

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

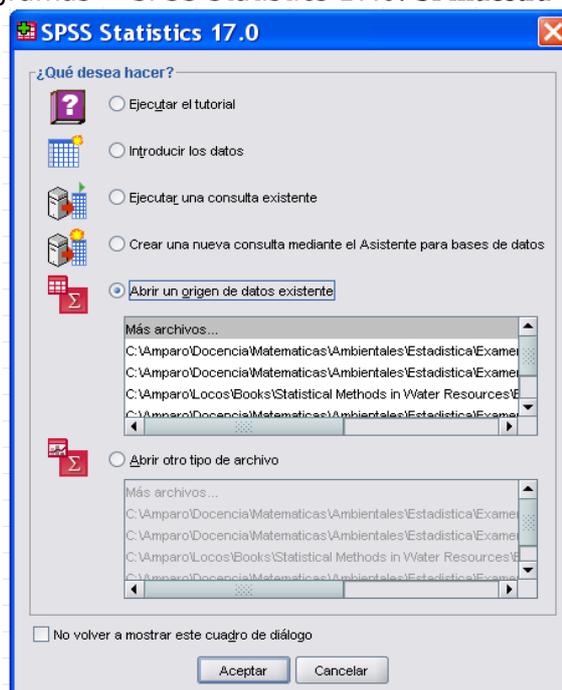
	Nombre en el fichero	Nombre en inglés	Nombre en latín	Nombre en español
1	MDW PIPIT	Meadow Pipit	<i>Anthus pratensis</i>	Bisbita común
2	TREE PIPIT	Tree Pipit	<i>Anthus trivialis</i>	Bisbita arbóreo
3	HDGE SPRW	Hedge Sparrow	<i>Prunilla modularis</i>	Acentor común
4	ROBIN	Robin	<i>Erithacus rubecula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

Referencias:

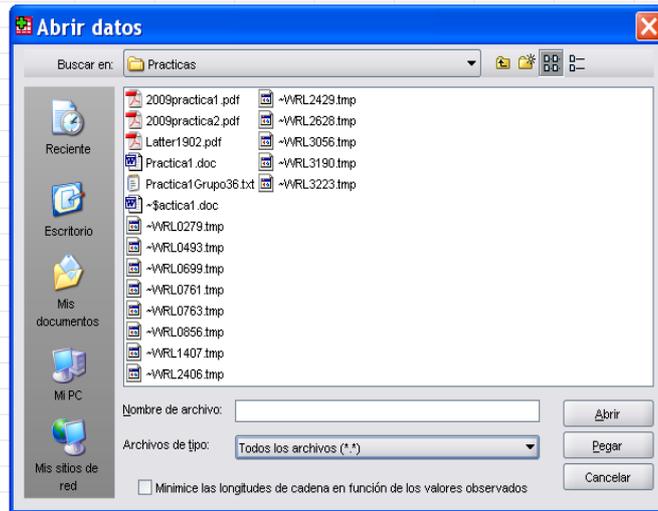
Latter, H. O. (1902). The egg of *Cuculus canorus*: an inquiry into the dimensions of the cuckoo's egg relation and the relation of the variations to the size of eggs of the foster-parent, with notes of coloration. *Biometrika* 1, 164–176.

Guión para realizar la práctica:

