seco tropical se completó un ciclo anual de muestreo con trampas de interceptación en seis parcelas localizadas a distintas altitudes para comprobar la existencia de: 1) relaciones de variaciones estacionales entre humedad y oscilaciones diarias de temperatura y humedad con el número y abundancia de taxa, 2) concordancia entre los patrones de distribución temporal de las comunidades a distintas altitudes, 3) diferentes amplitudes en los periodos de abundancia de las comunidades según la altitud y 4) influencia de la altitud en las relaciones expuestas en el objetivo 1). De manera consistente en todas las altitudes, el número de taxa se correlacionó negativamente con la humedad relativa y positivamente con las diferencias termohigrométricas diarias. La abundancia se correlacionó negativamente con la humedad en dos parcelas. Las correlaciones de abundancia de taxa con la humedad fueron negativas, salvo para Diptera. Las correlaciones con las fluctuaciones termohigrométricas diarias fueron de diferente signo. Salvo para Scorpiones, Pseudoescorpiones, Acariformes y Psocoptera, estas relaciones fueron consistentes en todas las altitudes. La amplitud de los periodos de abundancia no varió con la altitud pero la distribución temporal de las abundancias no fue concordante entre todas las parcelas. Por consiguiente, las diferencias en las dinámicas estacionales no fueron debidas a variaciones climáticas ligadas a la altitud.

3. PALABRAS CLAVE

Al final del resumen aparece una lista de unas cinco-nueve palabras (o expresiones de varias palabras) que permitan al lector tener una idea general de los temas que se van a tratar. Estas palabras se utilizan para hacer búsquedas bibliográficas automatizadas, de manera que uno puede buscar todos los trabajos publicados que contengan una determinada palabra clave.

Las palabras-clave del artículo anterior son:

artrópodos; Ecuador; estacionalidad; humedad; temperatura

4. INTRODUCCIÓN

La Introducción debe tener una estructura de pirámide invertida, empezando con una aproximación más general al tema, estrechando progresivamente el tema de trabajo hasta terminar planteando la pregunta concreta. Debe contener la siguiente información

- Antecedentes sobre el tema. ¿Qué se sabe sobre el tema? ¿Cuál va a ser nuestra aportación?
- Objetivo. ¿Cuál es la pregunta que se intenta responder en el trabajo? Justificar la importancia o interés de dicha pregunta, basándose en los antecedentes planteados.
- Hipótesis. ¿Qué resultado esperamos obtener y por qué? Es fundamental hacer una buena justificación razonada de la hipótesis citando la bibliografía en que se basa.

A lo largo de la introducción se deben citar, entre paréntesis y de forma abreviada, los trabajos científicos consultados que refrendan determinadas afirmaciones, y que contribuyen a establecer los antecedentes. Son las "citas bibliográficas". La referencia completa (no abreviada) de cada una de estas citas debe aparecer al final del trabajo, en el apartado de "Bibliografía". Tanto las citas como las referencias bibliográficas deben seguir un formato preciso que se detalla en el apartado "8. Bibliografía".

***Cuadro IV.2**: Primer párrafo de la introducción de Castro, A., Espinosa, C.I. 2016. Dinámica estacional de invertebrados en un matorral seco tropical a lo largo de un gradiente altitudinal. Ecosistemas 25(2): 35-45. Doi.: 10.7818/ECOS.2016.25-2.05*

Las respuestas de los organismos a variables climáticas a lo largo de gradientes ambientales se ha utilizado para predecir las posibles consecuencias del cambio climático y de las actividades humanas en los ecosistemas naturales (Crimmins et al. 2011; Denny et al. 2014). Uno de los organismos que responden más rápidamente a cambios ambientales son los invertebrados, por lo que se han empleado frecuentemente como indicadores (Prather et al. 2013). En el Trópico se prevé que las modificaciones climáticas ocurran antes que en otras regiones del planeta (Mora et al. 2013), siendo prioritario entender las posibles consecuencias sobre la dinámica estacional de grupos de los invertebrados. Además, el rápido calentamiento del planeta ha sido argumentado como uno de los principales motores de la acelerada extinción de especies (Thomas et al. 2004; Stork, 2010).

