



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

INFORME FINAL

Ayudas de la USAL para Proyectos de Innovación Docente

PROYECTO DE INNOVACIÓN DOCENTE

“Guía de aprendizaje de las prácticas con SPSS de las asignaturas de Estadística de los Grados en Químicas e Ingeniería Química de la Universidad de Salamanca”

Código del proyecto: ID2015/0063

RESPONSABLES DEL PROYECTO:

M ^a José Fernández Gómez (Coordinadora):	mjfg@usal.es
Inmaculada Barrera Mellado	ibm@usal.es
Javier Martín Vallejo	jmv@usal.es
Santiago Vicente Tavera	svt@usal.es
Rosa Amanda Sepúlveda Correa	a108813@usal.es
Jesús Martín Rodríguez	jmr@usal.es

Dirigido a la Sra. Vicerrectora de Docencia de la Universidad de Salamanca

1.- INTRODUCCIÓN

En este documento se presenta el informe final del trabajo realizado en el Proyecto de Innovación Docente “Guía de aprendizaje de las prácticas con SPSS de las asignaturas de Estadística de los Grados en Químicas e Ingeniería Química de la Universidad de Salamanca” (ID2015/0063), realizado dentro de la convocatoria de ayudas de la Universidad de Salamanca a Proyectos de Innovación y mejora Docente Curso 2015-2016.

Ambos grados presentan particularidades comunes en cuanto a la naturaleza de los datos que en ellos se analizan, y en cuanto a que la asignatura de Estadística pertenece en los dos al Bloque Formativo Básico, que hace que tenga un papel muy parecido dentro del bloque y de plan formativo de cada grado.

La idea surgió cuando los profesores del Departamento de Estadística que impartimos docencia en los grados de la Facultad de Química, nos dimos cuenta de que estábamos realizando de manera individual materiales docentes de prácticas de las asignaturas de forma que se multiplicaban los esfuerzos y no se optimizaban los recursos disponibles, ya que la búsqueda de bases de datos, creación de guiones de prácticas, material impreso y virtual, etc. se hacían por separado, y que aunque se realizaba con un criterio bastante similar, suponían una pérdida de tiempo y de esfuerzo para llegar a los mismos y quizás no óptimos resultados.

La docencia de estas asignaturas tiene un carácter marcadamente práctico, en la cual la utilización de un programa estadístico es fundamental para que el alumno adquiera todas las competencias fijadas en ellas.

La Universidad de Salamanca cuenta con numerosas aulas de informática, con un gran horario de acceso libre para todos los estudiantes, y en todas ellas se encuentra el programa SPSS, que es el que se utiliza en estas asignaturas para que los estudiantes aprendan como llevar a la práctica los conocimientos teóricos adquiridos en las clases de teoría y problemas. Esta ha sido otra de las razones que nos impulsaron a la elaboración del presente proyecto.

Además, se pretendía favorecer la comunicación entre los profesores para diseñar materiales y aunar criterios docentes, y poner en marcha en el próximo curso académico una visión global de la docencia de las prácticas de esta asignatura en la Facultad, evitando así que la comunicación sólo se diese entre profesor de teoría-profesor de prácticas.

Por eso todos los profesores que imparten docencia en esta materia en los grados en Químicas e Ingeniería Química han participado en el proyecto, tanto si eran profesores responsables o no de la asignatura y tanto si impartían clases de teoría y/o prácticas, ya que la mejor forma de que el proyecto tuviese éxito es conociendo la opinión de todos y unificar criterios a la hora de seleccionar los contenidos y a la hora de elaborar los materiales.

2.- OBJETIVOS

Este proyecto se planteó con dos objetivos fundamentales: primero unificar criterios docentes y elaborar un material de prácticas común, con contenidos comunes que pudiese ser utilizado en los dos grados que se imparten en la Facultad de Químicas y segundo, que ese primer objetivo fuese la excusa para involucrar a todos los profesores que impartían docencia de Estadística en los grados de la Facultad de Químicas para rentabilizar tanto el esfuerzo y los recursos entre todos los docentes, con el fin de conseguir mayor eficacia en el proceso de enseñanza-aprendizaje de esta materia.

Tras la materialización del proyecto consideramos que dichos objetivo han sido exitosamente alcanzados.

3.- ACTUACIONES REALIZADAS

3.1.- Actividades de coordinación

Tal como se estableció en la memoria de solicitud del proyecto.

PRIMERA REUNIÓN: Una vez que se nos notificó la concesión del Proyecto, la primera reunión se realizó justo al regreso de vacaciones de Navidad, en la primera semana docente. En dicha reunión se distribuyó el trabajo de preparación del material según los bloques temáticos en los que se subdividió la elaboración de la guía de prácticas. Así cada profesor se le asignó la realización de un bloque temático.

También se consensuó la forma y el modo de preparar el material, se aunaron criterios en la selección de los datos y en la selección y elaboración de los contenidos.

También se preparó una plantilla para que todos los profesores la utilizaran a la hora de elaborar el proyecto.

SEGUNDA REUNIÓN: Se realizó en la primera semana del mes de Febrero, en la que se presentaron diferentes bases de datos por parte de los profesores, y se eligieron dos de ellas una para la elaboración de la guía aprendizaje y otra para el documento de trabajo autónomo del estudiante.

TERCERA REUNIÓN: Se realizó en la primera semana del mes de Abril. En dicha reunión los profesores compartieron sus impresiones y comentaron cómo iban desarrollando el trabajo que se estaba realizando. Se discutió también sobre algún contenido más puntual de alguno de los bloques temáticos, y se fijaron los plazos para la entrega de todo el material a uno de los miembros del equipo para la maquetación final de las guías de aprendizaje. Esta reunión favoreció la valoración secuencial de los resultados y permitió realizar un feedback en la elaboración de los contenidos.

CUARTA REUNIÓN: En la segunda semana de Junio nos reunimos por última vez para comentar la experiencia establecer alguna sugerencia de última hora para la elaboración final del material, alguna corrección, etc. y tratar de establecer las líneas generales del informe final.

3.2.-Trabajo individual de cada profesor

Cada profesor trabajó de forma individual en la confección del material sobre el bloque temático asignado, de forma autónoma pero bajo los criterios comunes establecidos en las reuniones de coordinación.

Todo el material se entregó a uno de los miembros del equipo para la maquetación final de la guía de prácticas.

3.3.-Material elaborado

En los ANEXOS I y II se presenta el material confeccionado. Las bases de datos empleadas para el análisis estadístico: EMBALSES.sav y Cereal.sav al ser archivos de datos de SPSS no se adjuntan al informe.

El primer Anexo hace referencia a la Guía de aprendizaje propiamente dicha (en el que se ha trabajado con los datos EMBALSES.sav) y el segundo a un documento de ampliación de objetivos y preguntas para que los estudiantes afiancen de forma autónoma las competencias adquiridas (en el que el estudiante tiene que utilizar el archivo Cereal.sav).

Señalar que el material elaborado se implementará en STUDIUM el próximo curso académico.

1.-GUÍA DE APRENDIZAJE: Este primer documento se estructura de la siguiente forma:

1.1.- Introducción: en ella se plantea forma detallada la base de datos –de un problema real- y la problemática de dichos datos. Esta base de datos se utilizará para aplicar las distintas técnicas de análisis estadístico aprendidas en el curso y la problemática de los mismos.

En la figura 1 se presenta una parte de esta introducción sobre la descripción de los datos y en la 2 una captura de pantalla de la base de datos en el SPSS.

Los datos de los que dispone hacen referencia al análisis de las aguas de dos embalses, uno en la provincia de León y otro en la provincia de Cáceres. ¶

*La toma de muestras se realizó durante 10 semanas, el mismo día cada semana en cada embalse, en una estación representativa de cada embalse, situada en posición central. El muestreo se extendió, desde el 16 de abril al 25 de junio, de modo que las 5 primeras semanas se corresponden con un período de **mezcla** que se debe a que las capas superficiales que se van enfriando se hunden, produciéndose así constantemente la mezcla con el agua profunda, y las últimas 5 semanas se corresponden con una fase de **estratificación**, que en general es debida al calentamiento diferencial de las capas superficiales con respecto a las profundas. ¶*

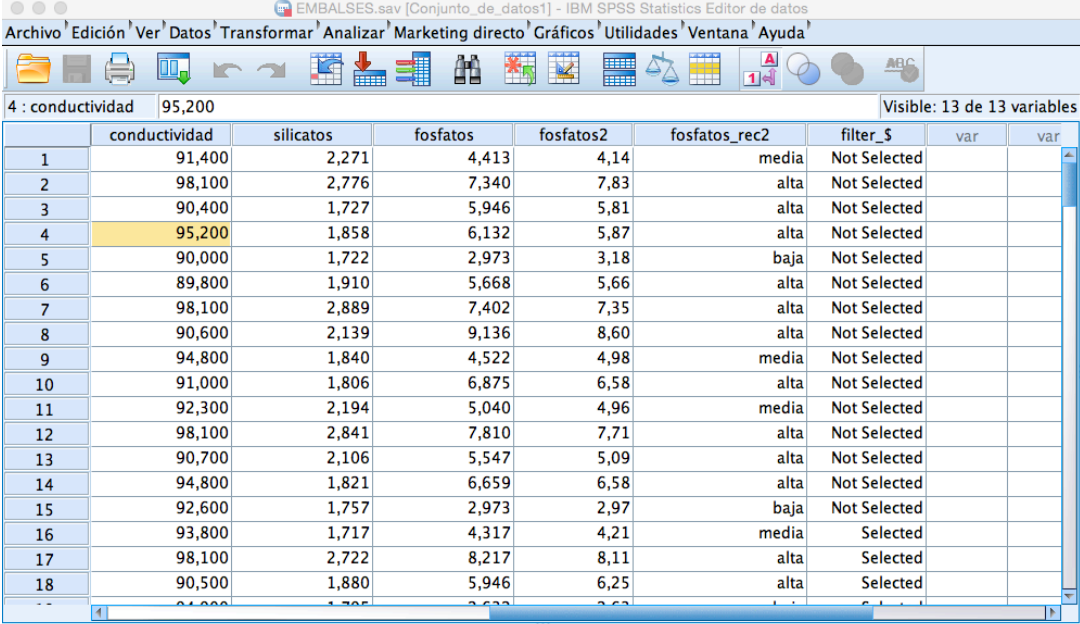
Las muestras se tomaron desde la superficie hacia el fondo a: 0, 2, 5, 10, 15, 20 metros de profundidad y fondo (2 metros por encima de la profundidad máxima de cada embalse). ¶

Las variables físico-químicas analizadas en las muestras fueron las siguientes: ¶

Temperatura	Oxígeno	pH	Conductividad	Silicatos	Fosfatos1	Fosfatos2
--------------------	----------------	-----------	----------------------	------------------	------------------	------------------

Figura 1: Descripción de la base de datos (información parcial de una página)

En la figura 2 se muestra una captura de pantalla de las base de datos utilizada.



	conductividad	silicatos	fosfatos	fosfatos2	fosfatos_rec2	filter_\$	var	var
1	91,400	2,271	4,413	4,14	media	Not Selected		
2	98,100	2,776	7,340	7,83	alta	Not Selected		
3	90,400	1,727	5,946	5,81	alta	Not Selected		
4	95,200	1,858	6,132	5,87	alta	Not Selected		
5	90,000	1,722	2,973	3,18	baja	Not Selected		
6	89,800	1,910	5,668	5,66	alta	Not Selected		
7	98,100	2,889	7,402	7,35	alta	Not Selected		
8	90,600	2,139	9,136	8,60	alta	Not Selected		
9	94,800	1,840	4,522	4,98	media	Not Selected		
10	91,000	1,806	6,875	6,58	alta	Not Selected		
11	92,300	2,194	5,040	4,96	media	Not Selected		
12	98,100	2,841	7,810	7,71	alta	Not Selected		
13	90,700	2,106	5,547	5,09	alta	Not Selected		
14	94,800	1,821	6,659	6,58	alta	Not Selected		
15	92,600	1,757	2,973	2,97	baja	Not Selected		
16	93,800	1,717	4,317	4,21	media	Selected		
17	98,100	2,722	8,217	8,11	alta	Selected		
18	90,500	1,880	5,946	6,25	alta	Selected		

Figura 2: Vista parcial de la base de datos EMBALSES.sav (en SPSS)

1.2.-Entorno del SPSS: Entrada de dato y operaciones básicas: En este apartado se explican con detalle las características generales del programa SPSS 23, y actuaciones básicas que los estudiantes deben saber realizar para poder llevar a cabo los análisis estadísticos de los diferentes bloques temáticos que se desarrollan en la guía de aprendizaje. Se presentan capturas de pantalla de los pasos que hay que seguir con el programa.

1.3.-Bloques temáticos: Se han elaborado 5 bloques temáticos correspondientes a los bloques más importantes de la asignatura.

BLOQUE 1: Estadística Descriptiva.

BLOQUE 2: Regresión y Correlación Lineal

BLOQUE 3: Introducción a la Inferencia Estadística: Intervalos de Confianza y Contrastes de Hipótesis .

BLOQUE 4: Análisis de la Varianza.

BLOQUE 5: Análisis de Tablas de Contingencia.

