

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 seis siguientes (las que tienen un mayor número de casos)

	Nombre en data	Nombre inglés	Nombre científico	Nombre español
1	MDWPIPIT	Meadow Pipit	<i>Anthus pratensis</i>	Bisbita común
2	TREEPIPIT	Tree Pipit	<i>Anthus trivialis</i>	Bisbita arbóreo
3	HDGESPRW	Hedge Sparrow	<i>Prunilla modularis</i>	Acentor común
4	ROBIN	Robin	<i>Erithacus rubecula</i>	Petirrojo
5	PIEDWTAIL	Pied Wagtail	<i>Motacilla alba</i>	Lavandera
6	WREN	Wren	<i>Troglodytes troglodytes</i>	Chochín

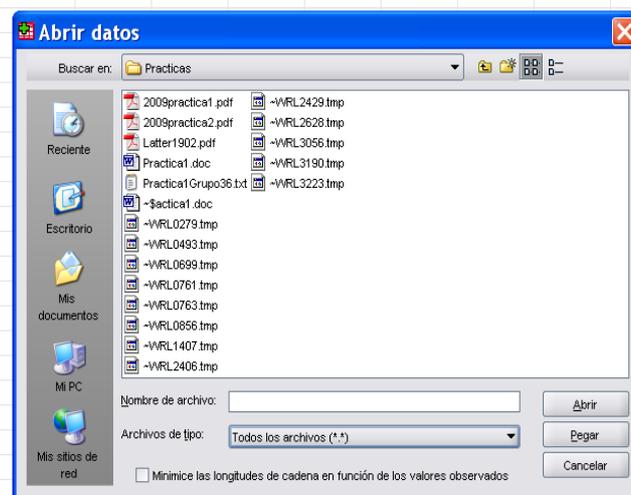
Referencias:

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. Se trata de un fichero de texto `Latter_data_length.csv` en el que aparecen las longitudes de los huevos y la especie en cuyo nido se encontró.
2. Entrar en Inicio → Programas → SPSS Statistics 21.0.