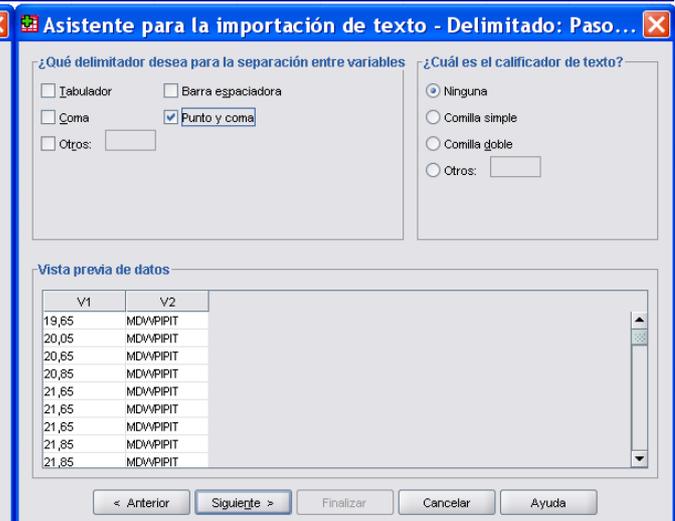
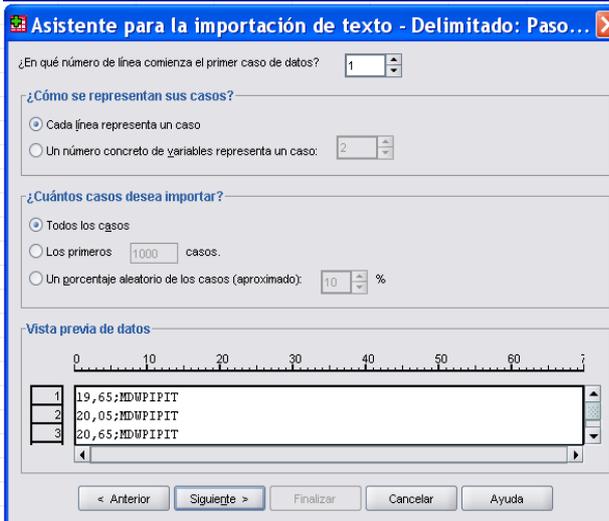
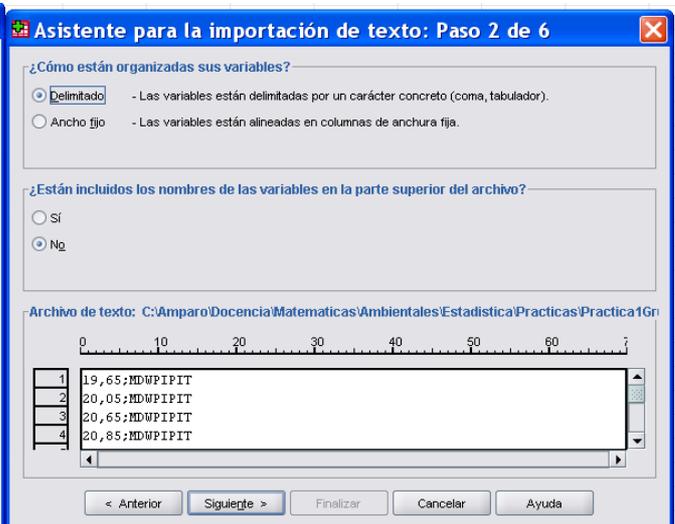
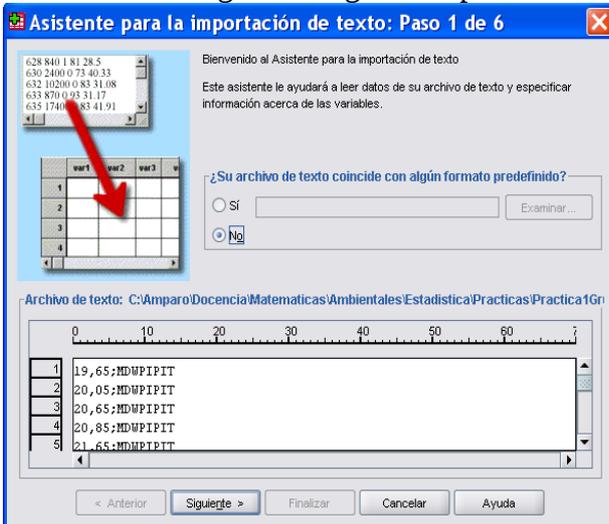
1. Descargar los datos de la red al escritorio o a una carpeta del disco duro (en cuyo caso es conveniente crear antes una carpeta para la práctica, por ejemplo en C:\Temp). Se trata de un fichero de texto Practica2.txt en el que aparecen las longitudes de los huevos y la especie en cuyo nido se encontró.
2. Entrar en Inicio → Programas → SPSS Statistics 17.0. Si muestra el siguiente cuadro de diálogo

