## ANÁLISIS DE DATOS – 2º BIOLOGÍA PRÁCTICA 2

## 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 Chance (1940) titulado "La verdad sobre el cuco" 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 Latter (1902). Las especies son

	Nombre en el fichero	Nombre en inglés	Nombre en latín	Nombre en español
1	MDW PIPIT	Meadow Pipit	Anthus pratensis	Bisbita común
2	TREE PIPIT	Tree Pipit	Anthus triviales	Bisbita arbóreo
3	HDGE SPRW	Hedge Sparrow	Prunilla modularis	Acentor común
4	ROBIN	Robin	Erithacus rubecula	Petirrojo
5	PIED WTAIL	Pied Wagtail	Motacilla alba	Lavandera
6	WREN	Wren	Troglodytes	Chochín
			troglodytes	

## **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.

## Guión para realizar la práctica:

**1.** Descargar los datos de la red al escritorio o a una carpeta del disco duro (en cuyo caso es conveniente crear antes una carpeta para la práctica, por ejemplo en C:\Temp). Se trata de un fichero de texto Practica2.txt en el que aparecen las longitudes de los huevos y la especie en cuyo nido se encontró.

2. Entrar en Inicio → Programas → SPSS Statistics 17.0. Si muestra el siguiente cuadro de diálogo



elegir Abrir un origen de datos existente. Si no, ir a Archivo  $\rightarrow$  Abrir  $\rightarrow$  Datos. En el cuadro de diálogo

Buscar en:	🗧 🗀 Practicas 👻 🖻 🔛 🖿
	🔁 2009practica1.pdf 🛛 = -WRL2429.tmp
	🛃 2009practica2.pdf 🛛 🖬 -/WRL2628.tmp
Reciente	Latter1902.pdf 🔤 -/VRL3056.tmp
	Practica1.doc 🔤 ~WRL3190.tmp
	🗐 Practica1Grupo36.txt 🔟 ~WRL3223.tmp
	🗐 ~\$actica1.doc
Escritorio	🖬 ~WRL0279.tmp
	🖬 ~WRL0493.tmp
	📼 ~WRL0699.tmp
	🖬 ~WRL0761 tmp
MIS	-//RL0763.tmp
oumoritoo	□ ~\/RL0856.tmp
	🖬 ~WRL1407.tmp
3	國 ~WRL2406.tmp
Mi PC	Nomirro do orobiuo:
	Abrir
	Archivos de tino: Tadas las anchivos (t.t)

seleccionar Archivos de tipo: Todos los archivos (o Texto (\*.txt,\*.dat)) y abrir el fichero Practica2.txt. Seguir los siguientes pasos:

Asistente para la importación de texto: Paso 1 de 6	😫 Asistente para la importación de texto: Paso 2 de 6 🛛 🛛 🔀
C28 ki0 1 s1 28.5 G32 ki00 0 73 40 33 G33 ki00 0 33 10.8 G33 ki00 0 33 11.0 G33 fix0 0 3 41.0 G33 fix0 0 3 41.0 G33 fix0 0 5 41.0 G33 f	¿Cómo están organizadas sus variables? <li>Delimitado         <ul> <li>Las variables están delimitadas por un carácter concreto (coma, tabulador).</li> <li>Ancho tijo             <li>Las variables están alineadas en columnas de anchura fija.</li> </li></ul> </li>
ver1       ver2       ver3	r¿Están incluidos los nombres de las variables en la parte superior del archivo? ○ sí ⊙ N <u>o</u>
Archivo de texto: C:Amparo/Docencia/Matematicas/Ambientales/Estadística/Practicas/Pra	Archivo de texto: C:\Amparo\Docencia\Matematicas\Ambientales\Estadistica\Practicas\Pra
🕿 Asistente para la importación de texto - Delimitado: Paso 🗙	🛿 Asistente para la importación de texto - Delimitado: Paso 🔀
Asistente para la importación de texto - Delimitado: Paso   Can qué número de línea comienza el primer caso de datos?  Cada línea representa sus casos?  Cada línea representa un caso  Un número concreto de variables representa un caso:  Cuíntos casos desea importar?	Asistente para la importación de texto - Delimitado: Paso      Sué delimitador desea para la separación entre variables     Sué delimitador desea para la separación entre variables     Sué delimitador de texto?     Sué delimitador     Sué de la separación entre variables     Sué delimitador     Sué de la separación entre variables     Sué de la separación de texto?
Asistente para la importación de texto - Delimitado: Paso   En qué número de línea comienza el primer caso de datos?  Conso se representan sus casos?  Conso se representa un caso  Un número concreto de variables representa un caso:  Cuántos casos desea importar?  Tados los casos	Asistente para la importación de texto - Delimitado: Paso ¿Qué delimitador desea para la separación entre variables [] Isbulador    Barra espaciadora ] Coma    Punto y coma ] Otros:
Asistente para la importación de texto - Delimitado: Paso × ¿En qué número de línea comienza el primer caso de datos? ¿Cómo se representan sus casos? ③ Cada línea representa un caso ③ Cada línea representa un caso ② Un número concreto de variables representa un caso: ② Todos los casos ③ Los primeros 1000 casos. ③ Un gorcentaje aleatorio de los casos (aproximado): 10 20 30 40 50 60 10 19,655:2000/PIPTT 20,055:2000/PIPTT 20,055:2000/PIPTT 20,055:2000/PIPTT	Asistente para la importación de texto - Delimitado: Paso          ¿Qué delimitador desea para la separación entre variables         [ Jabulador       Barra espaciadora         © Oma       Punto y coma         Otgos:       Otros:         Vista previa de datos         20,65       MOVMPIPT         20,65       MOVMPIPT         20,65       MOVMPIPT         21,85       MOVMPIPT

