

# ANÁLISIS DE DATOS – 2º BIOLOGÍA

## PRÁCTICA 1

### Descripción de los datos

Al menos desde 1892 se sabía que los huevos de cuco (*Cuculus canorus*) eran característicos de la localidad en la que se encontraban. Un estudio de Edgard P. Chance (1940) titulado *The Truth About the Cuckoo* demuestra que el cuco vuelve año tras año al mismo territorio y pone sus huevos en nidos de una especie concreta de huésped. Además parece que el cuco sólo se aparea en su territorio. Por esta razón, se desarrollan subespecies geográficas, cada una de ellas con su especie de padres adoptivos. La selección natural asegura la supervivencia de los cucos mejor adaptados a poner huevos que sean adoptados e incubados por una especie en particular.

Los datos son la longitud (en mm) de los huevos de cuco según la especie del huésped. Las observaciones son originales de Oswald H. Latter (1902). En el archivo aparecen una gran cantidad de especies huéspedes pero seleccionaremos solamente las cuatro siguientes (las que tienen un mayor número de casos)

	Nombre inglés	Nombre científico	Nombre español
1	Meadow Pipit	<i>Anthus pratensis</i>	Bisbita común
2	Tree Pipit	<i>Anthus trivialis</i>	Bisbita arbóreo
3	Hedge Sparrow	<i>Prunilla modularis</i>	Acentor común
4	Robin	<i>Erithacus rubecula</i>	Petirrojo

### Referencias

Latter, H. O. (1902). The egg of *Cuculus canorus*: an inquiry into the dimensions of the cuckoo's egg relation and the relation of the variations to the size of eggs of the foster-parent, with notes of coloration. *Biometrika* **1**, 164–176.

### Guion para realizar la práctica

1. Descarga («baja») los datos de la red al escritorio.

Se trata de un fichero de texto `Latter_data_comma` en el que aparecen las longitudes de los huevos y la especie en cuyo nido se encontró.

2. Entra en Inicio → Programas → SPSS Statistics 21.0.

3. Cierra el cuadro de diálogo «¿Qué desea hacer?».

4. Selecciona Archivo → Leer datos de texto...

En el cuadro de diálogo selecciona: **Archivos de tipo: Todos los archivos**

5. Abre el archivo: `Latter_data_comma` que acabas de bajar.

Aparecerán una serie de pantallas (Pasos de 1 a 6). Debes poner especial cuidado en

**Paso 2 de 6** ¿Están incluidos los nombres...? Selecciona «Sí».

**Paso 4 de 6** Elige «Tabulador».

6. Aparecerán dos pantallas que se mantendrán durante todo el manejo de SPSS:

El **editor de datos** en el que es conveniente seleccionar Archivo → Guardar como... y guardar el archivo como *Practica1.sav* (que es el formato de fichero de datos en SPSS). En la parte inferior del editor aparecen dos pestañas: *Vista de datos* y *Vista de variables*. La pestaña *Vista de datos*

corresponde a la pantalla anterior y nos muestra los nombres y valores de las variables. La pestaña *Vista de variables* muestra las características de cada variable (Nombre, Tipo, Anchura, . . .).

El **Visor de resultados**, en él irán apareciendo los resultados de nuestro análisis estadístico (tablas, gráficos, . . .). Los resultados se pueden guardar al final en un fichero llamado *Practica1.spv*. Las tablas se pueden copiar como texto y los gráficos se pueden extraer en formato Postscript (.eps) o en formato pdf (.pdf).

6. En el editor o en el visor selecciona Gráficos → Generador de gráficos → Diagramas de caja

Arrastra el primero de los diseños que se muestra. Arrastra la variable *Longitud* al eje vertical del gráfico y la variable *Especie* al eje horizontal. Pincha en **Aceptar**. Con un doble clic en el gráfico resultante aparece el **Editor de gráficos**, que nos permite modificar los límites de los ejes, colores, etc.

### Cuestión 1

¿Para qué sirve este gráfico de los datos?

*Responde en la hoja de respuestas*

7. Hay otra manera de obtener el mismo gráfico y los estadísticos descriptivos (media, mediana, . . .) de la variable respuesta calculados por grupos. Selecciona Analizar → Estadísticos descriptivos → Explorar

Puedes dejar todas las opciones seleccionadas por defecto o cambiarlas (por ejemplo, seleccionar «Histograma» y «Gráficos con pruebas de normalidad»).