El material para cada bloque temático se ha dividido en dos partes, una primera parte en la que se plantean uno o varios objetivos a alcanzar –con preguntas dirigidas para guiar al estudiante hacia la correcta elección de la técnica de análisis que tienen que utilizar, focalizando para que el estudiante consiga aplicarla correctamente, interpretar los resultados de forma apropiada y precisa. En la figura 3, aparecen por ejemplo algunos de los objetivos y preguntas planteados en el bloque de Regresión y Correlación.

Práctica 3 REGRESIÓN

OBJETIVOS y PREGUNTAS

Se piensa que existe una relación lineal entre la temperatura y los Silicatos, de forma que la Temperatura depende de los Silicatos del agua. Ajuste un Modelo de regresión lineal simple sólo para los datos correspondientes al embalse de LEÓN siguiendo los pasos que se piden a continuación:

1. Represente el diagrama de dispersión correspondiente, y comente cómo podría ser el tipo de relación entre ambas variables.
2. ¿Cuál es la magnitud de la relación. ¿Por qué? ¿Qué coeficiente utiliza para responder?
 1. Ajuste un modelo de regresión lineal. Escriba el modelo.
 2. Interpretación experimental de los coeficientes del modelo.
3. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.
4. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.
 - Calcule el valor pronosticado de para la Temperatura si el valor de los Silicatos es de 2.5. Comente si la predicción es adecuada y por qué. (Realice el cálculo con el ordenador utilizando 5 decimales). **No se olvide de poner, además, en el archivo de datos que ese dato que pretende predecir pertenece al embalse de LEÓN en la columna embalse.**
 - Realice ahora el ajuste también para el embalse de CÁCERES. ¿En qué embalse el ajuste lineal es mejor?. ¿Por qué?.

Figura 3: Vista parcial de la Guía de Aprendizaje.

Esta parte de cada bloque temático, cuando este material se implemente en STUDIUM el próximo curso académico, se pondrá a disposición del estudiante antes de la clase presencial correspondiente de prácticas.

Una vez que se haya impartido la clase presencial de prácticas, se pondrá a disposición del estudiante la parte del bloque en el que se da solución a dichas preguntas, en el que se presentan las capturas de pantalla con las opciones que se tienen que elegir con el programa para aplicar la técnica de análisis correcta, así como las capturas de los resultados obtenidos y la explicación de la interpretación de dichos resultados tanto desde el punto de vista estadístico como desde el punto de vista práctico.

En la figura 4 se muestran parte de los resultados correspondientes a la figura 3.

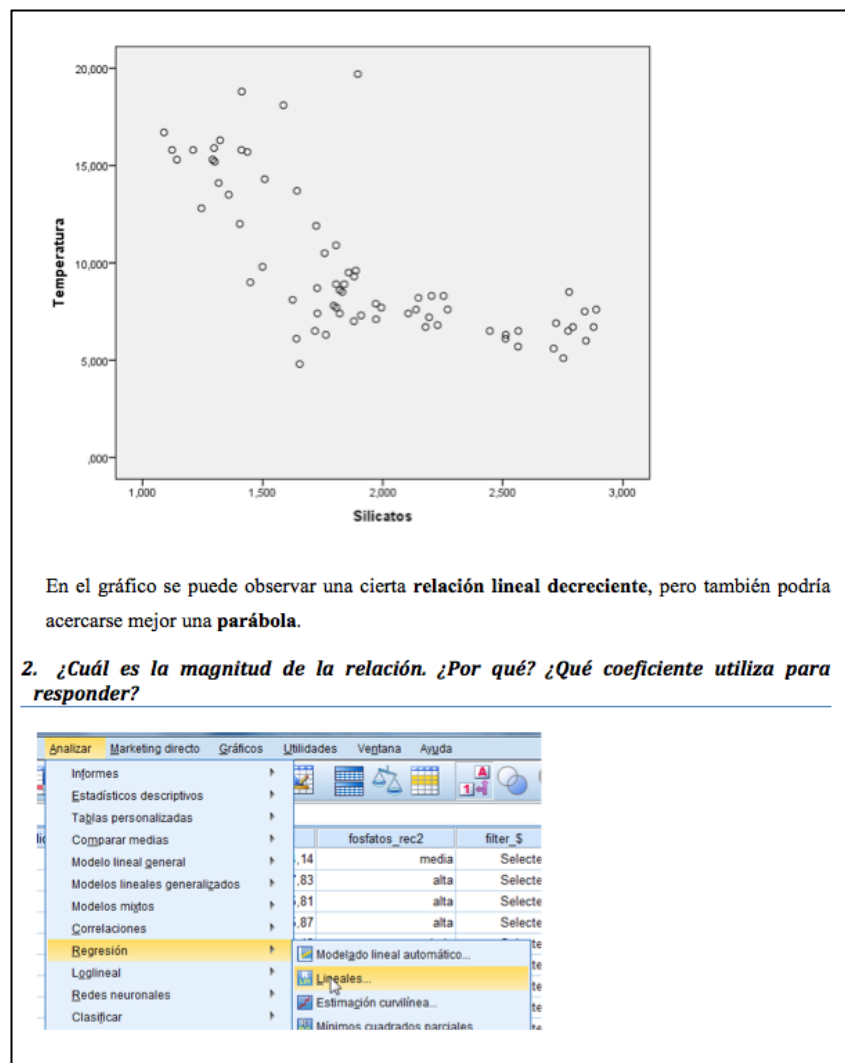


Figura 4: Vista parcial de la Guía de Aprendizaje.

2.- AMPLIACIÓN DE OBJETIVOS Y PREGUNTAS: TRABAJO AUTÓNOMO DEL ESTUDIANTE. El documento se estructura de la siguiente forma:

2.1.- Introducción: De igual forma que en la Guía de aprendizaje, se presenta un problema real y una base de datos que los estudiantes deberán utilizar de forma autónoma para tratar de asentar sus conocimientos y poner a prueba las competencias adquiridas. En esta base de datos se utilizarán las mismas técnicas de análisis estadístico que se han aprendido en el curso y se han utilizado en la Guía de aprendizaje. Ver figura 5


INTRODUCCIÓN

Imagine que usted trabaja en una compañía alimentaria donde se pretende producir una marca especial de cereales. El gerente de la empresa le manda realizar un estudio los cereales que se venden en el mercado para poder conocer mejor este mercado y fabricar un tipo de cereales que pueda competir con los que ya llevan mucho tiempo en venta al público.

Para ello, le proporcionan un archivo donde se han elegido al azar 76 marcas de cereales donde se estudian características relacionadas con la composición nutritiva, en términos de oligoelementos, proteínas y grasas. También se tiene en cuenta las calorías, si están enriquecidos con vitaminas (Bajo=0, Medio=1, Alto=2). Por último se ha considerado el precio de los cereales de forma cualitativa (0=Baratos, 1=Caros) y si se toman fríos (C) con leche en los desayunos o calientes (H) como ingredientes de comidas.

En el archivo "Cereals.sav", aparece la información de los cereales con las siguientes variables

***Nombre** de la marca del cereal, **Empresa** que fabrica la marca de cereal, **Mfr:** la primera inicial de cada empresa, **Consumo:** si se toman en frío (C) o en caliente (H), **Calorías**, **Proteínas:** número de proteínas que contienen, **Grasa (gr)**, **Sodio (mgr)**, **Fibra(gr):** fibra(gr) **Carbohidratos complejos(gr)**, **Azúcar (gr)** **Caloríasgrasa**, **Potasio (mgr)** **Enriquecidos:** si están enriquecidos con vitaminas, **Fibrarec:** fibra recodificada en baja, media, alta, **Precio.** Además se han **considerado** dos variables nuevas donde se pretende comprobar si los valores que dicen que son en términos medios correctos o no. Para ello la propia empresa realiza sus mediciones del sodio (**ValoraciónSodio**) y del potasio (**ValoraciónPotasio**).*



INFORMACIÓN NUTRICIONAL Valores medios aproximados				
	Por 100g	Por porción (1.13g)	%CDO por porción	%CDO por 100g
Valor energético	439 kcal	55 kcal	3%	12%
	1848 kJ	231 kJ		
Proteínas	4,5 g	1,1 g	2%	8%
Hidratos de carbono de los cuales azúcares	65 g	8,1 g	3%	12%
	29 g	3,6 g	4%	16%
Grasas	16 g	2 g	3%	12%
de las cuales saturadas	4,3 g	0,5 g	3%	12%
Fibra alimentaria	6,8 g	0,9 g	3%	12%
Sodio	0,26 g	0,03 g	1%	4%
Vitaminas y Minerales				
	Por 100g	%CDB	Por porción (1.13g)	%CDO
Vitamina E	4,70 mg	39%	0,59 mg	5%
Vitamina B1	0,507 mg	46%	0,06 mg	5%
Vitamina B3	4,02 mg	25%	0,50 mg	3%
Magnesio	128 mg	34%	15,75 mg	4%
Hierro	7,70 mg	55%	0,86 mg	7%

*CDO - Cantidad Diaria Orientativa, que representa la cantidad aproximada de calorías y nutrientes para un adulto medio, calculada sobre la base de una dieta de 2.000 kcal.
 **CDB - Cantidad diaria recomendada.

Para realizar el informe conteste a los siguientes bloques de prácticas:

Figura 5: Descripción de la base de datos (información parcial de una página) del documento de trabajo autónomo del estudiante

2.2.-*Bloques temáticos*: Se sigue de forma análoga a la Guía de aprendizaje el esquema de los 5 bloques temáticos, de forma que ahora sólo se fijan objetivos y se plantean preguntas para que el estudiante valore las competencias adquiridas. (ver figura 6)

PRÁCTICA 4 – ANOVA.

1. Ahora se pretende conocer si hay diferencias entre tres de enriquecimiento de los cereales (Bajo, Medio, Alto), con relación a la variable calorías grasa. Para ello rellene los siguientes apartados.

- H_0
- H_a
- Niveles de significación
- Defina el Estadígrafo de contraste y su distribución.
- Valor experimental del estadígrafo de contraste para la muestra analizada. Escriba la tabla típica de este análisis.
- *p*-valor
- Conclusión estadística
- Conclusión experimental.

2. Si ha encontrado diferencias en el apartado anterior. ¿A que son debidas esas diferencias?. Plantee el contraste de forma general

- H_0
- H_a
- Niveles de significación
- Defina el Estadígrafo de contraste y su distribución.
- Valores experimentales para los distintos contrastes y su *p*-valores.
- Conclusión estadística
- Conclusión experimental.

3. ¿Qué contraste ha utilizado y por qué?

4. ¿Qué penalización ha utilizado y por qué.

5. Se han cumplido Los supuestos de partida para hacer este análisis. En que se basa para responder a esta pregunta. ¿Qué test ha utilizado?. ¿Cuáles han sido los resultados de las pruebas?

Figura 6: Vista parcial del documento para trabajo autónomo del estudiante.

3.4.-Implementación en STVDIVM y metodología de enseñanza.

Tal como apuntábamos en la memoria de solicitud del proyecto, el próximo curso académico se implementará en la plataforma de docencia virtual de la USAL, todo el material elaborado.

Como se dijo, antes de cada clase práctica presencial de cada bloque temático, cada estudiante tendrá a su disposición en la plataforma el material correspondiente a las bases de datos y a las preguntas y objetivos planteados en la Guía de aprendizaje y en la de Trabajo autónomo.

Con la explicación del profesor se irán resolviendo en clase las preguntas planteadas en la guía de aprendizaje, de modo interactivo con los estudiantes, de forma que el profesor sea el vehículo para fomentar el razonamiento de los estudiantes y el apoyo para la toma de decisiones al elegir las técnicas estadísticas de análisis, fomentando la crítica y la correcta interpretación de los resultados.

Una vez visto presencialmente, el material con la solución y las capturas de pantalla se pondrá a disposición de los estudiantes para que puedan estudiar y/o repasar el tema, de forma que puedan asentar más sus conocimientos.

Entonces será e momento en el que el estudiante deberá –de forma individual, y autónoma- tratar de dar solución a las correspondientes preguntas planteadas para cubrir los objetivos establecidos en la guía de trabajo autónomo. Por supuesto que siempre podrán consultar al profesor las dudas que les surjan.

Una vez finalizada la docencia, y antes de las pruebas de evaluación ordinaria, se pondrá a disposición del alumno todo el material unificado en un solo documento.

4.- RESULTADOS OBTENIDOS

Se ha conseguido un trabajo colaborativo entre todos los profesores que imparten docencia en las asignaturas de Estadística de la Facultad de Químicas de la USAL, lo que ha aumentado la motivación a la hora de elaborar un material didáctico novedoso que repercutirá favorablemente en el proceso de enseñanza-aprendizaje.

Se ha optimizado y mejorado la fase de preparación del material docente de prácticas para el próximo curso académico, lo que esperamos que influya de forma favorable en los estudiantes matriculados en dichas asignaturas, puesto que tendrán a su disposición un material concienzudamente elaborado, en el que se ha implicado un equipo de profesores trabajando bajo una misma perspectiva, y que ha dado como resultado un material didáctico que les va a facilitar el aprendizaje de la puesta en práctica de las técnicas estadísticas aprendidas a lo largo del curso, que aumentará su motivación y mejorará su aprendizaje autónomo afianzando destrezas y competencias.