📾 Asistente para la importación de texto: Paso 5 de 6 🛛 🛛 🔀	🖾 Asistente para la importación de texto: Paso 6 de 6 🛛 🛛 🔀
Especificaciones para las variables seleccionadas en la vista previa de datos Nombre de la yariable: Especie Formato ge datos: Cadena Caracteres: g	Image: State Stat
Vista previa de datos Longitud V2 19,65 MDW/PIPIT 20,05 MDW/PIPIT 20,85 MDW/PIPIT 20,85 MDW/PIPIT 21,65 MDW/PIPIT 21,65 MDW/PIPIT 21,65 MDW/PIPIT 21,65 MDW/PIPIT 21,65 MDW/PIPIT 21,65 MDW/PIPIT	Vista previa de datos Longitud Especie 19,65 MCVAPPIT 20,05 MCVAPPIT 20,05 MCVAPPIT 20,85 MCVAPPIT 21,85 MCVAPIT 21,85 M

**3.** Aparecen dos pantallas que se mantendrán durante todo el manejo de SPSS: El editor de datos

🖬 *Sin tít	tulo2 [Co	njunto_de	_datos1	] - 9
Archivo <u>E</u> dic	ión <u>V</u> er <u>D</u> at	os <u>T</u> ransformar	<u>A</u> nalizar	<u>G</u> ráfi
🗁 📙 🔔 🛛	📴 🔶 🖶	🔚 📭 🔐 🦊	•	
1 : Longitud	19,6	65		
	Longitud	Especie	var	
1	19,65	MDWPIPIT		
2	20,05	MDWPIPIT		
3	20,65	MDWPIPIT		
4	20,85	MDWPIPIT		
5	21,65	MDWPIPIT		
6	21,65	MDWPIPIT		
7	21,65	MDWPIPIT		
8	21,85	MDWPIPIT		
9	21,85	MDWPIPIT		
10	21,85	MDWPIPIT		
11	22,05	MDWPIPIT		
12	22,05	MDWPIPIT		
13	22,05	MDWPIPIT		
14	22,05	MDWPIPIT		

en el que es conveniente seleccionar Archivo  $\rightarrow$  Guardar como... y guardar el fichero 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,...).

La segunda pantalla es el Visor de resultados

🕮 *R	esulta	do 1	[Do	ocum	ento1]	- SPS	S Stat	istics	Visor					
Archivo	Edición	⊻e	Da	tos <u>I</u>	ransformar	Insertar	Formato	Analiza	r <u>G</u> ráficos	: Utilidades	Complement	ntos '	Ventana	Ayuda
🗁 🖬	🔔 🙆	3		<b>•</b> d	) III III III III III III III III III I	<b>.</b>	ବ୍ତ 🌑	星『	i 🖌 🗗	+++	+ -		1 7	<b></b>
	Resultad	0			GET	DATA TYPE=T FILE=' DELCAS DELIMI ARRANG FIRSTO 'IMPORT 'VARIAE congitu specia HE. COUTE. TE OUTF av'	XT C:\Ampo TERS="; TERS="; TERS="; XASE=1 VASE=AI LES= d F5.2 A9. TLE='C: SSED.	ro/Doo  ELIMI L	encia\M ED	atematica cia\Natema	s\Ambien aticas\A	mbien	s\Este htales	distica)

en el que 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 Postcript (.eps) o en formato pdf (.pdf).