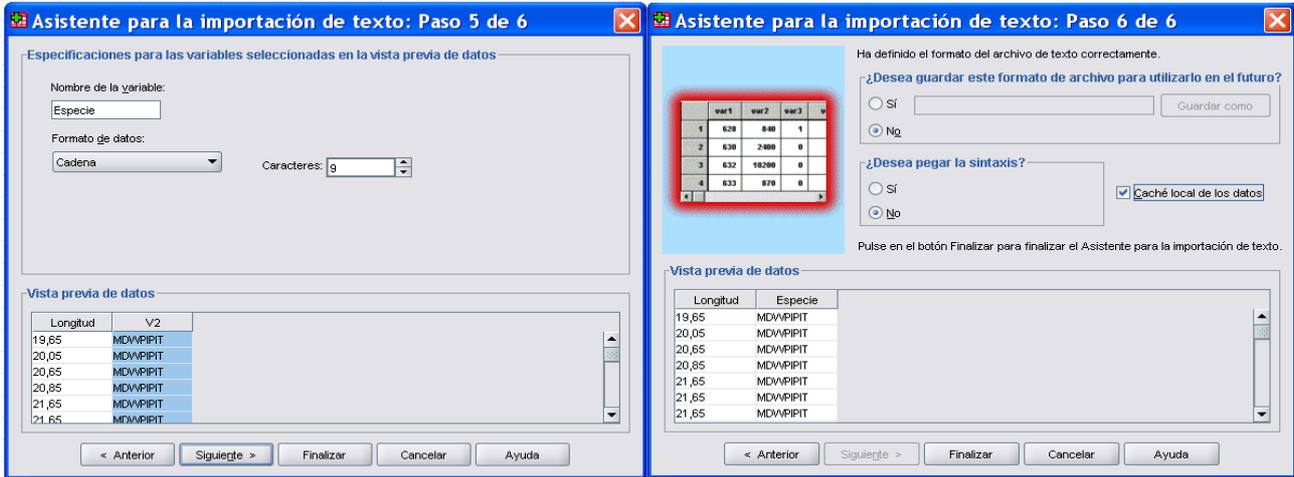


elegir Abrir un origen de datos existente. Si no, ir a Archivo → Abrir → Datos. En el cuadro de diálogo

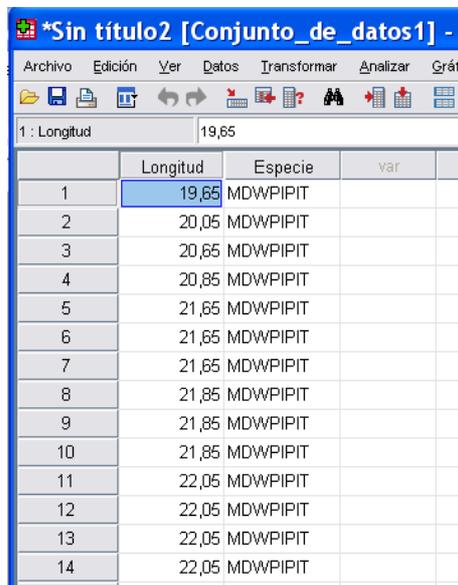


seleccionar Archivos de tipo: Todos los archivos (o Texto (*.txt,*.dat)) y abrir el fichero Practica2.txt. Seguir los siguientes pasos:





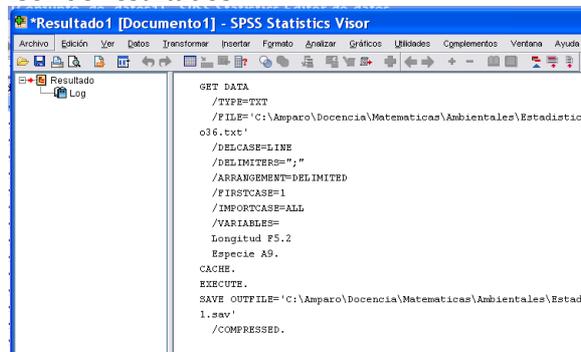
3. Aparecen dos pantallas que se mantendrán durante todo el manejo de SPSS: El editor de datos



en el que es conveniente seleccionar Archivo → Guardar como... y guardar el fichero como Practica1.sav (que es el formato de fichero de datos en SPSS).

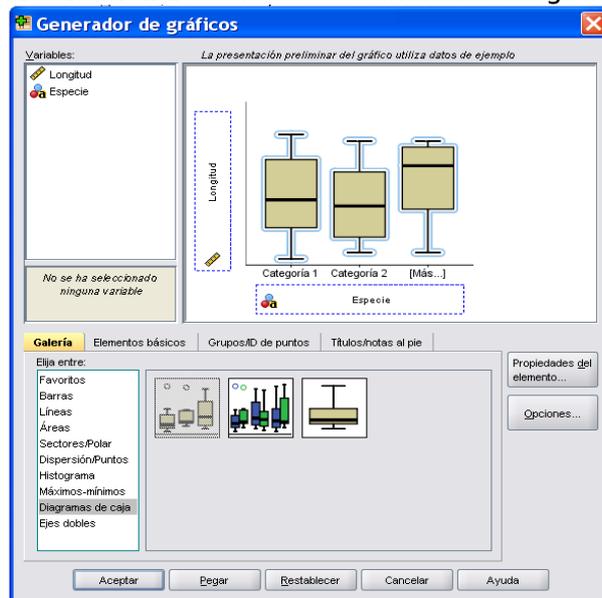
En la parte inferior del editor aparecen dos pestañas: Vista de datos y Vista de variables. La pestaña Vista de datos corresponde a la pantalla anterior y nos muestra los nombres y valores de las variables. La pestaña Vista de variables muestra las características de cada variable (Nombre, Tipo, Anchura,...).

La segunda pantalla es el Visor de resultados



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 Postscript (.eps) o en formato pdf (.pdf).

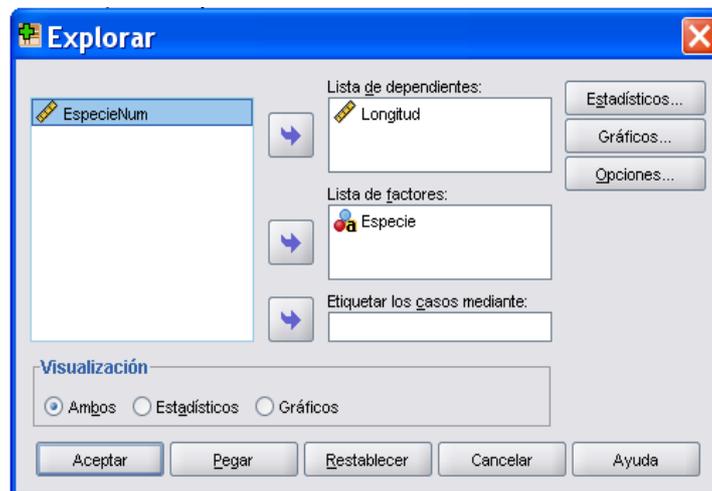
4. En el editor o en el visor seleccionar Gráficos → Generador de gráficos → Diagramas de caja