Anexo I

GUÍA DE APRENDIZAJE DE LAS PRÁCTICAS CON SPSS DE LAS ASIGNATURAS DE ESTADÍSTICA DE LA FACULTAD DE QUÍMICAS (I)

DEPARTAMENTO DE ESTADÍSTICA

UNIVERSIDAD DE SALAMANCA

PROYECTO DE INNOVACIÓN DOCENTE: **ID2015/0063**

Coordinadora:

M^a José Fernández Gómez

Miembros del Equipo:

Inmaculada Barrera Mellado

Javier Martín Vallejo

Santiago Vicente Tavera

Rosa Amanda Sepúlveda Correa

Jesús Martín Rodríguez

INTRODUCCIÓN

...“Se define calidad del agua como la adecuación a un determinado uso (por ejemplo, uso para abastecimiento urbano, regadío, aptitud para la vida de los peces, etc.). Para establecer el grado de adecuación que presenta el agua en relación con el uso al que se va a destinar han de medirse una serie de atributos que, de forma objetiva, verifiquen dicha adecuación. Los atributos a medir son parámetros físicoquímicos y biológicos, fundamentalmente...”

...”La Directiva Marco del Agua (Directiva 2000/60/CE, conocida como DMA), así como la normativa estatal mediante la cual se transpone dicha norma comunitaria, establece la obligatoriedad de contar con Programas de seguimiento del estado de las masas de agua, tanto superficiales como subterráneas (Tablas 5.13 y 5.14).

Por otro lado, la DMA establece la necesidad de alcanzar, como mínimo, un “buen estado” de las aguas superficiales en toda la cuenca hidrográfica del Duero para el año 2015. En el caso de las masas de agua artificiales (por ejemplo el Canal de Castilla) o muy modificadas (como por ejemplo los embalses), la Directiva establece como objetivo la consecución de un buen potencial ecológico y un buen estado químico.

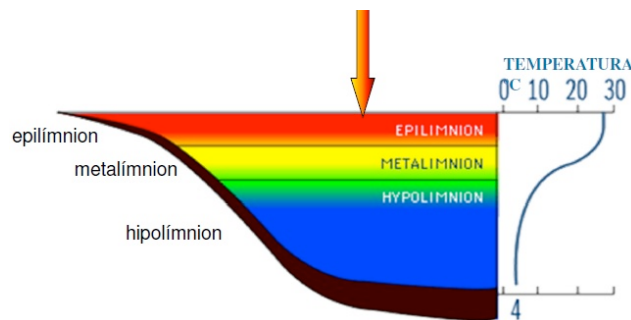
Para las aguas superficiales, el estado viene definido por la suma del estado ecológico y del estado químico. Para la determinación del estado ecológico entran en juego numerosos indicadores de diferente naturaleza (indicadores biológicos, físicoquímicos e hidromorfológicos). Para la determinación del estado químico se deben tener en cuenta únicamente unas determinadas sustancias químicas denominadas sustancias prioritarias, que son aquellas sustancias que presentan un riesgo significativo para el medio acuático o a través del mismo (ver el recientemente aprobado Real Decreto 60/2011).”... (Tomado de la Memoria 2009-2010. Confederación Hidrográfica del Duero (CDH))

Las propiedades de solubilidad de sales y gases en agua y la mayor parte de las propiedades químicas del agua se ven afectadas por las variaciones de temperatura. Es tal vez el factor que más influencia tiene en los lagos, pues determina la densidad, viscosidad y movimiento del agua. La temperatura del agua se establece por la absorción de la radiación que incide en las capas más superficiales, estando relacionada con la energía cinética media de sus moléculas.

En un ambiente lacustre o de un lago, sin embargo, la temperatura de la columna de agua sigue una variación cíclica en dos fases: (a) la fase de “mezcla” con temperatura similar en todas las profundidades debida a la circulación e intercambio del agua entre la superficie y el fondo, de forma que no sólo la temperatura sino también los parámetros físico-químicos son más o menos homogéneos a lo largo de toda la columna de agua, y (b) la fase de “estratificación” con aguas más cálidas en las capas más superficiales y más frías en las capas más profundas, de modo que se establece un gradiente de temperatura y de densidad

desde la superficie hasta el fondo, de forma que el resultado final es el establecimiento de una termoclina que se convierte incluso en una barrera física infranqueable por ciertos organismos fitoplanctónicos.

En la fase de estratificación el embalse o lago queda estratificado en tres zonas:



- La capa superficial de mayor temperatura, en contacto con la radiación solar directa. Registra temperaturas elevadas de forma progresiva según avanza la primavera y el verano. (Epilimnion).
- La zona de transición entre la capa superficial y la capa inferior. Constituye el punto de inflexión del perfil de temperaturas. Además, es en esta capa donde se sitúa la Termoclina (la parte donde se produce un descenso brusco de la temperatura). (Metalimnion)
- La capa inferior de baja y homogénea temperatura. (Hipolimnion)

Se producirán así dos gradientes. Uno hace referencia al gradiente de temperatura (máximo en la parte alta y constante en el hipolimnion), y otro gradiente de densidad, que al contrario que el gradiente de temperatura, tendrá más densidad en la parte baja y menos densidad en la parte alta.

En conclusión: la acumulación del calor en un lago/embalse estratificado es una combinación de:

- La radiación solar directa.
- La conducción turbulenta del calor (flujo de calor. / gradiente térmico). El viento en superficie lo provoca.
- Las corrientes de densidad. Hay movimientos en profundidad.

La temperatura es un factor abiótico que controla procesos vitales para los organismos, y también afecta a las propiedades químicas y físicas de otros factores abióticos en un ecosistema. A su vez, un aumento notable de este parámetro puede alargar los períodos de estratificación, potenciando el efecto de los contaminantes aportados por la cuenca. (Armengol y García, 1997).

El conocimiento de las características físico químicas del agua de un embalse es una herramienta fundamental para gestionar adecuadamente este recurso. Dentro de los indicadores que deben tenerse en consideración, la temperatura, el potencial redox (Eh), el oxígeno disuelto (OD) y el pH, resultan de gran interés porque se incluyen dentro de los factores que regulan la liberación de nutrientes, metales y materia

orgánica desde los sedimentos (Bostrom et al., 1988; Aminot & Andrieux, 1996; Appan & Ting, 1996; Harris, 1999).

Los datos de los que dispone hacen referencia al análisis de las aguas de dos embalses, uno en la provincia de León y otro en la provincia de Cáceres.

La toma de muestras se realizó durante 10 semanas, el mismo día cada semana en cada embalse, en una estación representativa del cada embalse, situada en posición central. El muestreo se extendió, desde el 16 de abril al 25 de junio, de modo que las 5 primeras semanas se corresponden con un período de **mezcla** que se debe a que las capas superficiales que se van enfriando se hunden, produciéndose así constantemente la mezcla con el agua profunda, y las últimas 5 semanas se corresponden con una fase de **estratificación**, que en general es debida al calentamiento diferencial de las capas superficiales con respecto a las profundas.

Las muestras se tomaron desde la superficie hacia el fondo a: 0, 2, 5, 10, 15, 20 metros de profundidad y fondo (2 metros por encima de la profundidad máxima de cada embalse).

Las variables físico-químicas analizadas en las muestras fueron las siguientes:

Temperatura	Oxígeno	pH	Conductividad	Silicatos	Fosfatos1	Fosfatos2
--------------------	----------------	-----------	----------------------	------------------	------------------	------------------

La variable fosfato se determinó con dos procedimientos de medición diferentes, de ahí que aparezcan en el archivo las variables Fosfatos1 y Fosfatos2. La cantidad de Fosfatos1 aparece también categorizada de la siguiente forma: **Baja - Media - Alta** atendiendo a si los valores estaban entre [mín-3.5], (3.5-5.5] y (5.5,max.] respectivamente.

En el archivo de datos EMBALSES.sav, además de los datos relativos a todas estas variables cuenta, además con la información en otras 4 columnas en las que aparece la información sobre el embalse (León, Cáceres), la Profundidad, la Semana y la Fase a la que corresponde cada dato. (Deberá chequear en la Vista de variables cómo están definidas las categorías de cada una de ellas para cuando necesite trabajar con dichas variables en los análisis planteados).

Práctica 1

ENTORNO DEL SPSS

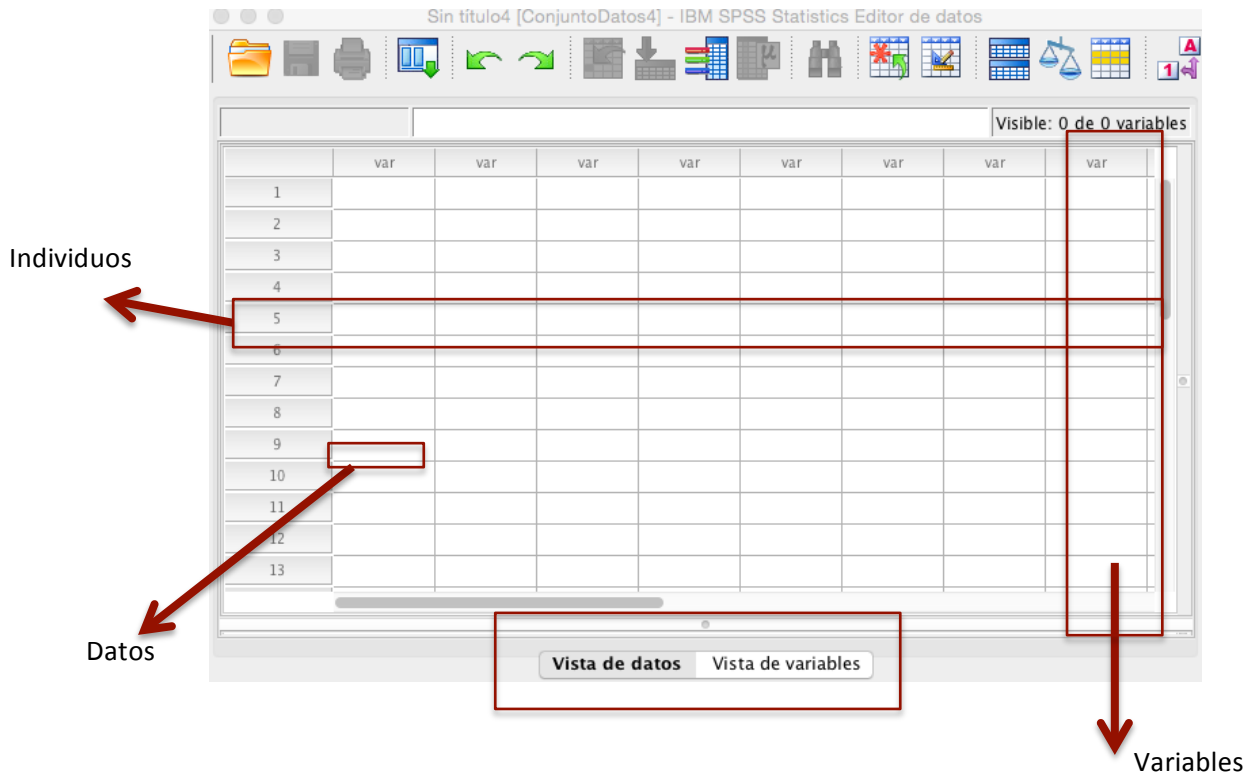
ENTRADA DE DATOS Y OPERACIONES BÁSICAS

1. INSPECCIÓN DEL PROGRAMA	6
2. DEFINIR VARIABLES: PESTAÑA DE VISTA DE VARIABLES.....	7
3. INTRODUCIR DATOS : PESTAÑA DE VISTA DE DATOS	9
4. INSPECCIONANDO EL MENÚ EDICIÓN.....	9
5. INSPECCIONANDO EL MENÚ DATOS.....	10
5.1.- Seleccionar datos	
5.2.- Segmentar archivo	
6. INSPECCIONANDO EL MENÚ TRANSFORMAR.....	12
6.1.- Calcular nuevas variables	
6.2.- Recodificar en distintas variables	

1.- INSPECCIÓN DEL PROGRAMA

Cuando abordamos un trabajo lo primero que hemos de hacer es **introducir los datos en el programa**.

Al abrir por primera vez SPSS nos aparece un pantalla con el aspecto siguiente:



Puede observarse que disponemos de una matriz (tabla) de datos con **n** filas y **p** columnas, vacías.

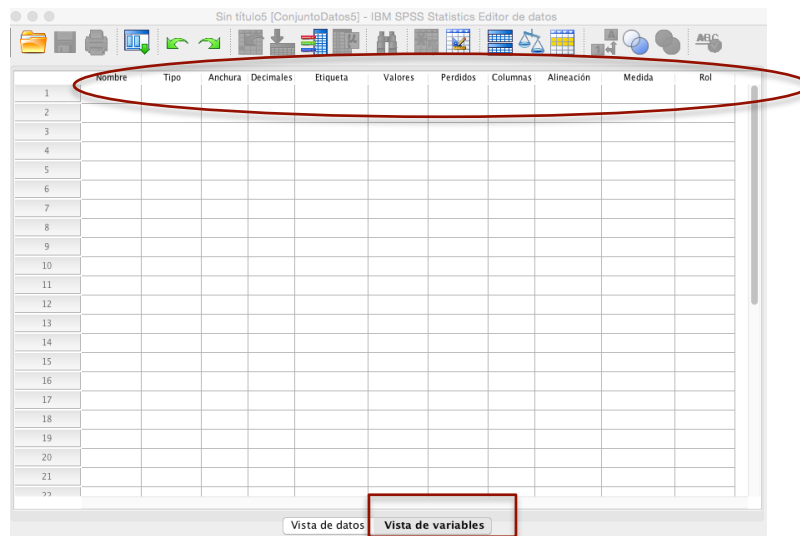
El programa funciona de forma que **las variables** del estudio se localizarán en las **columnas** y los **individuos** en las **filas**.