Ir a Archivo → Abrir → Datos. En el cuadro de diálogo



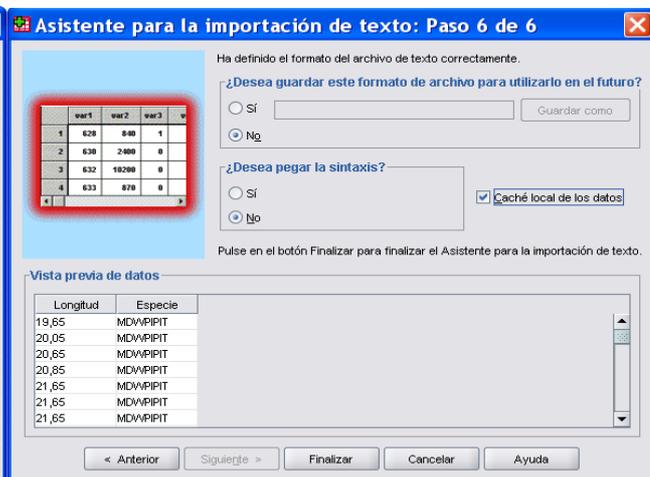
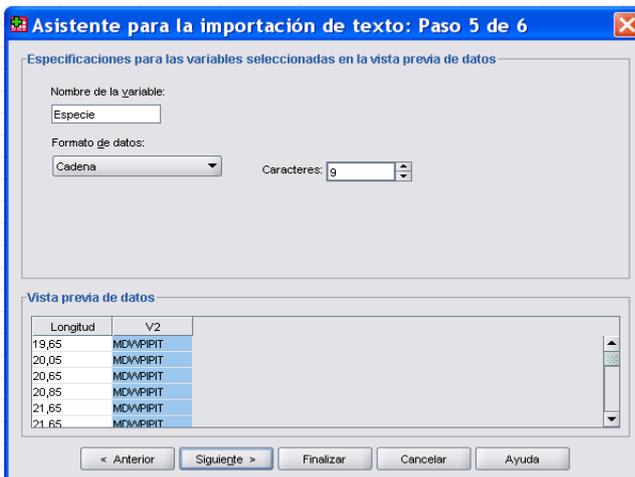
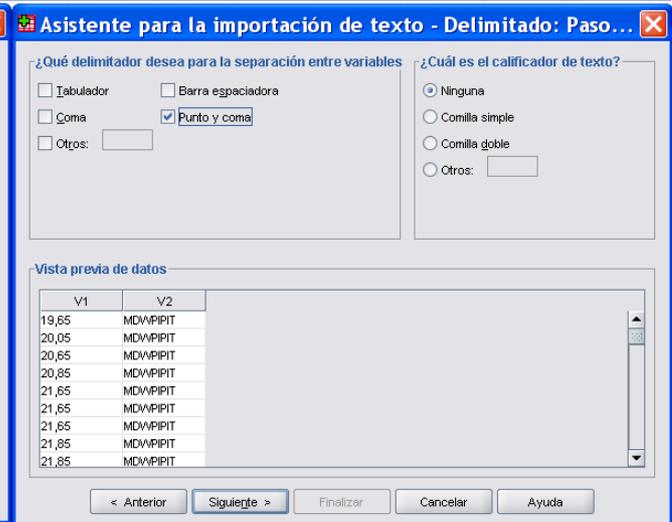
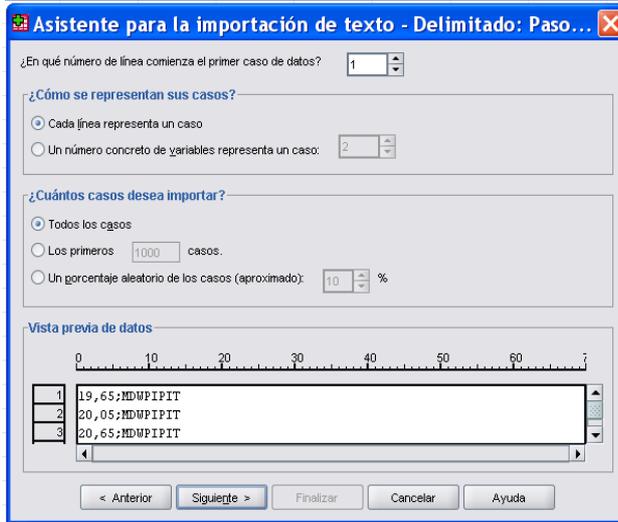
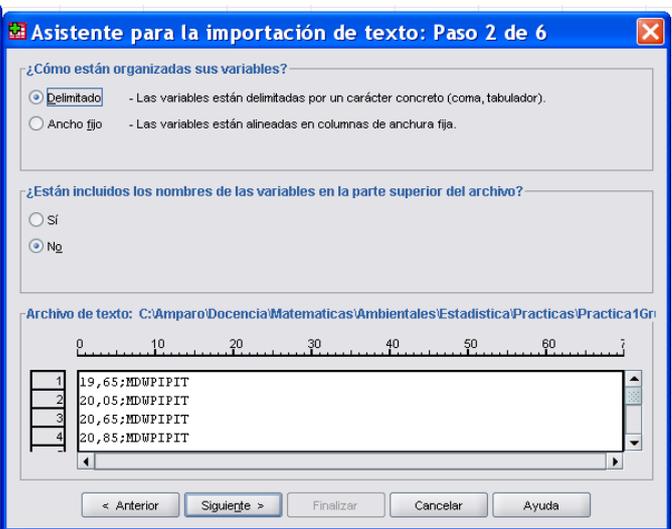
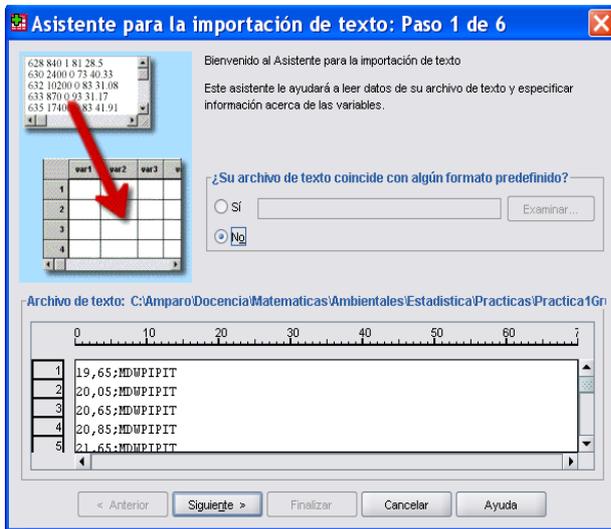
seleccionar