8. Queremos ver si hay diferencias entre la longitud media de un huevo de cuco en función de la especie huésped.

### Cuestión 2

Escribe el nombre y expresión matemática del modelo lineal que debemos emplear para analizar estos datos. Indica claramente quiénes son las variables del modelo en estos datos (cuál es la variable respuesta, etc.). Escribe el problema estadístico que nos planteamos en términos de un contraste de hipótesis.

*Responde en la hoja de respuestas*

9. Para poder realizar con SPSS un análisis de la varianza unifactorial tenemos que recodificar la variable *Especie* en una variable numérica ficticia que llamaremos (por ejemplo) *EspecieNum*. En la **Vista de variables** del **Editor de datos** crea la variable y luego recodifica la especie según los números asignados en la primera tabla de la práctica (también se puede utilizar la «Recodificación automática»).

A continuación selecciona Analizar → Comparar medias → ANOVA de un factor. En el cuadro de diálogo que aparecerá pasa *Longitud* a la *Lista de dependientes* y *EspecieNum* a la *Lista de Factores*. El botón **Contrastes** no hay que tocarlo. En **Post hoc** marca la opción *Bonferroni* y selecciona el nivel de significación global  $\alpha_T$  que deseamos (0,05 por ejemplo). En **Opciones** marca *Estadísticos descriptivos* y *Prueba de homogeneidad de varianzas*. No cambies nada donde pone Valores perdidos. No marques la opción *Gráfico de las medias*.

Al presionar **Aceptar** obtenemos la tabla ANOVA (donde pone *Sig.* ha calculado el *p*-valor del contraste) y en **Pruebas Post-Hoc**, los intervalos de confianza múltiples para las diferencias de medias.

### Cuestión 3

¿Qué conclusión se obtiene de la tabla ANOVA?

*Responde en la hoja de respuestas*

### Cuestión 4

¿Cuántas comparaciones de parejas de medias habría que realizar?

*Responde en la hoja de respuestas*

### Cuestión 5

En consecuencia, ¿cuál es el nivel individual de confianza con el que se han calculado los intervalos para  $\mu_i - \mu_j$ ?

*Responde en la hoja de respuestas*

Donde pone *Diferencia de medias (I - J)* se calcula la diferencia de medias muestrales  $\bar{y}_i - \bar{y}_j$ ; donde pone *Error típico* aparece  $S_r \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ ; la columna *Sig.* da el  $p$ -valor del contraste de igualdad de las medias; donde pone *Intervalo de confianza al 95 %* calcula los intervalos de confianza para las diferencias de las medias con esa confianza conjunta calculada por el método de Bonferroni.

### Cuestión 6

¿Qué conclusiones acerca de las relaciones entre las medias obtienes de los intervalos de confianza?

*Responde en la hoja de respuestas*

10. El procedimiento **ANOVA con 1 factor** de SPSS no calcula los residuos. Para calcularlos utilizamos el procedimiento del **modelo lineal general**. Pincha en Analizar → Modelo lineal general → Univariante. Pasa la *Longitud* a *Variable dependiente* y la *Especie* a *Factor*.

Pinchar en **Guardar** y elegir *Valores pronosticados no tipificados* y *Residuos no tipificados*. Pinchar en **Continuar** y luego en **Aceptar**.

Vemos que en el editor de datos han aparecido dos nuevas columnas con los valores predichos  $\hat{y}_{ij}$  en la columna PRE\_1 y los residuos  $e_{ij}$  en la columna RES\_1. Para representar los residuos frente a valores predichos, pincha en Gráficos → Generador de gráficos y elige *Diagrama de dispersión*.

### Cuestión 7

¿Qué conclusiones acerca de la hipótesis de homocedasticidad extraes del gráfico?

*Responde en la hoja de respuestas*

Para dibujar un histograma de los residuos pincha en Analizar → Estadísticos descriptivos → Frecuencias y pasa el *Residuo* a *Variables*. Pincha en el botón **Gráficos** y selecciona *Histograma con curva normal*. Luego pincha en **Continuar** y después en **Aceptar**.