El interior de ellas aparecen las **celdas** que será el espacio destinado para introducir nuestros **datos**

La pantalla muestra también, en la parte inferior, **dos pestañas (Vista de datos, Vista de variables)**, y observamos que la pestaña que está activa es la de vista de datos.

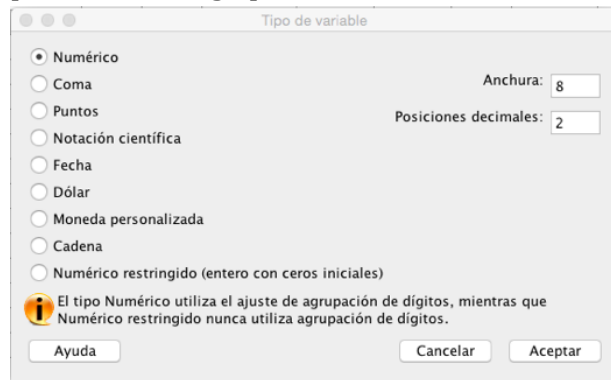
Antes de proceder a introducir los datos en esta ventana tenemos que definir cuidadosamente nuestras variables, para lo que presionaremos sobre la pestaña **vista de variables**.

2.- DEFINIR VARIABLES: PESTAÑA DE VISTA DE VARIABLES

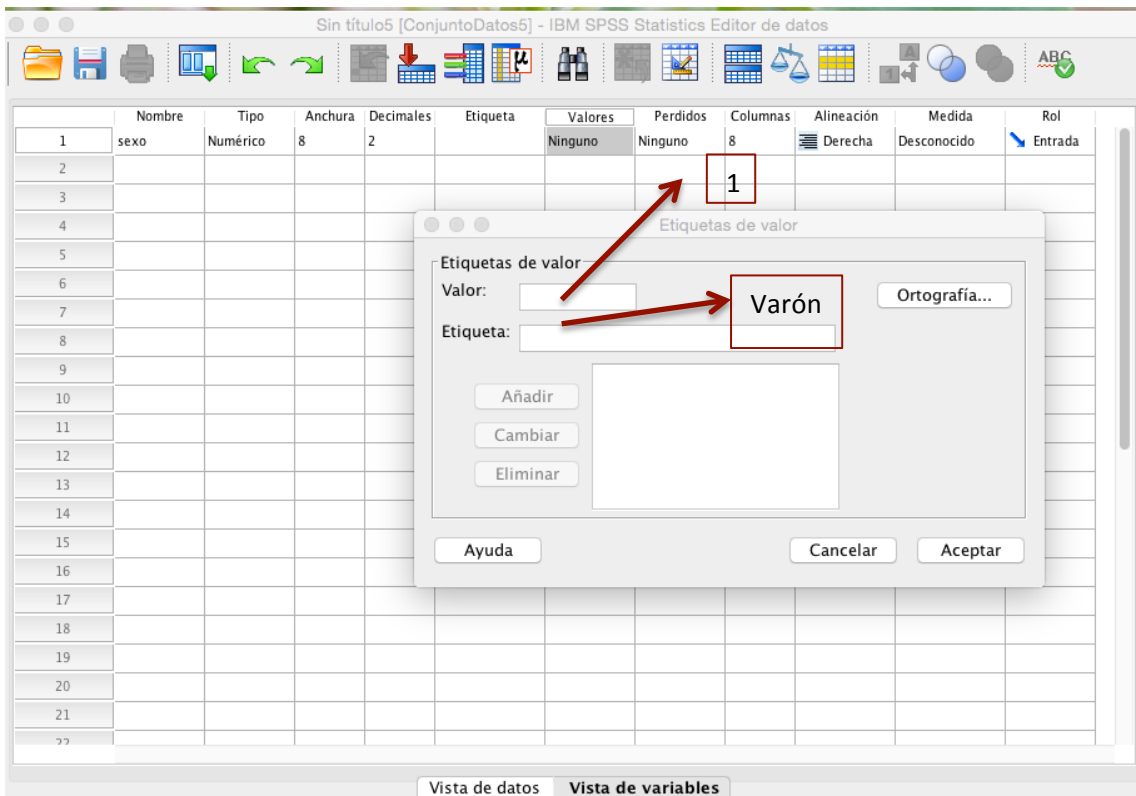


Matriz de definición de características de cada variable:

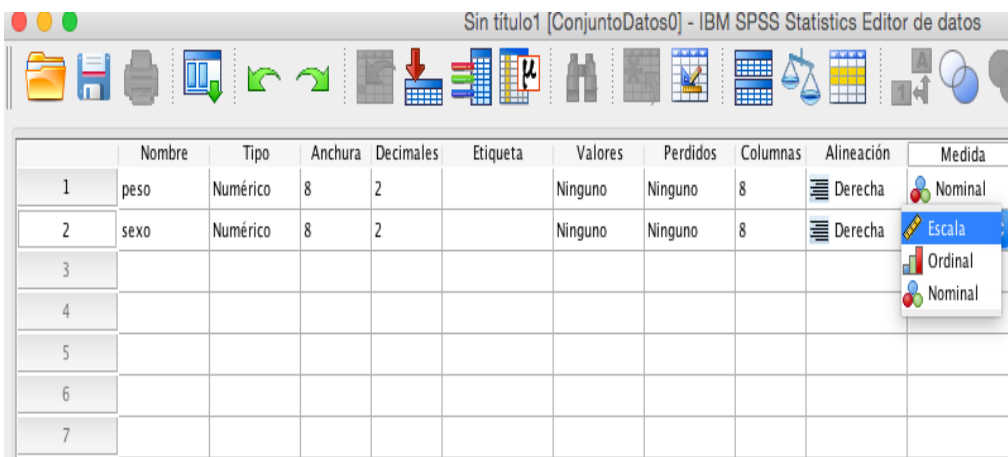
- **En cada fila definiremos las características de cada variable: Nombre, Tipo,**
- **Nombre:** teclearemos el nombre sin caracteres especiales, ni espacios en blanco.
- **Tipo:** en la opción tipo de variable elegimos el tipo **numérica siempre** (salvo cadenas y fechas o monedas) para evitar luego problemas al realizar los análisis.



- Resulta interesante la columna **Etiqueta**. Como cuando se define el nombre de la variable no admite ni caracteres especiales y espacios en blanco, en esta opción nos permite escribir texto normal, por lo que resulta muy útil para que luego en los resultados aparezca correctamente escrito.
- Cuando es de tipo **categórico** la elegimos como numérica pero en la **columna de VALORES** definimos los **valores de las clases**, por ejemplo para la variable sexo definimos los valores: 1= varón y 2= mujer (o al revés).

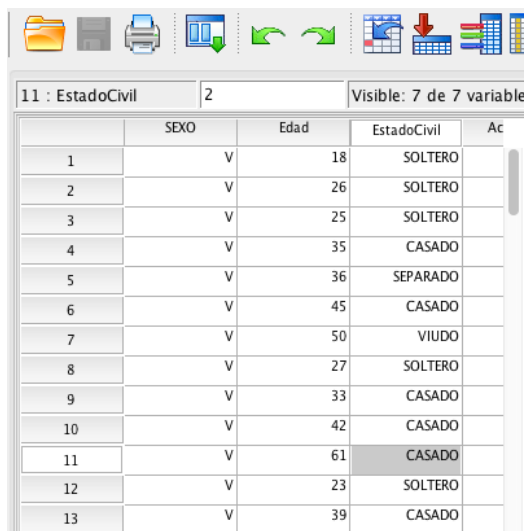


- Una de las columnas de esta matriz (**MEDIDA**) nos permite identificar el tipo de medida que la variable en concreto presenta en nuestro estudio (nominal, ordinal, escala). Cada una viene identificada con un símbolo lo que facilitará a partir de ahora la identificación de la escala de medida de nuestras variables.



3.- INTRODUCIR DATOS: PESTAÑA DE VISTA DE DATOS

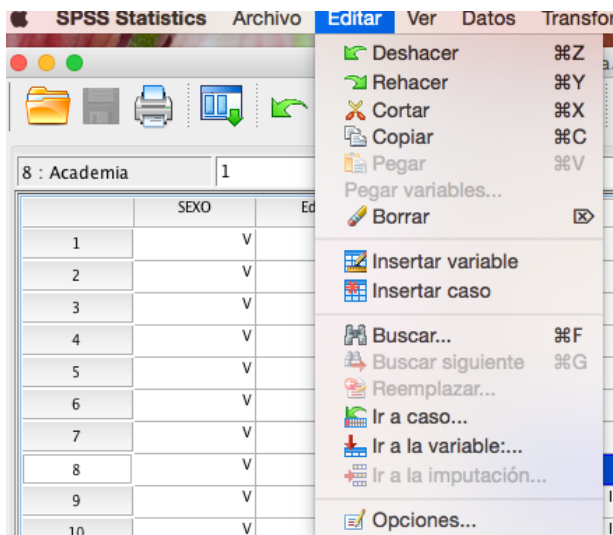
Una vez terminada la definición de nuestras variables estamos en disposición de introducir nuestros datos del estudio, sobre los que queremos trabajar en adelante.



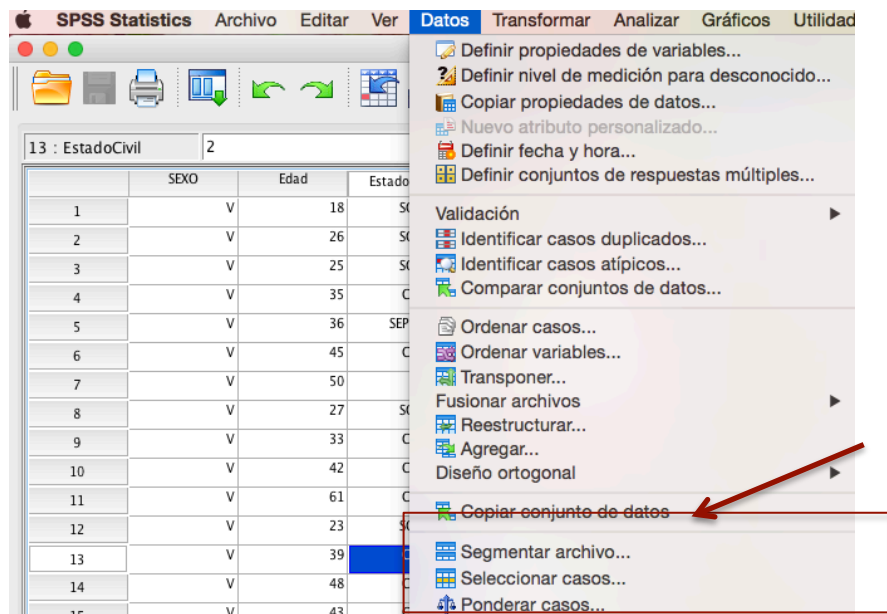
	SEXO	Edad	EstadoCivil	Ac
1	V	18	SOLTERO	
2	V	26	SOLTERO	
3	V	25	SOLTERO	
4	V	35	CASADO	
5	V	36	SEPARADO	
6	V	45	CASADO	
7	V	50	VIUDO	
8	V	27	SOLTERO	
9	V	33	CASADO	
10	V	42	CASADO	
11	V	61	CASADO	
12	V	23	SOLTERO	
13	V	39	CASADO	

4.- INSPECCIONANDO EL MENÚ EDICIÓN

Además de las opciones básicas de cualquier menú edición (cortar, copiar, pegar) el programa nos permite insertar variables o casos en una posición particular, buscar un valor concreto de la tabla o posicionarnos con el cursor en una variable o caso concreto.



5.- INSPECCIONANDO EL MENÚ DATOS

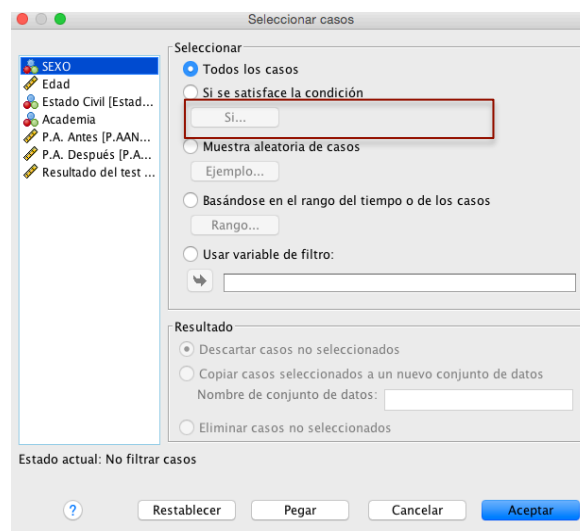


Al final del menú **DATOS** aparecen dos opciones de gran utilidad: **Segmentar archivo** y **Seleccionar datos**

5.1.- Seleccionar datos

En la opción de menú datos se puede seleccionar casos para el estudio posterior cuando no se requieran análisis del fichero completo. Por ejemplo si queremos analizar los resultados del grupo varones y no el de mujeres.