Archivos de tipo: Todos los archivos

y abrir el fichero

`Latter_data_length.csv`.

Seguir los siguientes pasos:



3. Aparecen dos pantallas que se mantendrán durante todo el manejo de SPSS:

1. El editor de datos

*Sin título2 [Conjunto_de_datos1] - SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos

1: Longitud 19,65

	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 → Guardar como... y guardar el fichero como Practical.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,...).

2. La segunda pantalla es el Visor de resultados

*Resultado1 [Documento1] - SPSS Statistics Visor

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Complementos Ventana Ayuda

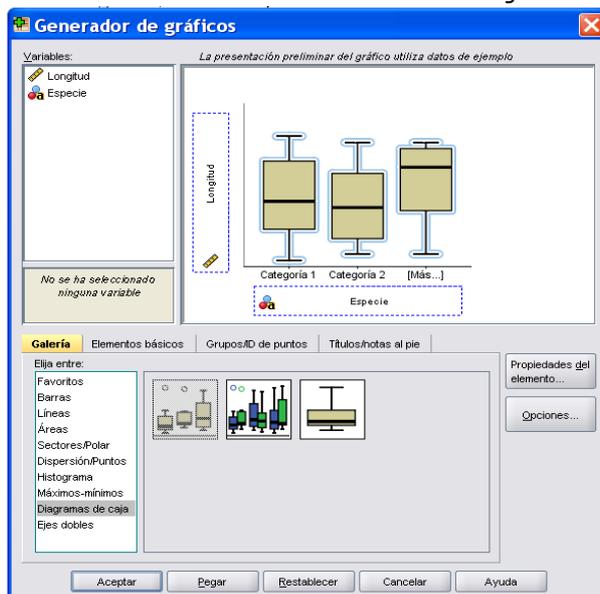
```

GET DATA
  /TYPE=TEXT
  /FILE='C:\Amparo\Docencia\Matemáticas\Ambientales\Estadística\o36.txt'
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
    Longitud F5.2
    Especie A9.
CACHE.
EXECUTE.
SAVE OUTFILE='C:\Amparo\Docencia\Matemáticas\Ambientales\Estadística\Practical.sav'
/COMPRESSED.

```

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

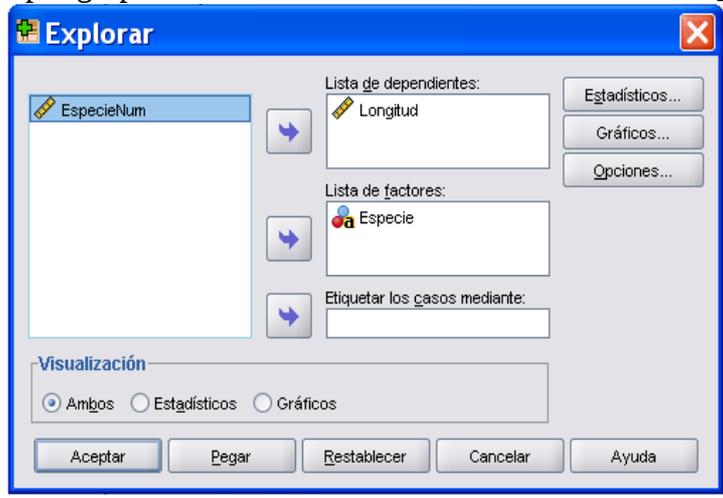
4. En el editor o en el visor seleccionar Gráficos → Generador de gráficos → Diagramas de caja



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` → `Estadísticos descriptivos` → `Explorar`



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.