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

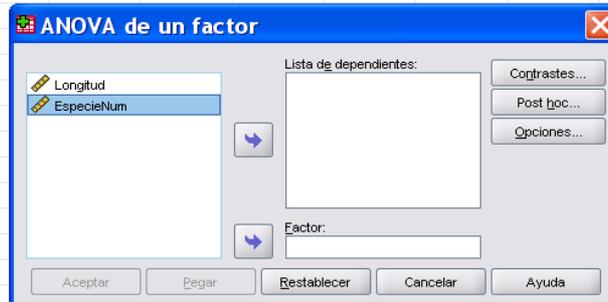
7. Para poder realizar un análisis de la varianza con un factor con SPSS tenemos que recodificar la variable Especie en una variable numérica ficticia que llamaremos EspecieNum. En la Vista de variables del Editor de datos introducir la variable como se indica a continuación



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

	Longitud	Especie	EspecieNum	var
64	22,86	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 → 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 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 $\mu_i - \mu_j$?

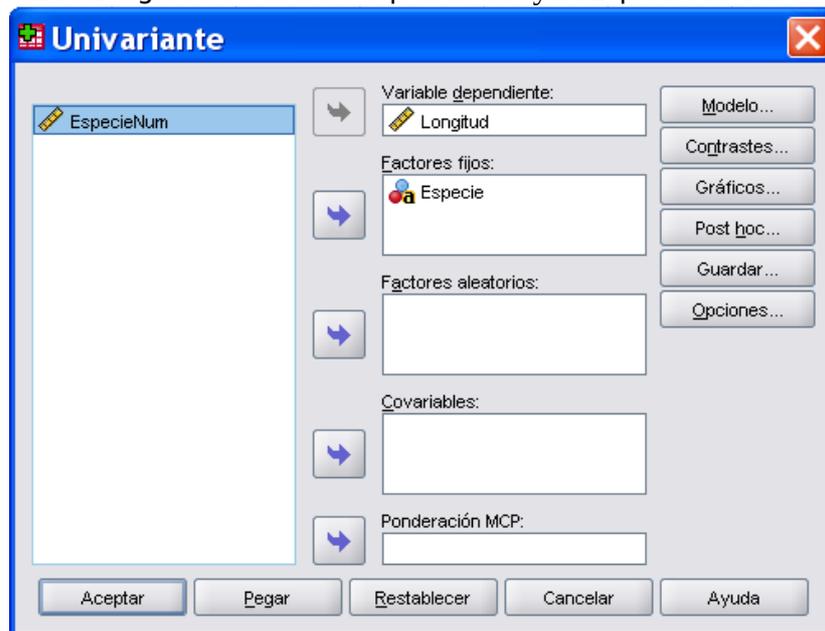
Responde en la hoja de respuestas

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 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.

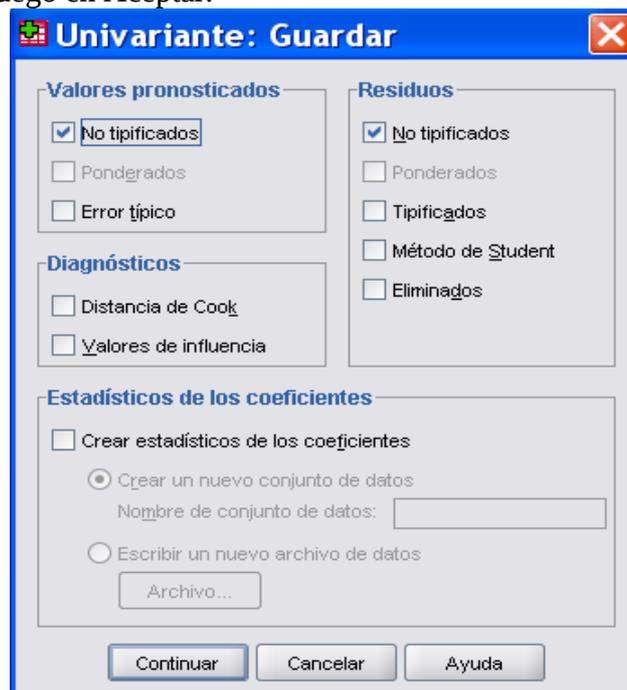
Cuestión 6: ¿Qué conclusiones acerca de las relaciones entre las medias μ_i obtienes de los intervalos de confianza?