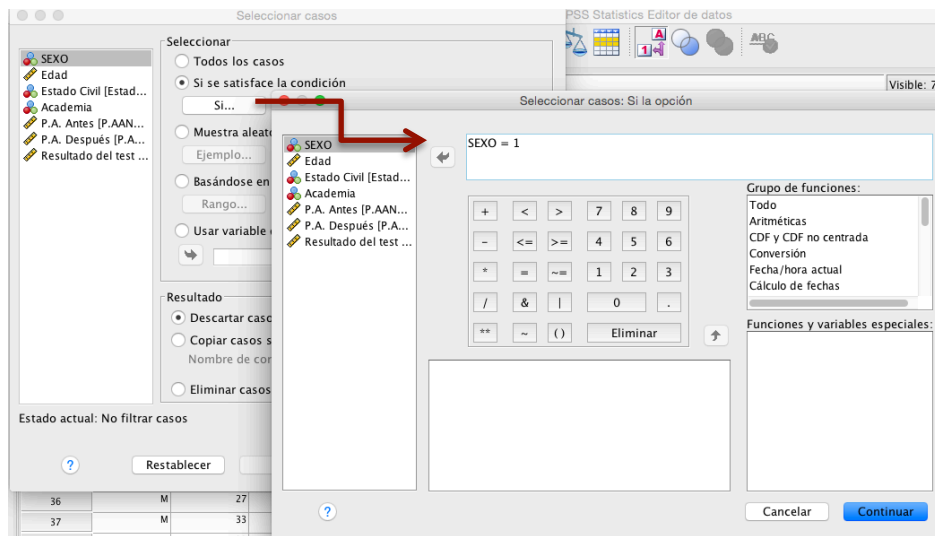
Al marcar esta opción nos aparece otra ventana donde nos pregunta si queremos analizar todos los datos (opción por defecto), sólo los datos que satisfacen una opción, una muestra aleatoria de casos y si queremos usar una variable filtro. Si marcamos esta última nos permite descartar estos datos en el resultado, copiar los datos a un nuevo fichero o eliminar los datos filtrados.



Si elegimos la opción segunda tenemos que definir la opción de selección, supongamos que en nuestro caso queremos que nos seleccione sólo a los **varones (sexo =1)**.

En la última columna de la tabla de datos aparece una nueva variable **filtro** donde aparecen los datos seleccionados y no seleccionados.

En la última opción se nos permite utilizar esta nueva variable filtro creada para generar un nuevo fichero de datos sólo con los chicos si lo deseamos.



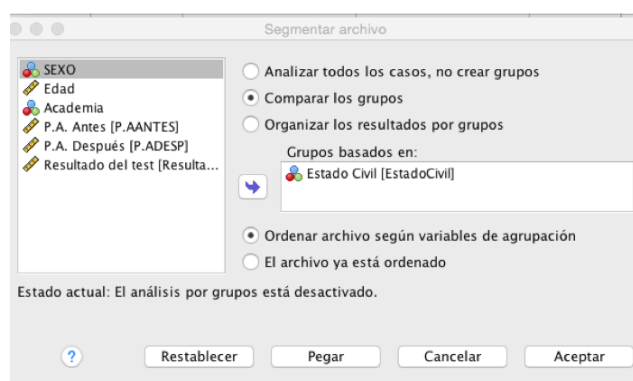
5.2.- Segmentar archivo

Esta opción del menú de datos nos permite hacer análisis de todos los datos o separar los análisis los distintos niveles de una variable categórica.

Hay dos posibles casos: **Comparar grupos** u **organizar los resultados por grupos**.

Supongamos que queremos hacer el estudio descriptivo de la edad pero por separado para distintos grupos del estado civil:

- Si seleccionamos la opción **comparar grupos** nos aparecerá en la misma tabla y/o gráfica la información correspondiente a todos los grupos del estado civil y podemos comparar los resultados.
- La opción **organizar resultados por grupos** nos realiza tablas y/o gráficos separados para cada nivel de la variable categórica.



6.- INSPECCIONANDO EL MENÚ TRANSFORMAR

El menú transformar nos permite **calcular nuevas variables** a partir de las existentes. (Por ejemplo Logaritmos de los valores de una variable ya existente).

Otra opción interesante es la de **recodificar los valores** de una variable para generar otra variable con nuevas clases. Útil para transformar una variable numérica a categórica.

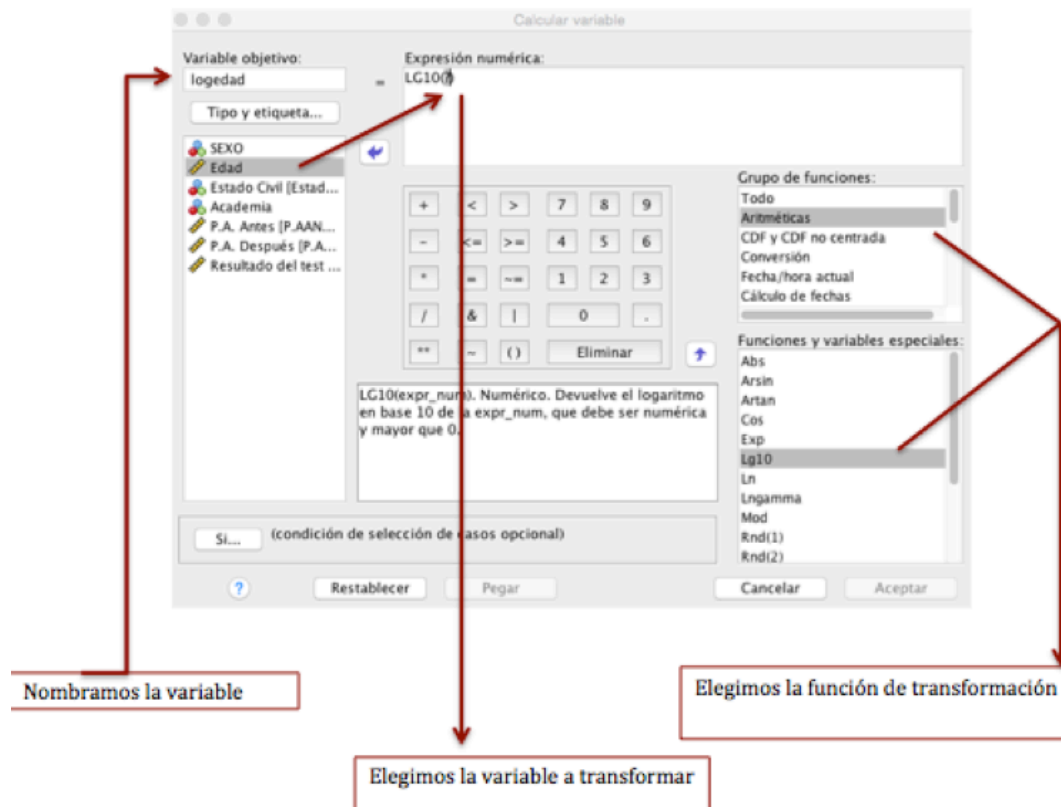


6.1.- Calcular nuevas variables

Deberemos poner un nombre a la variable donde queremos que se escriban los resultados de los cálculos (con las reglas de nombre que dijimos al principio) donde pone **Variable objetivo**. (por ejemplo si queremos calcular el logaritmo en base 10 de la variable edad escribimos: *logedad*).

A continuación seleccionamos la opción correspondiente en el menú Grupo de funciones (en este caso *log10*). Observamos que se escribe LG10(?)

Después seleccionamos la variables con la que queremos realizar los cálculos en el recuadro **Expresión numérica**. (en este caso damos doble click sobre *Edad*, o bien la seleccionamos y la añadimos con la flecha, de modo que el interrogante de la fórmula se sustituye por los valores de la variable edad). Al aceptar, el programa realizará el cálculo de la operación aritmética indicada para todos los individuos.

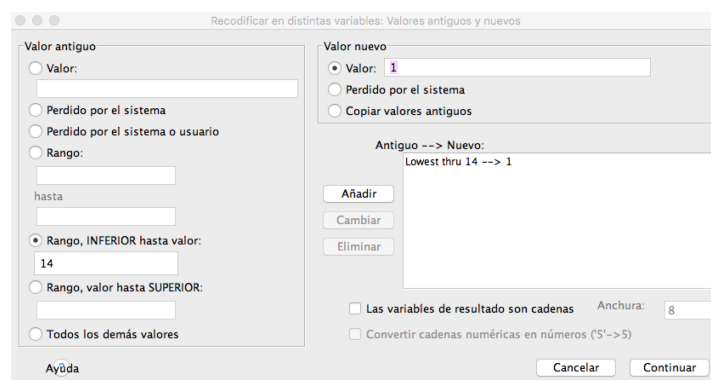


6.2.- Recodificar en distintas variables

Así por ejemplo si tenemos la variable edad como numérica podemos transformarla en una nueva variable categórica con unas nuevas clases establecidas por nosotros por ejemplo: menores de 15 (niños), de 15-25 (jóvenes), mayores de 25 (adultos).

Para ello tenemos que añadir la variable numérica que queremos recodificar, en este caso la Edad. Le damos un **nombre** y si queremos le ponemos una *etiqueta*. Le damos al botón **CAMBIAR** y luego al botón: **VALORES ANTIGUOS Y NUEVOS** donde definiremos el cambio de valores a categorías.

- Así todos los valores de edad hasta los 14 años, van a ser los niños, por tanto escribimos y seleccionamos tal como aparece en la captura de pantalla:



- Los valores de edad de entre los 15-25 años, van a ser los jóvenes, por tanto escribimos y seleccionamos tal como aparece en la captura de pantalla.

Recodificar en distintas variables: Valores antiguos y nuevos

Valor antiguo

Valor:

Perdido por el sistema

Perdido por el sistema o usuario

Rango:

hasta

Rango, INFERIOR hasta valor:

Rango, valor hasta SUPERIOR:

Todos los demás valores

Ayuda

Valor nuevo

Valor:

Perdido por el sistema

Copiar valores antiguos

Antiguo --> Nuevo:

Lowest thru 14 --> 1

15 thru 25 --> 2

Añadir

Cambiar

Eliminar

Las variables de resultado son cadenas Anchura: 8

Convertir cadenas numéricas en números (*5-->5)

Cancelar Continuar

- Por último los valores de edad mayores de 25 años, van a ser los adultos, por tanto escribimos y seleccionamos tal como aparece en la captura de pantalla.

Recodificar en distintas variables: Valores antiguos y nuevos

Valor antiguo

Valor:

Perdido por el sistema

Perdido por el sistema o usuario

Rango:

hasta

Rango, INFERIOR hasta valor:

Rango, valor hasta SUPERIOR:

26

Todos los demás valores

Ayuda

Valor nuevo

Valor: 3

Perdido por el sistema

Copiar valores antiguos

Antiguo --> Nuevo:

Lowest thru 14 --> 1

15 thru 25 --> 2

26 thru Highest --> 3

Añadir

Cambiar

Eliminar

Las variables de resultado son cadenas Anchura: 8

Convertir cadenas numéricas en números (*5-->5)

Cancelar Continuar

Práctica 2

DESCRIPTIVA e INTERVALOS DE CONFIANZA

OBJETIVOS y PREGUNTAS

1. Explique la estructura de la tabla de datos.
2. Se desea comparar los resultados para las variables silicatos y pH entre la fase de Mezcla y de Estratificación. Rellene la siguiente tabla:

Variable	FASE	Media	Desv. Típica	Error Estándar	Mediana	Rango Intercuartílico
SIO	MEZCLA					
	ESTRATIFICACIÓN					
pH	MEZCLA					
	ESTRATIFICACIÓN					

En base a la información presentada en la tabla anterior, y a la de otros análisis que considere oportunos responda a las siguientes preguntas:

3. ¿En qué FASE hay mayor contenido en silicatos? Justifique su respuesta.
4. ¿En qué FASE hay mayor variabilidad de pH? Justifique la respuesta e indique qué coeficiente ha utilizado para responder a la pregunta y por qué lo ha elegido.
5. ¿Cuál de las dos variables presenta más variabilidad en la fase de MEZCLA? Justifique la respuesta e indique qué coeficiente ha utilizado para responder a la pregunta y por qué lo ha elegido.
6. Interprete el significado de las medianas de ambas variables en la fase de MEZCLA.
7. Calcule los siguientes percentiles en la fase de MEZCLA interprete después el P90 y el P15 para cada variable.

Variable	Percentiles				
	P ₁₅	P ₂₅	P ₄₀	P ₇₅	P ₉₀
SIO					
pH					

Interpretación para ambas variables de los P₁₅ y del P₉₀

8. Realice la representación gráfica que considere más adecuada para analizar si la distribución del pH es aproximadamente simétrica en ambas FASES. Justifique la respuesta. Comente los resultados gráficos.