**4.** En el editor o en el visor seleccionar Gráficos → Generador de gráficos → Diagramas de caja

Yearlables:       La presentación preliminar del gráfico utiliza datos de ejemplo	Generador de g	ráficos 🔀
Gatería         Elementos básicos         Grupos/ID de puntos         Titulos/notas al pie           Elja entre:         Favortos         Baras         Lineas         Popiedades del elemento           Lineas         Áreas         Opciones         Qpciones	Yeriables: Congitud Especie No se ha seleccionado ninguna variable	La presentación preliminar del gráfico utiliza datos de ejemplo
Sectores/Polar Dispersión/Puntos Histograma Méximos-mínimos Diagramas de caja Ejes dobles	Galería Elementos básico Elja entre: Pavortos Barras Lineas Areas Sectores:Polar Dispersión/Puntos Histograma Máximos-mínimos Diagramas de caja Ejes dobles	s Grupos/D de puntos Titulos/hotas al pie Propiedades del elemento Qpciones

Arrastrar la variable Longitud al eje vertical del gráfico y la variable Especie al eje horizontal. Pinchar en Aceptar. Haciendo 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

**5.** Hay otra manera de obtener el mismo gráfico y los estadísticos descriptivos (media, mediana, ...) de la variable respuesta calculados por grupos. Ir a Analizar  $\rightarrow$  Estadísticos descriptivos  $\rightarrow$  Explorar

😤 Explorar	X
	Estadísticos Gráficos Opciones
• Visualización       • Ambos     Estadísticos       Gráficos         Aceptar     Pegar       Restablecer     Cancelar	Ayuda

Se pueden dejar todas las opciones seleccionadas por defecto o cambiarlas.

**6.** 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

**7.** Para poder realizar un análisis de la varianza con un factor con SPSS tenemos que recodificar la variable Especie en una variable <u>numérica</u> ficticia que llamaremos EspecieNum. En la Vista de variables del Editor de datos introducir la variable como se indica a continuación

📴 Pract	ica1.sav [C	onjunto_	de_dato	s1] - SPS	SS Statistics	Editor de	datos				
Archivo <u>E</u> c	lición <u>∨</u> er <u>D</u> at	os <u>T</u> ransformar	<u>A</u> nalizar	<u>G</u> ráficos <u>U</u>	țilidades C <u>o</u> mplemei	ntos Ventana	Ayuda				
🗁 📕 🚑	📴 🔶 🔿	🏪 📭 🔐 🦊	i 📲 📩	🔡 🥼 📑	🏾 🐳 🍖 🦠 🖏						
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	
1	Longitud	Numérico	5	2		Ninguna	Ninguna	8	'≣ Derecha	🛷 Escala	
2	Especie	Cadena	9	0		Ninguna	Ninguna	9	📰 Izquierda	\delta Nominal	
3	EspecieNum	Numérico	1	0		Ninguna	Ninguna	8	'≡ Derecha	🛷 Escala	
4											
5											

Luego recodificar la especie según los números asignados en la primera tabla de la práctica:

	Longitud	Especie	EspecieNum	var
64	22,85	HDGESPRW	3	
65	23,05	HDGESPRW	3	
66	23,05	HDGESPRW	3	
67	23,05	HDGESPRW	3	
68	23,05	HDGESPRW	3	
69	23,45	HDGESPRW	3	
70	23,85	HDGESPRW	3	
71	23,85	HDGESPRW	3	
72	23,85	HDGESPRW	3	
73	24,05	HDGESPRW	3	
74	25,05	HDGESPRW	3	
75	21,05	ROBIN	4	
76	21,85	ROBIN	4	
77	22,05	ROBIN	4	
78	22,05	ROBIN	4	
79	22,05	ROBIN	4	
80	22,25	ROBIN	4	

A continuación en el Editor de datos o en el Visor de resultados seleccionar Analizar  $\rightarrow$  Comparar medias  $\rightarrow$  ANOVA de un factor. En el cuadro de diálogo que aparece

🖬 ANOVA de un fac	tor	X
Congitud	Lista d <u>e</u> dependientes:	Contrastes Post hoc Opciones
Aceptar <u>P</u> egar	Restablecer Cancelar	Ayuda

pasar Longitud a la Lista de dependientes y EspecieNum a la lista de Factor. El botón Contrastes no hay que tocarlo. En Post hoc marcar la opción Bonferroni y seleccionar el nivel de significación global  $\alpha_T$  que deseamos (0,05 por ejemplo). En Opciones marcar Estadísticos descriptivos y Prueba de homogeneidad de varianzas. No Cambiar nada donde pone Valores perdidos. No marcar la opción Gráfico de las medias.