5. MATERIAL Y MÉTODOS

Debe contener la siguiente información:

- Descripción del área de estudio si es un estudio de campo (situación geográfica, clima, suelo, vegetación...) o descripción del experimento (unidad experimental, réplicas, etc.)
- Variables implicadas. En un experimento controlado, condiciones en que se realiza.
- Método de muestreo (sectorizado, al azar o en gradiente, número de réplicas, fechas de muestreo, etc.).
- Material utilizado (solo si se trata de aparatos especializados, es decir, no es necesario mencionar materiales como regla, sobres, bolsas, lápiz, papel, etc.).
- Método de análisis elegido en función del tipo de variables (no hay que explicar cómo se hace el análisis). Programa estadístico utilizado.

Se recomienda pedir a algún colega que desconozca el tema que lea este apartado para comprobar si con la información proporcionada se siente capaz de repetir ese mismo estudio (pues es la prueba de una buena redacción del apartado).

6. RESULTADOS

Debe contener la siguiente información:

- Este apartado debe aportar una descripción escueta y sencilla (pero bien redactada) de los resultados obtenidos, sin interpretarlos (eso se hace en la discusión). En primer lugar se presentan los resultados de la estadística descriptiva (por ej. "Tras dos meses de crecimiento, las plantas cultivadas a 30ºC alcanzaron una altura casi el doble que las cultivadas a 15ºC") y en segundo lugar los resultados de la estadística inferencial (test de contraste de hipótesis). Por ejemplo: "Las diferencias entre ambos tratamientos fueron altamente significativas ($p<0.001$)". Es fundamental tener en cuenta que los resultados estadísticos no aportan ninguna información útil si no se acompañan de la estadística descriptiva.
- El texto debe reforzase con Tablas y/o Figuras que faciliten la visualización de los resultados. Éstas han de llevar una leyenda (texto que aparece encima de la tabla o debajo de la figura y que explica su contentido), precedida de su número de referencia (por ej. Tabla 2; Figura 1). Desde el texto se debe hacer referencia a todas las tablas y/figuras (por ej. "en la Figura 1 se puede observar la altura media de las plantas en cada tratamiento").
- Las tablas y/o figuras que se presentan en los resultados deben contener resultados de la estadística descriptiva (por ej. media aritmética de una medida en cada tratamiento ± desviación típica o desviación estándar), así como los resultados de la estadística inferencial (parámetro estadístico del test y valor de p).
- No debe haber redundancia entre tablas y figuras, ni entre éstas y el texto. Por ej. no se puede mostrar la misma información en forma de tabla y de figura.
- Las salidas estadísticas que ofrecen los programas, como RStudio, suelen contener más información de la que es necesaria para interpretar el resultado. Por tanto, hay que elaborar estas salidas, en vez de copiarla y pegarlas directamente.
- Nunca se deben poner las tablas de datos brutos recogidos por el investigador (estadillo). Para poder interpretarlas el lector necesitará coger la calculadora y empezar a hacer cálculos. Esa labor ya la ha hecho el investigador, por tanto se la debe dar hecha al lector.
- Las tablas y figuras deben indicar los nombres de las variables que aparecen, así como las unidades de las magnitudes que recogen. Por ej. "Peso (g)", "Altura (cm)".
- Las figuras deben mostrar el nombre completo de las variables representadas en cada eje, con sus unidades. Si se utilizan abreviaturas, hay que explicarlas en la leyenda.
- El número de decimales con los que aparecen los valores de las tablas y figuras debe ser homogéneo dentro de una misma variable (por ej. un valor no puede tener un decimal y el siguiente tres decimales). Además debe estar en concordancia con la precisión del instrumento de medida. Por ej. si medimos espesores foliares con un calibre que da una precisión de centésimas de mm, los resultados deben darse en mm con dos decimales (no con cinco).

Cuadro IV.3. Ejemplo de elaboración de resultados a partir de la salida estadística de RStudio

```
cor.test( ~ longitud.pico + peso.corporal, data=datos.pico, method = "pearson",
continuity = FALSE,conf.level = 0.95)
Pearson's product-moment correlation
data:  longitud.pico and peso.corporal
t = 3.5194, df = 8, p-value = 0.007853
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2942560 0.9452104
sample estimates:
      cor
0.7794691
```

Esta información se puede resumir en la siguiente frase: "La longitud del pico y el peso corporal de la población mostraron una correlación positiva y significativa (r=0.78, p=0.008, N=10)". En este caso no es necesario realizar ninguna tabla ni figura porque el resultado se sintetiza en una línea de texto.