9. Realice la representación gráfica que considere más adecuada para analizar si la distribución de la Conductividad y de los Fosfatos es aproximadamente simétrica en la Fase de ESTRATIFICACIÓN. Justifique la respuesta. Comente los resultados gráficos.
10. Escriba el valor de la media muestral y el intervalo de confianza al 95% para la media poblacional de la variable Oxígeno disuelto en la Fase de ESTRATIFICACIÓN (suponga normalidad). INTERPRETE LOS RESULTADOS.

Práctica 2 DESCRIPTIVA

SOLUCIÓN

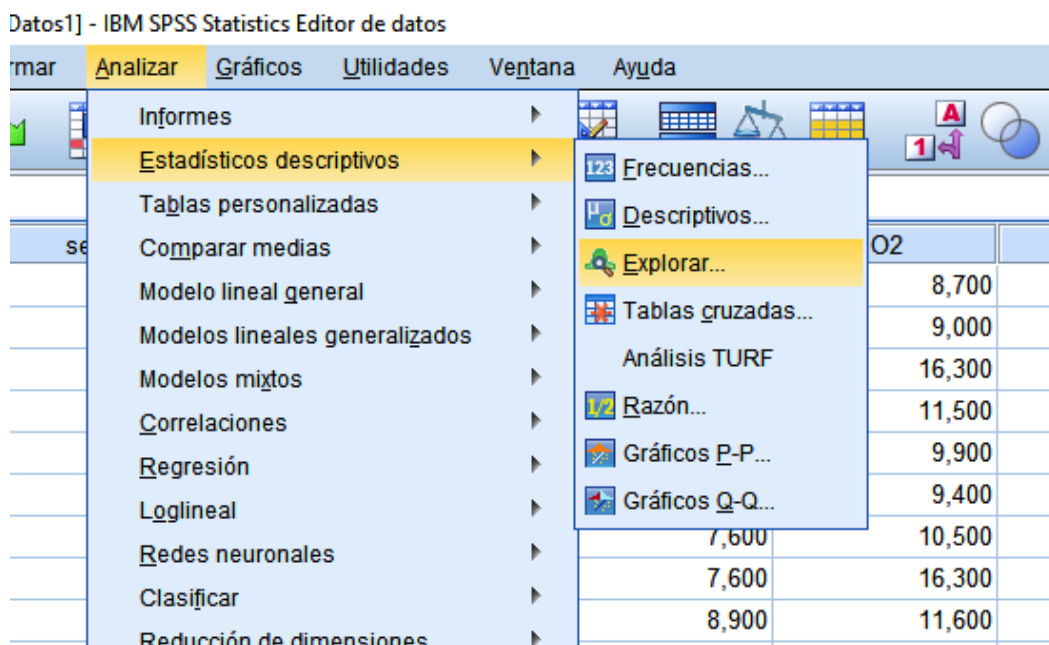
1. Explique la estructura de la tabla de datos.

En nuestra base de datos tenemos las siguientes variables:

- **EMBALSE:** variable nominal (cualitativa) que toma dos valores, 1 = LEÓN, 2 = CÁCERES.
- **Profundidad:** variable ordinal (cualitativa), 1 = 0m, 2 = 2m, 3 = 5m, 4 = 10m, 5 = 15m, 6 = 20m y 7 = fondo.
- **Semana:** variable ordinal (cualitativa).
- **FASE:** variable nominal (cualitativa), 1 = MEZCLA y 2 = ESTRATIFICACIÓN.
- **T:** Temperatura, variable cuantitativa continua.
- **O2:** Oxígeno disuelto, variable cuantitativa continua.
- **pH:** variable cuantitativa continua.
- **Conductividad:** variable cuantitativa continua.
- **silicatos:** variable cuantitativa continua.
- **fosfatos:** variable cuantitativa continua.
- **fosfatos2:** variable cuantitativa continua.
- **Fosfatos_rec2:** variable ordinal (cualitativa), 1 = baja, 2 = media y 3 = alta.

2. Se desea comparar los resultados para las variables silicatos y pH entre la fase de Mezcla y de Estratificación. Rellene la siguiente tabla:

En el menú nos vamos a Analizar -> Estadísticos descriptivos -> Explorar



Datos1] - IBM SPSS Statistics Editor de datos

mar Analizar Gráficos Utilidades Ventana Ayuda

Informes

Estadísticos descriptivos

Tablas personalizadas

Comparar medias

Modelo lineal general

Modelos lineales generalizados

Modelos mixtos

Correlaciones

Regresión

Loglineal

Redes neuronales

Clasificar

Reducción de dimensiones

123 Frecuencias...

Descriptivos...

Explorar...

Tablas cruzadas...

Análisis TURF

Razón...

Gráficos P-P...

Gráficos Q-Q...

7,600

7,600

8,900

O2

8,700

9,000

16,300

11,500

9,900

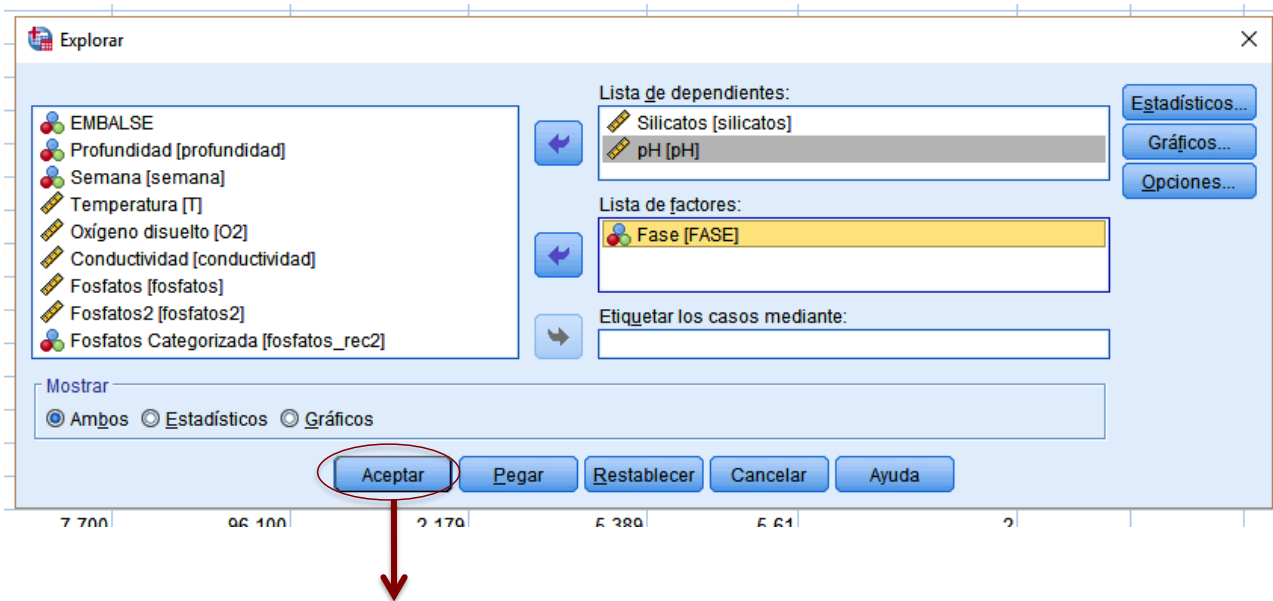
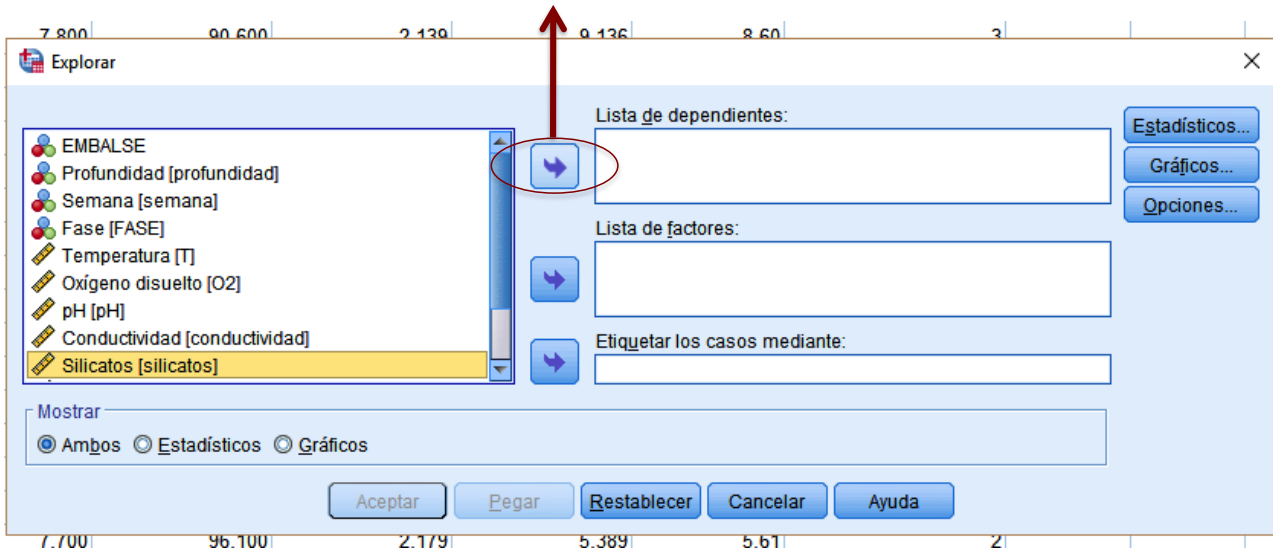
9,400

10,500

16,300

11,600

Con este botón
seleccionamos las
variables para el
análisis



No olvidar
presionar aceptar
para llevar a cabo
el análisis

De esta forma obtenemos los estadísticos descriptivos para cada **Fase por separado**, y con esta información completamos la tabla (aproximando nuestros resultados a 3 cifras decimales).

Descriptivos

Fase			Estadístico	Error estándar	
Silicatos	MEZCLA	Media	1,98594	,055319	
		95% de intervalo de confianza para la media	Límite inferior	1,87558	
			Límite superior	2,09630	
		Media recortada al 5%	1,97962		
		Mediana	1,86114		
		Varianza	,214		
		Desviación estándar	,462830		
		Mínimo	1,110		
		Máximo	2,889		
		Rango	1,779		
	Rango intercuartil	,664			
	Asimetría	,364	,287		
	Curtosis	-,840	,566		
	ESTRATIFICACION	Media	1,53302	,064306	
		95% de intervalo de confianza para la media	Límite inferior	1,40473	
			Límite superior	1,66131	
		Media recortada al 5%	1,52400		
		Mediana	1,44200		
		Varianza	,289		
		Desviación estándar	,538022		
Mínimo		,525			
Máximo		2,878			
Rango		2,353			
Rango intercuartil	,652				
Asimetría	,448	,287			
Curtosis	-,159	,566			
pH	MEZCLA	Media	7,69307	,032330	
		95% de intervalo de confianza para la media	Límite inferior	7,62857	
			Límite superior	7,75757	
		Media recortada al 5%	7,69183		
		Mediana	7,70000		
		Varianza	,073		
		Desviación estándar	,270493		
		Mínimo	7,100		
		Máximo	8,266		
		Rango	1,166		
	Rango intercuartil	,416			
	Asimetría	,053	,287		
	Curtosis	-,708	,566		
	ESTRATIFICACION	Media	7,96672	,028884	
		95% de intervalo de confianza para la media	Límite inferior	7,90910	
			Límite superior	8,02434	
		Media recortada al 5%	7,97426		
		Mediana	8,00000		
		Varianza	,058		
		Desviación estándar	,241659		
Mínimo		7,300			
Máximo		8,425			
Rango		1,125			
Rango intercuartil	,323				
Asimetría	-,511	,287			
Curtosis	,030	,566			

Variable	FASE	Media	Desv. Típica	Error Estándar	Mediana	Rango Intercuartílico
SIO	MEZCLA	1,986	0,463	0,055	1,861	0,664
	ESTRATIFICACIÓN	1,533	0,538	0,064	1,442	0,652
pH	MEZCLA	7,693	0,270	0,032	7,700	0,416
	ESTRATIFICACIÓN	7,967	0,242	0,029	8,000	0,323

En base a la información presentada en la tabla anterior, y a la de otros análisis que considere oportunos responda a las siguientes preguntas:

3. ¿En qué FASE hay mayor contenido en silicatos? Justifique su respuesta.

Consideramos las medias muestrales para la variable "Silicatos" en las distintas fases: 1,986 y 1,533, para las fases de Mezcla y Estratificación, respectivamente. Por lo tanto, hay mayor contenido en silicatos en la fase de Mezcla.

4. ¿En qué FASE hay mayor variabilidad de pH? Justifique la respuesta e indique qué coeficiente ha utilizado para responder a la pregunta y por qué lo ha elegido.

Consideraremos las desviaciones típicas para comparar la variabilidad ya que tenemos las mismas unidades de medida en ambos grupos (las unidades de medida del pH). Para la fase de mezcla tenemos una desviación típica de 0,270 y para la fase de Estratificación de 0,242. Por lo tanto, la variable pH tiene mayor variabilidad en la fase de Mezcla.