Responde en la hoja de respuestas

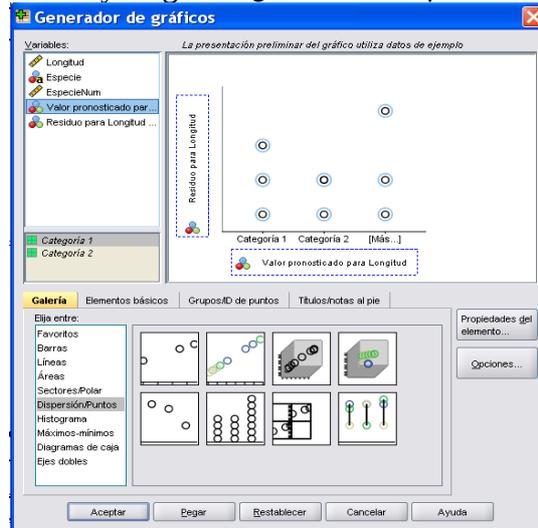
8. El procedimiento ANOVA con 1 factor de SPSS no calcula los residuos. Para calcularlos utilizamos el procedimiento del modelo lineal general. Pinchar en Analizar → Modelo lineal general → Univariante. Pasar la Longitud a Variable dependiente y la Especie a Factor



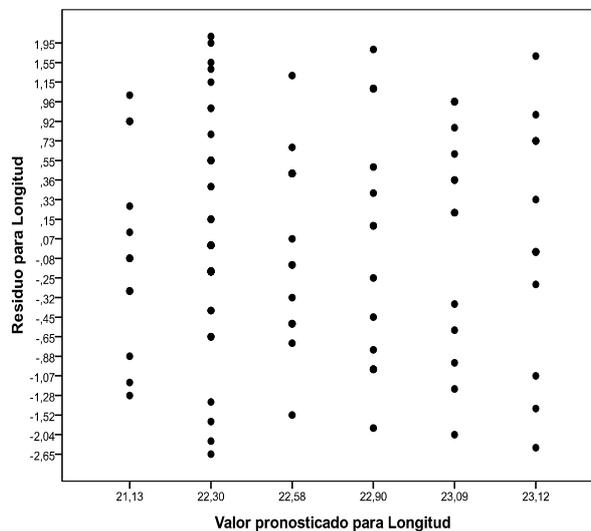
Pinchar en Guardar y elegir Valores pronosticados no tipificados y Residuos no tipificados. Pinchar en Continuar y luego en Aceptar.