Cuadro IV.4: Ejemplo de figura. Tomado de Castro-Díez et al. 2012. Effects of exotic and native tree leaf litter on soil properties of two contrasting sites in the Iberian Peninsula. *Plant Soil* 350: 179-191. DOI.: 10.1007/s11104-011-0893-9

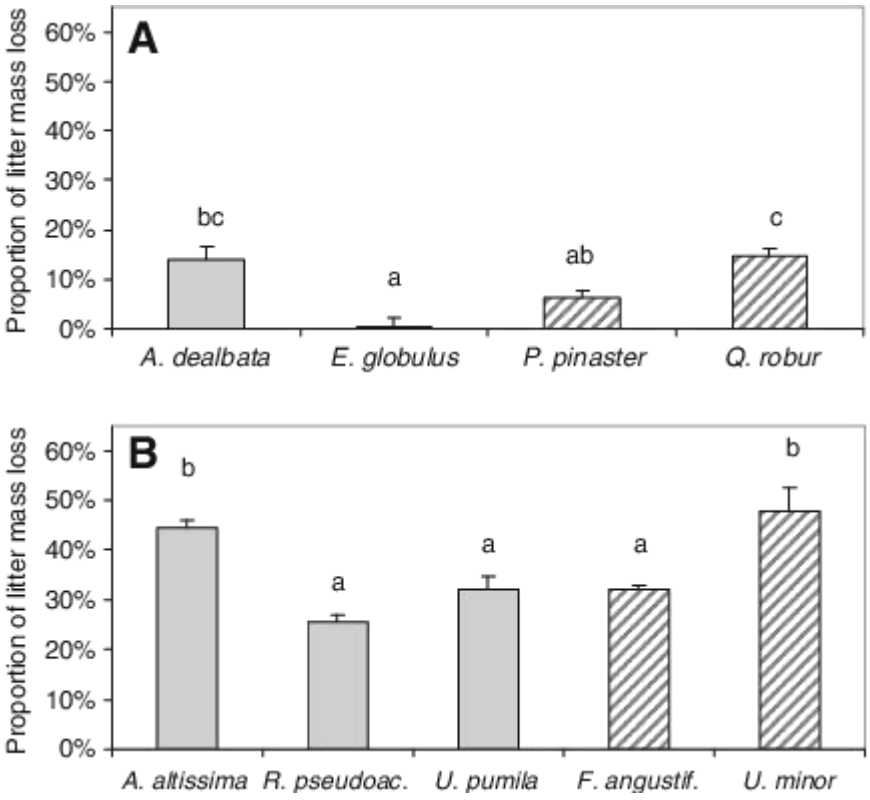

Fig. 1. Percentage of leaf litter mass loss after 9 months of incubation in a growth chamber. a) Species from Poulo (mesic forest of NW Spain) and b) species from Alcalá (riparian forest of central Spain). Grey columns represent exotic species and dashed columns native species. Different letters between columns in each graph means significant differences on the basis of an ANOVA test followed by a post-hoc Tukey test (Significant values P≤0.05)

Cuadro IV.5: Ejemplo de tabla tomada de Castro, A., Espinosa, C.I. 2016. Dinámica estacional de invertebrados en un matorral seco tropical a lo largo de un gradiente altitudinal. Ecosistemas 25(2): 35-45. DOI.:

Tabla 3. Valores mínimos y máximos (Min-Max) de los promedios de las variables climáticas para cada fecha de muestreo y cocientes (R) entre los rangos de valores intraparcelares e interparcelares. Abreviaturas como en la Tabla 1.

Table 3. Minimum and maximum values (Min-Max) of the averages of climatic variables for each sampling date, and quotient (R) between intra- and inter-plot ranges of values. Abbreviations as Table 1.