5. ¿Cuál de las dos variables presenta más variabilidad en la fase de MEZCLA? Justifique la respuesta e indique qué coeficiente ha utilizado para responder a la pregunta y por qué lo ha elegido.

En este caso tenemos variables con distintas unidades de medida (Silicatos y pH) por lo que tenemos que considerar el coeficiente de variación para comparar la variabilidad de ambas.

Para los silicatos en la fase de Mezcla tenemos:

$$CV = \frac{0,463}{1,986} = 0,233 = 23,3\%$$

Para el pH en la fase de Mezcla tenemos:

$$CV = \frac{0,270}{7,693} = 0,035 = 3,5\%$$

Por lo tanto, en la fase de Mezcla presenta mayor variabilidad la variable silicatos.

6. Interprete el significado de las medianas de ambas variables en la fase de MEZCLA.

Para los silicatos en la fase de Mezcla tenemos:

$$Me = 1,861$$

En la fase de Mezcla, el 50% de las observaciones tiene un valor de silicatos inferior o igual a 1,861.

Para el pH en la fase de Mezcla tenemos:

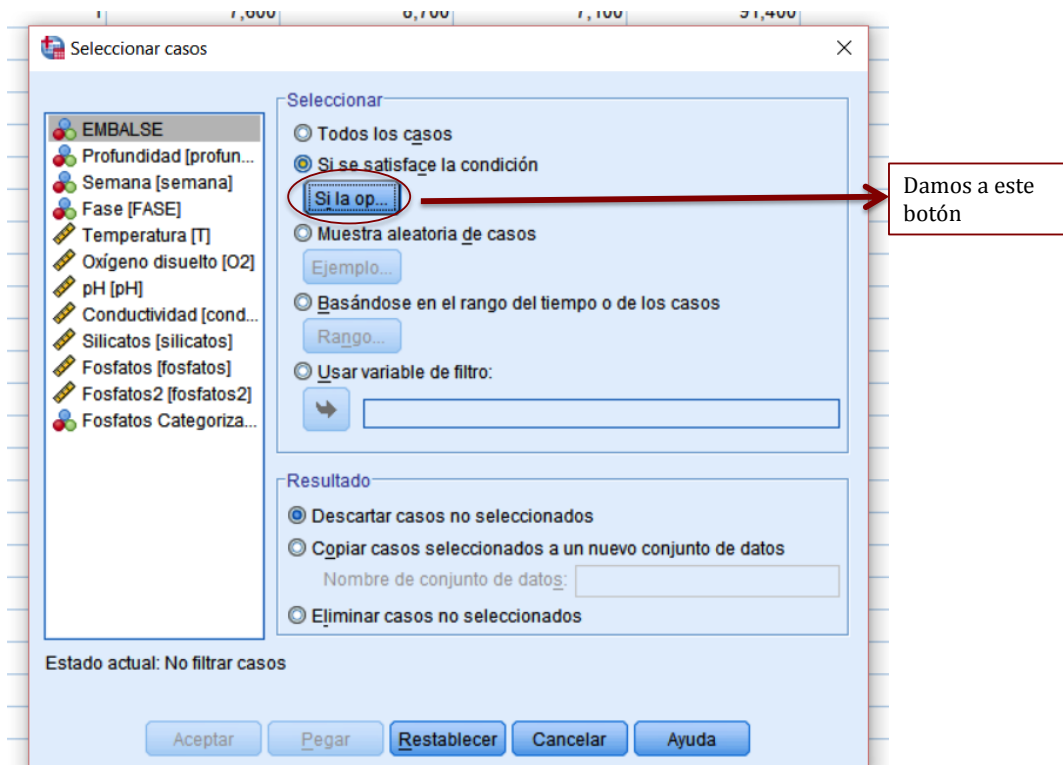
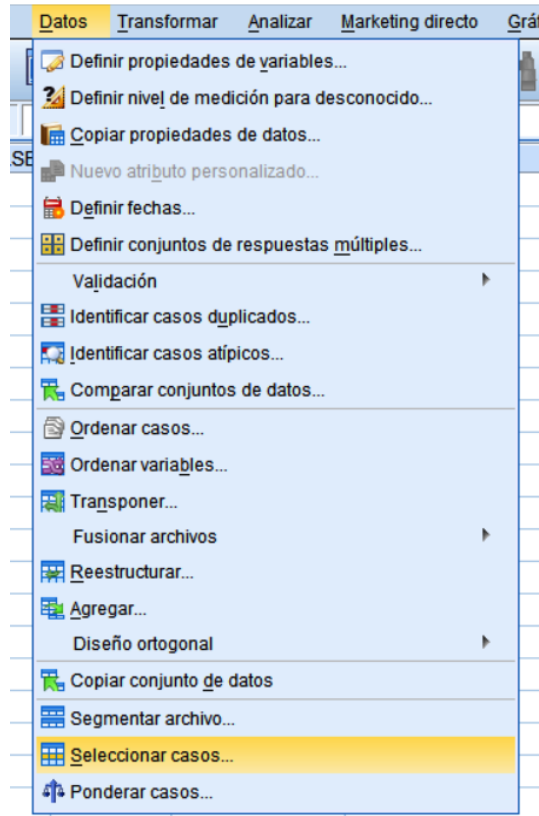
$$Me = 7,700$$

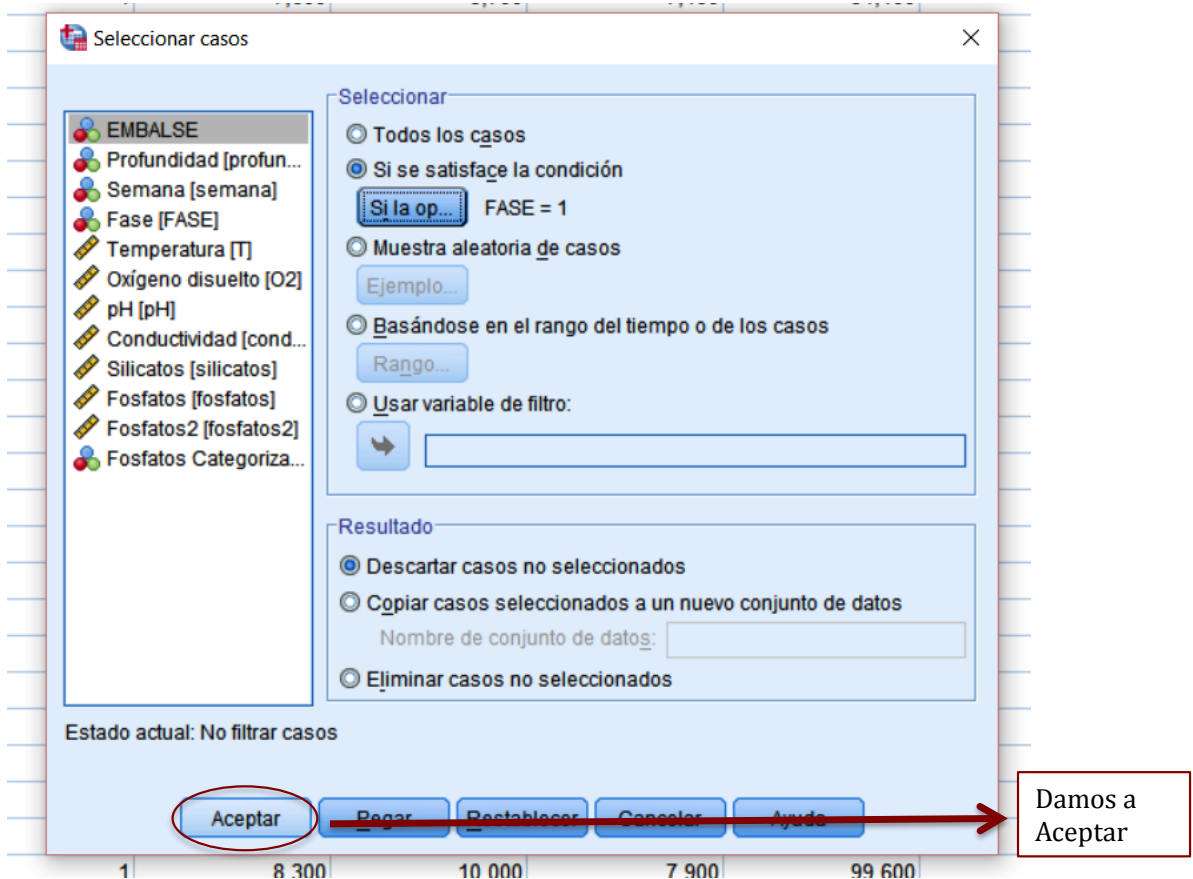
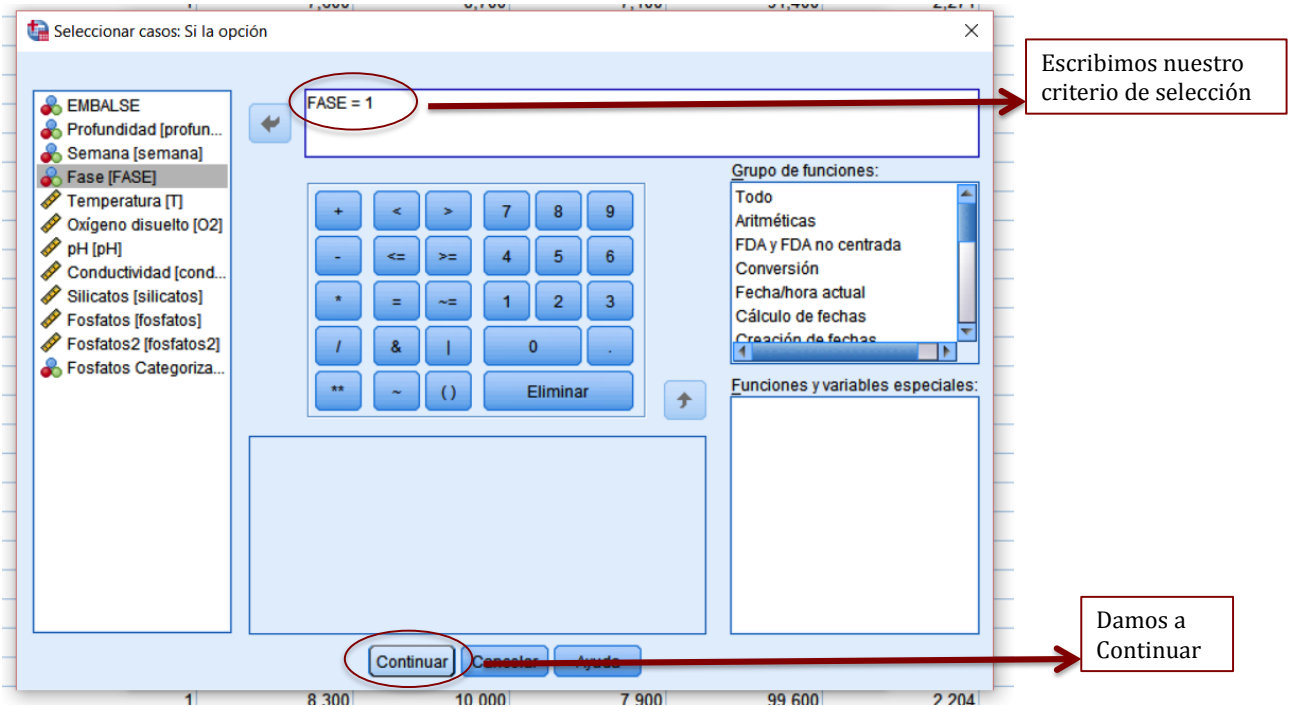
En la fase de Mezcla, el 50% de las observaciones tiene un nivel de pH inferior o igual a 7,7.

7. Calcule los siguientes percentiles en la fase de MEZCLA, interprete después el P90 y el P15 para cada variable.

Como solo queremos los percentiles para la fase de Mezcla (FASE=1), tenemos que trabajar **solo** con estos datos.