Al presionar Aceptar obtenemos la tabla ANOVA (donde pone Sig. ha calculado el p-valor del contraste) y 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  $\mu_i - \mu_i$ ?

Responde en la hoja de respuestas

Donde pone Diferencia de medias (I-J) está calculando la diferencia de medias muestrales  $\overline{y_{F}}$ ,  $\overline{y_{j}}$ ,

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 y 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  $\mu_i$  obtienes de los intervalos de confianza?

Responde en la hoja de respuestas

**8.** El procedimiento ANOVA con 1 factor de SPSS no calcula los residuos. Para calcularlos utilizamos el procedimiento del modelo lineal general. Pinchar en Analizar  $\rightarrow$  Modelo lineal general  $\rightarrow$  Univariante. Pasar la Longitud a Variable dependiente y la Especie a Factor

🗟 Univariante		X
🔗 EspecieNum	Variable <u>d</u> ependiente:	<u>M</u> odelo Co <u>n</u> trastes
	Eactores fijos:	Gráficos
		Post hoc
	F <u>a</u> ctores aleatorios:	Guardar
	•	Opciones
	Covariables:	]
	Ponderación MCP:	
Aceptar <u>P</u> ega	r <u>R</u> establecer Cancelar	Ayuda

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

🖬 Univariante: Gua	rdar 🛛 🔀
Valores pronosticados     No tipificados     Pondgrados     Error típico     Diagnósticos     Distancia de Cook     Valores de influencia	Residuos         ✓ No tipificados         ○ Ponderados         ○ Tipificados         ○ Método de Student         ○ Eliminados
Estadísticos de los coeficie Crear estadísticos de los co Crear un nuevo conjunto Nombre de conjunto de d Escribir un nuevo archiv Archivo	ntes eficientes o de datos datos: ro de datos

Vemos que en el editor de datos han aparecido dos nuevas columnas con los valores de los valores predichos  $\tilde{y}_{ij}$  y los residuos  $e_{ij}$ . Para representar los residuos frente a valores predichos, pincha en Gráficos  $\rightarrow$  Generador de gráficos y elegir Diagrama de dispersión.



Responde en la hoja de respuestas

Para dibujar un histograma de los residuos pincha en Analizar  $\rightarrow$  Estadísticos descriptivos  $\rightarrow$  Frecuencias y pasa el Residuo a Variables.

🔁 Frecuencias		X
<ul> <li>✓ Longitud</li> <li>➡ Especie</li> <li>✓ EspecieNum</li> <li>➡ Valor pronosticado para</li> </ul>	Variables:	Estadísticos Gráficos Eormato
🗹 Mostrar tablas <u>d</u> e frecuencias		
Aceptar <u>P</u> egar	<u>R</u> establecer Cancelar	Ayuda

Pincha en el botón Gráficos y seleccionar Histograma con curva normal. Luego pincha en Continuar y después en Aceptar.

Histograma



Para calcular el coeficiente de asimetría y la curtosis de los residuos pincha en Analizar  $\rightarrow$  Estadísticos descriptivos  $\rightarrow$  Descriptivos y pasa el Residuo a Variables. Pincha en el botón Opciones y selecciona los estadísticos descriptivos de interés.

🔁 Descriptivos	•	×
<ul> <li>✓ Longitud</li> <li>✓ EspecieNum</li> <li>✓ Valor pronosticado para</li> </ul>	Variables:	Opciones
Guardar valores tipificados c	omo variables	
Aceptar <u>P</u> egar	Restablecer Cancelar	Ayuda

Dependiendo de los estadísticos que hayas seleccionado obtendrás una tabla de este estilo: Estadísticos descriptivos

	N	Media	Desv. típ.	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Error típico	Estadístico	Error típico
Residuo para Longitud	120	,0000	,88995	-,321	,221	,610	,438
N válido (según lista)	120						

Como regla aproximada, si el valor absoluto de la asimetría dividida por su error típico es mayor que 1,96 entonces la distribución de los residuos es asimétrica. Asimismo, si el valor absoluto de la curtosis dividida por su error típico es mayor que 1,96 entonces la distribución de los residuos no sigue una distribución normal.

<b>Cuestión 8:</b> ¿Qué conclusiones acerca de la hipótesis de normalidad extraes del histograma de los
residuos y de sus estadísticos descriptivos?
Pospondo on la boia do rospuestas

Responde en la hoja de respuestas