	T		HR		DTE		DHE	
	Min-Max (°C)	R	Min-Max (%)	R	Min-Max (°C)	R	Min-Max (%)	R
Parcela 1	24.9 - 26.4	0.2	43.8 - 79.8	4.5	17.6 - 23.6	0.7	36.2 - 55.9	2.5
Parcela 2	23.0 - 24.6	0.2	47.1 - 80.7	4.2	14.5 - 22.0	0.9	31.8 - 54.3	2.8
Parcela 3	21.3 - 24.4	0.4	48.7 - 82.0	4.2	11.3 - 23.7	1.6	27.4 - 54.5	3.4
Parcela 4	21.3 - 23.5	0.3	48.2 - 83.9	4.5	12.6 - 20.6	1.0	30.8 - 54.9	3.0
Parcela 5	20.9 - 23.3	0.3	49.1 - 82.0	4.1	13.3 - 21.8	1.1	31.6 - 54.8	2.9
Parcela 6	20.8 - 23.5	0.3	49.4 - 81.8	4.1	14.9 - 24.1	1.2	36.9 - 58.3	2.7

7. DISCUSIÓN

Debe contener la siguiente información

- Interpretación racional de los resultados: ¿por qué obtenemos tales resultados? ¿Se ajustan a nuestras expectativas? En caso positivo explicar por qué esperábamos esos resultados. En caso negativo buscar explicaciones alternativas. Hay que mantener una actitud abierta a la posibilidad de rechazar nuestra hipótesis inicial.
- Contraste de nuestros resultados con los obtenidos por otros autores en experiencias similares.
- La discusión ha de basarse en los resultados, en la bibliografía y en razonamientos bien fundados.
- Es fácil caer en una repetición de los resultados. Aunque puede haber una cierta repetición, ésta debe ser la mínima necesaria para facilitar la lectura. La sección Resultados "describe" los mismos, y la sección Discusión los "interpreta".

Cuadro V.6: Primer párrafo de la discusión tomado de: Castro, A., Espinosa, C.I. 2016. Dinámica estacional de invertebrados en un matorral seco tropical a lo largo de un gradiente altitudinal. Ecosistemas 25(2): 35-45. Doi.: 10.7818/ECOS.2016.25-2.05

Relaciones entre dinámicas de invertebrados y variables meteorológicas

En el matorral seco de Alamala se evidenció la importancia de los factores meteorológicos sobre las dinámicas estacionales de la comunidad de invertebrados. Así, tanto la riqueza taxonómica como la mayoría de taxa más abundantes se relacionaron con alguna variable meteorológica. Además, se añaden evidencias a la hipótesis de que las fluctuaciones termohigrométricas muestran también relación con las dinámicas estacionales de la diversidad (Checa et al. 2014). Sin embargo hubo resultados inesperados que se tratan a continuación.

8. BIBLIOGRAFÍA

CITAS BIBLIGRÁFICAS EN EL TEXTO

La introducción (especialmente en antecedentes), la discusión y, a veces, material y métodos, deben llevar citarse los trabajos de los que procede la información aportada (ver ejemplos de citas en los cuadros anteriores). Estas citas se ponen entre paréntesis, de forma abreviada, al final de la frase o párrafo que refrendan. La forma de citar esos trabajos es la siguiente:

- Si el trabajo tiene solo un autor, se cita el apellido de ese autor (nunca el nombre ni sus iniciales) seguido del año de publicación, por ejemplo: "La temperatura favorece la germinación de las semillas de *Pinus sylvestris* **(Smith, 1987)"**.
- Si tiene dos autores, se menciona el apellido de ambos (separados por "&", "y" o "and"), seguidos del año de publicación. Por ejemplo: "La velocidad de carrera de las lagartijas está influida por la temperatura (**García & Ibáñez, 2008**)".
- Si hay más de dos autores, se indica el apellido del primero seguido de "et al." y el año de publicación, por ejemplo: "El tamaño de las hojas tiende a ser mayor en las regiones del planeta con precipitación más elevada **(Mooney et al. 1978)"**.

La referencia bibliográfica completa debe aparecer al final del artículo, en la sección "bibliografía".