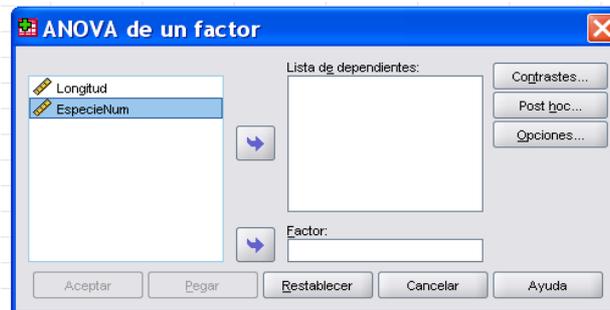
Cuestión 2: Escribe el nombre y expresión matemática del modelo lineal que debemos emplear para analizar estos datos. Indica claramente quiénes son las variables del modelo en estos datos (cuál es la variable respuesta, etc.). Escribe el problema estadístico que nos planteamos en términos de un contraste 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 numérica ficticia que llamaremos `EspecieNum`. En la Vista de variables del Editor de datos introducir la variable como se indica a continuación

	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	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.

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



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 α_T que deseamos (0,05 por ejemplo). En Opciones marcar Estadísticos descriptivos y Prueba de homogeneidad de varianzas. No cambies 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 $\mu_i - \mu_j$?

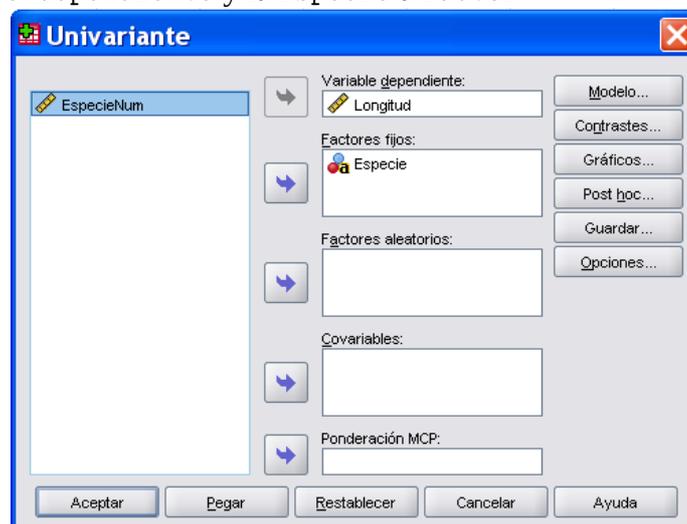
Responde en la hoja de respuestas

Donde pone Diferencia de medias (I-J) está calculando 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.

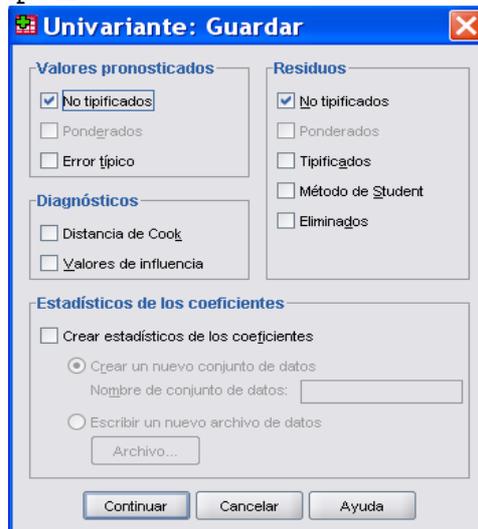
Cuestión 6: ¿Qué conclusiones acerca de las relaciones entre las medias μ_i obtienes de los intervalos de confianza?

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. 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 de los valores predichos \tilde{y}_{ij} y los residuos e_{ij} . Para representar los residuos frente a valores predichos, pincha en Gráficos → Generador de gráficos y elige Diagrama de dispersión.

Obtendrás el gráfico

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.