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



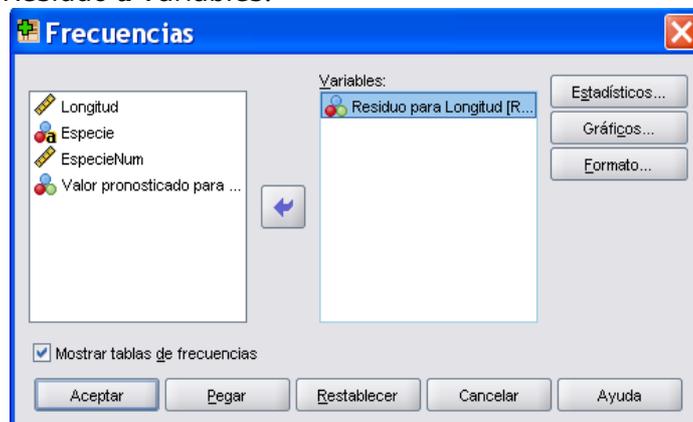
y obtendrás el gráfico



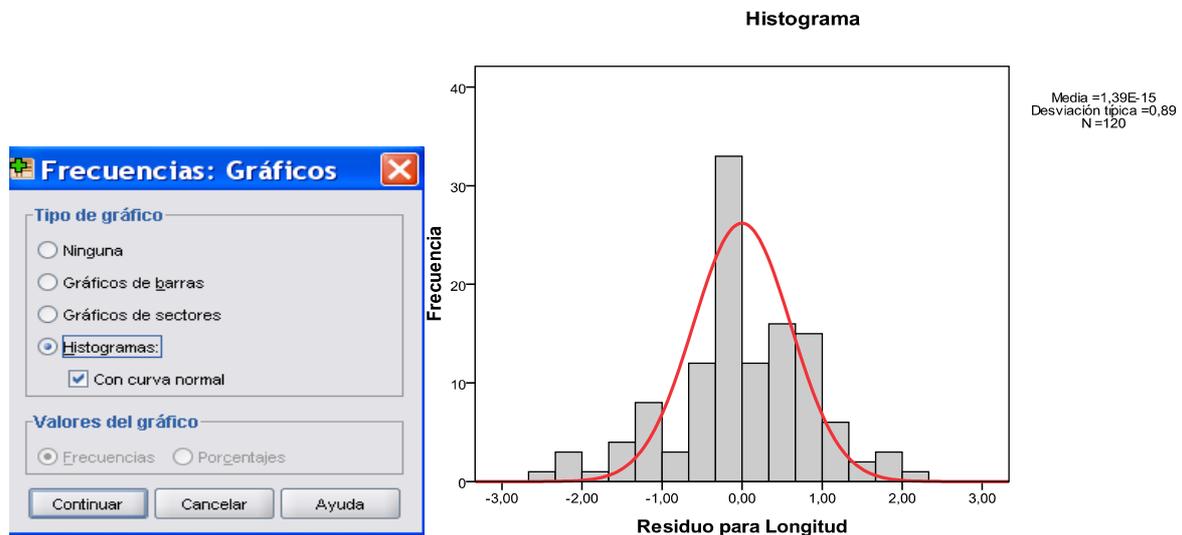
Cuestión 7: ¿Qué conclusiones acerca de la hipótesis de homocedasticidad extraes del gráfico?

Responde en la hoja de respuestas

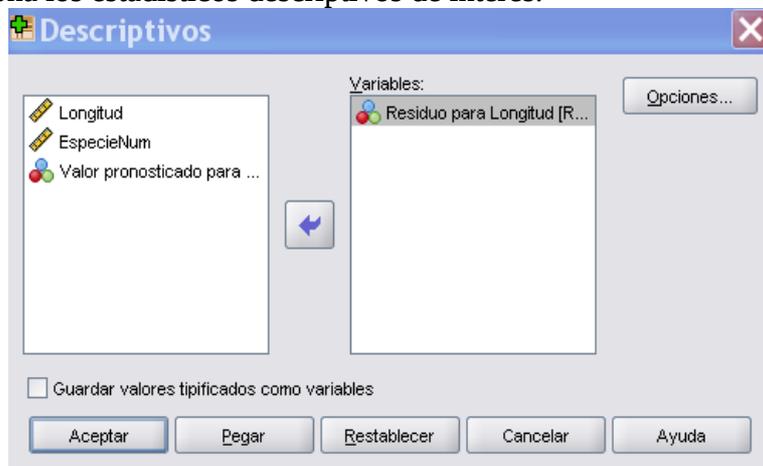
Para dibujar un histograma de los residuos pincha en Analizar → Estadísticos descriptivos → Frecuencias y pasa el Residuo a Variables.



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



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



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.

Cuestión 8: ¿Qué conclusiones acerca de la hipótesis de normalidad extraes del histograma de los residuos y de sus estadísticos descriptivos?

Responde en la hoja de respuestas