BIBLIOGRAFÍA

Al final del trabajo se expone una lista de las referencias bibliográficas **que han sido citadas a lo largo del texto** (y solo éstas), con un formato que incluya la siguiente información: Autor/es (apellido e inicial del nombre), año de publicación, título del artículo o capítulo de libro, título de la revista o del libro en que se publica. Si es una revista, ha de ir seguida del volumen y las páginas en que aparece el artículo. Si es un libro, detrás del título se pone el nombre de los editores del libro, la ciudad donde se ha publicado, la editorial y las páginas en que se encuentra el capítulo. Ejemplos:

- Mooney, H. A., Ferrar, P. J., Slatyer, R. O. (1978). Photosynthetic capacity and carbon allocation patterns in diverse growth forms of *Eucalyptus*. Oecologia 36: 103-111. (Referencia de artículo de una revista).
- Givnish TJ (1995). Plant stems: biomechanical adaptation for energy capture and influence on species distributions. En: Gartner BL (ed.) Plant stems: Physiology and functional morphology. San Diego, Academic Press: 3-49. (Referencia de un capítulo de libro editado (donde cada capítulo tiene un título y una autoría diferente).

Hay ciertos detalles de formato que dependen de cada revista. Por ej., el nombre de la revista y el título del libro a menudo se ponen en cursiva; en algunas revistas los apellidos de los autores van en mayúsculas; a veces el nombre de la revista está abreviado y otras veces no. Lo relevante en nuestro caso es que el formato sea siempre el mismo para todas las referencias.

Las referencias bibliográficas se ordenan alfabéticamente por el apellido del primer autor.

Nunca debe aparecer en la bibliografía un trabajo referencias que no ha sido citado en el texto. En otras palabras, debe haber una correspondencia plena entre los trabajos citados en el texto y los que aparecen en la bibliografía.

Cada vez en más frecuente en la bibliografía de artículos recientes encontrar el "doi" al final de la referencia (ver Cuadro V.7), que es la abreviatura de "*digital object identifier*". Se trata de una cadena de números/signos/caracteres que es única para cada trabajo disponible en medios digitales. Si

anteponemos http://doi.org/ a este identificador, accederemos directamente al sitio web donde se ha publicado el trabajo (por ej. https://doi.org/10.1016/j.scitotenv.2021.146141).

Cuadro V.7: Ejemplo de bibliografía tomado de: Castro, A., Espinosa, C.I. 2016. Dinámica estacional de invertebrados en un matorral seco tropical a lo largo de un gradiente altitudinal. Ecosistemas 25(2): 35-45. Doi.: 10.7818/ECOS.2016.25-2.05

Referencias

Aristophanous, M. 2010. Do your preservative preserve? A comparison of the efficacy of some pitfall traps solutions in preserving the internal reproductive organs of dung beetles. *ZooKeys* 34: 1-16. doi: 10.3897/zookeys.34.215

Barry, R.G. 2008. *Mountain Weather and Climate*. Cambridge University Press, New York, Estados Unidos.

Boinski, S., Fowler, N.L. 1989. Seasonal patterns in a tropical lowland forest. *Biotropica* 21: 223–233.

Castro, A., Espinosa, C.I. 2015. Seasonal diversity of butterflies and its relationship with woody-plant resources availability in an Ecuadorian tropical dry forest. *Tropical Conservation Science* 8 (2): 333-351. Disponible en: www.tropicalconservationscience.org

Checa, M.F., Rodriguez, J., Wilmott, K.R., Liger, B. 2014. Microclimate variability significantly affects the composition, abundance and phenology of butterfly communities in a highly threatened neotropical dry forest. *Florida Entomologist* 97 (1): 1-13.

Crimmins, T., Crimmins, M., Bertelsen, C.D. 2011. Onset of summer flowering in a 'Sky Island' is driven by monsoon moisture. *New Phytologist* 191: 468–479. doi: 10.1111/j.1469-8137.2011.03705.x

Denny, E.G., Gerst, K.L., Miller-Rushing, A.J., Tierney, G.L., Crimmins, T.M., Enquist, C.A.F. et al. 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology*. Doi: 10.1007/s00484-014-0789-5

9. BIBLIOGRAFÍA RECOMENDADA

Harvey, J. A. 2009. Preparing a paper for publication: an action plan for rapid composition and completion. Ann. Zool. Fennici, 46:158-164.