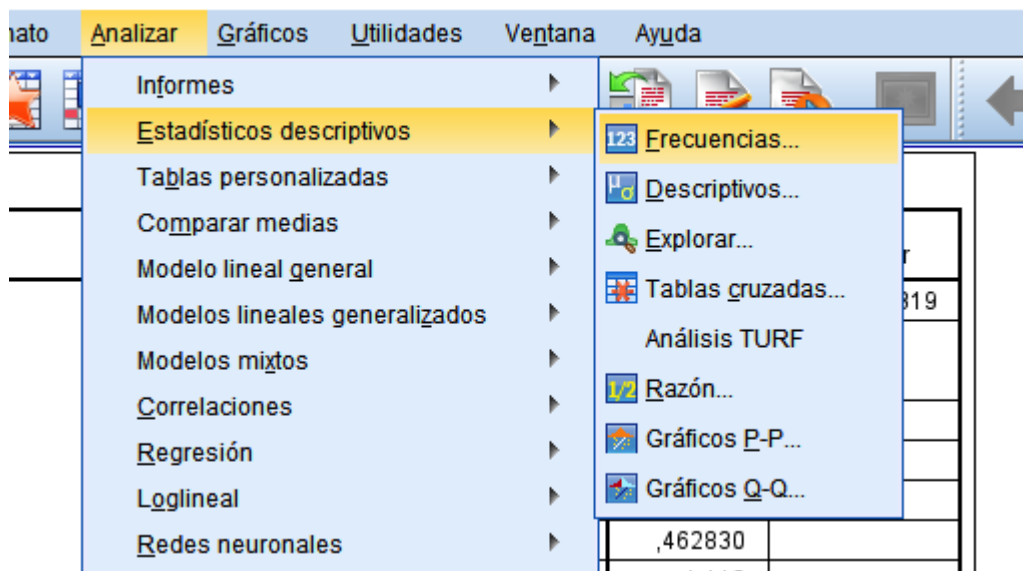
En el menú nos vamos a Datos -> Seleccionar casos





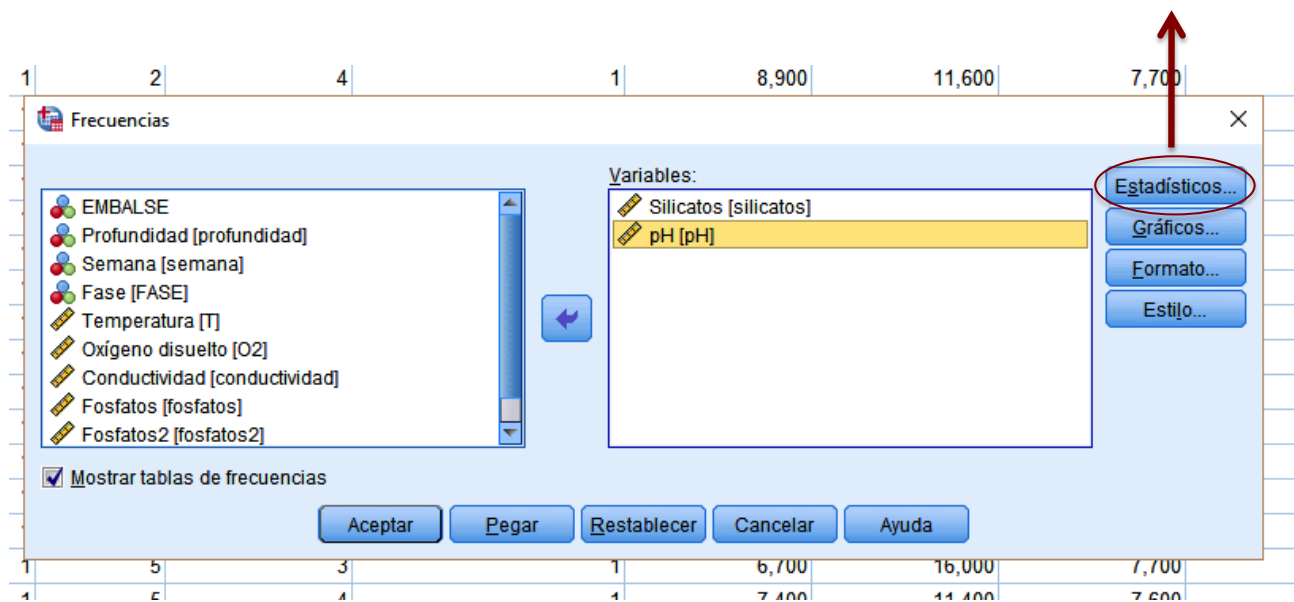
De esta forma, solo trabajaremos con los datos para los cuales FASE = 1 (fase de Mezcla).

En el menú nos vamos a Analizar -> Estadísticos descriptivos -> Frecuencias



Seleccionamos las variables para las que queremos calcular los estadísticos.

Con este botón seleccionamos los estadísticos a calcular



2,889	7,402	7,35	3
2,120	0,126	8,60	2

Frecuencias: Estadísticos

Valores percentiles

- Cuartiles
- Puntos de corte para: 10 grupos iguales
- Percentiles: 15

Añadir
Cambiar
Eliminar

Tendencia central

- Media
- Mediana
- Moda
- Suma

Los valores son puntos medios de grupos

Dispersión

- Desviación estándar
- Mínimo
- Varianza
- Máximo
- Rango
- Media de error estándar

Distribución

- Asimetría
- Curtosis

Continuar Cancelar Ayuda

Con este botón añadimos cada percentil a calcular

2,000	7,402	7,35	3
2,120	0,126	8,60	2

Frecuencias: Estadísticos

Valores percentiles

- Cuartiles
- Puntos de corte para: 10 grupos iguales
- Percentiles: 15,0

Añadir
Cambiar
Eliminar

Tendencia central

- Media
- Mediana
- Moda
- Suma

Los valores son puntos medios de grupos

Dispersión

- Desviación estándar
- Mínimo
- Varianza
- Máximo
- Rango
- Media de error estándar

Distribución

- Asimetría
- Curtosis

Continuar Cancelar Ayuda

Deseleccionamos esta opción porque no es correcto pedir una tabla de frecuencias

Estadísticos

		Silicatos	pH
N	Válido	70	70
	Perdidos	0	0
Percentiles	15	1,48856	7,40000
	25	1,65659	7,48415
	40	1,79172	7,60435
	75	2,32070	7,90000
	90	2,74900	8,09575

De esta forma, obtenemos los percentiles que se piden y completamos la tabla:

Variable	Percentiles				
	P ₁₅	P ₂₅	P ₄₀	P ₇₅	P ₉₀
SIO	1,489	1,657	1,792	2,321	2,749
pH	7,400	7,484	7,604	7,900	8,096

Interpretación para ambas variables de los P_{15} y del P_{90} :

En la fase de Mezcla, el 15% de las observaciones tiene un nivel de silicatos inferior o igual a 1,489.

En la fase de Mezcla, el 15% de las observaciones tiene un nivel de pH inferior o igual a 7,4.

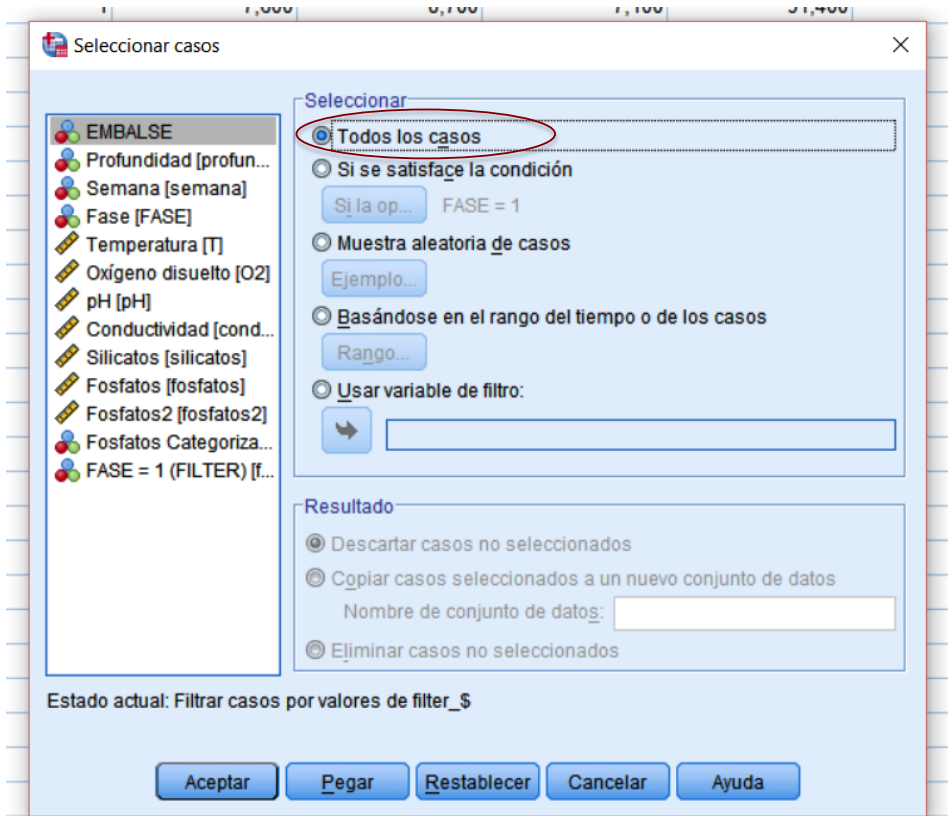
En la fase de Mezcla, el 90% de las observaciones tiene un nivel de silicatos inferior o igual a 2,749.

En la fase de Mezcla, el 90% de las observaciones tiene un nivel de pH inferior o igual a 8,096.

8. Realice la representación gráfica que considere más adecuada para analizar si la distribución del pH es aproximadamente simétrica en ambas FASES. Justifique la respuesta. Comente los resultados gráficos.

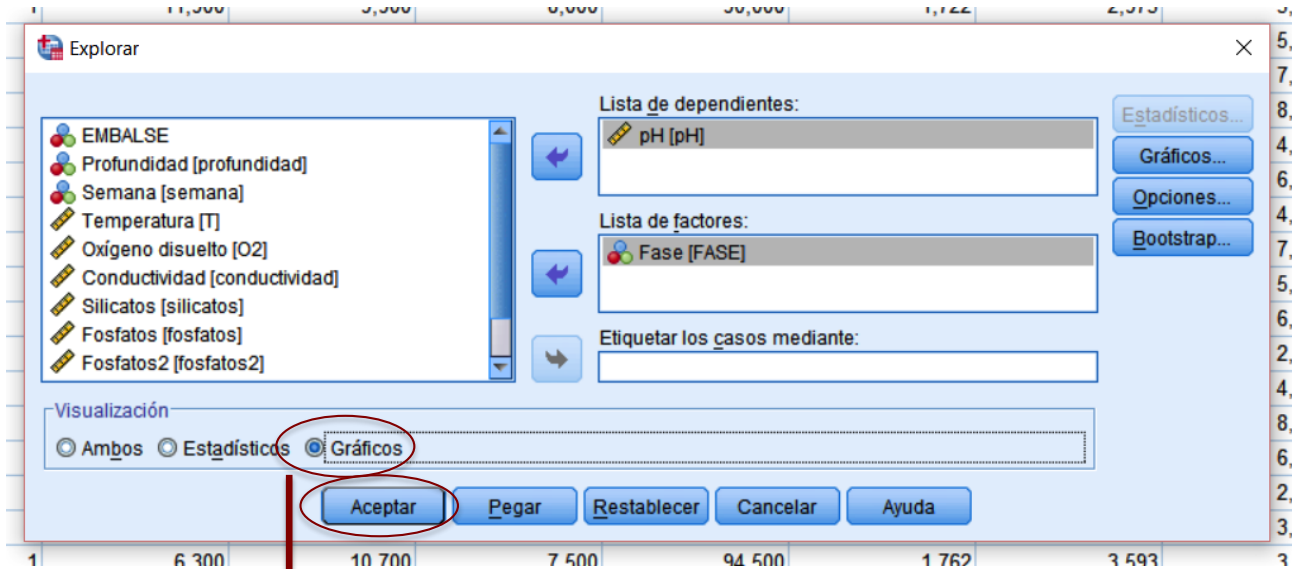
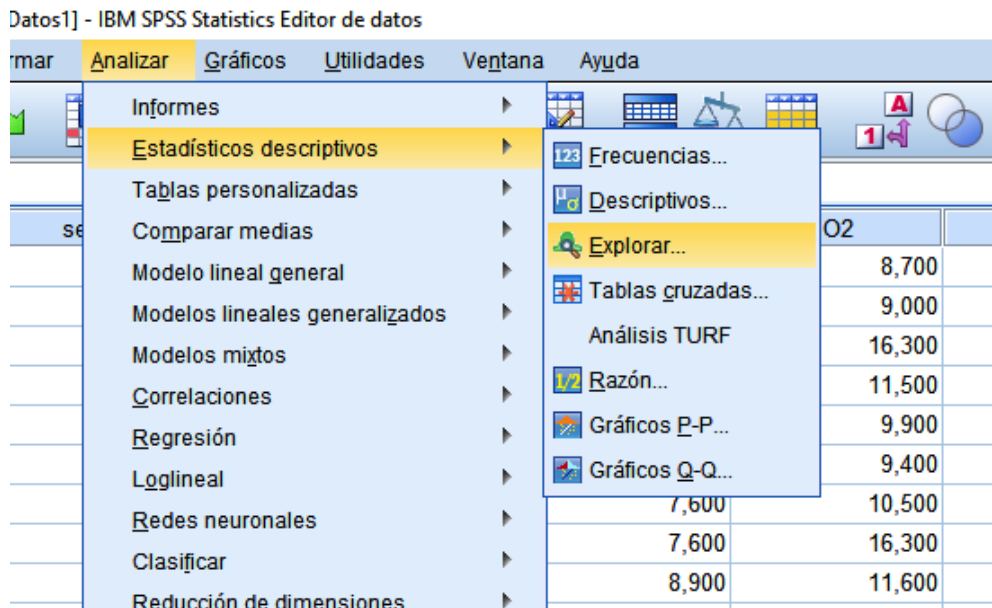
Recordemos que anteriormente habíamos seleccionado solo los datos de la fase de Mezcla, como ahora tenemos que trabajar con ambas fases tenemos que quitar esta selección.

Igual que antes, en el menú nos vamos a Datos -> Seleccionar casos, y marcamos la opción de **Todos los casos**, luego damos a aceptar.



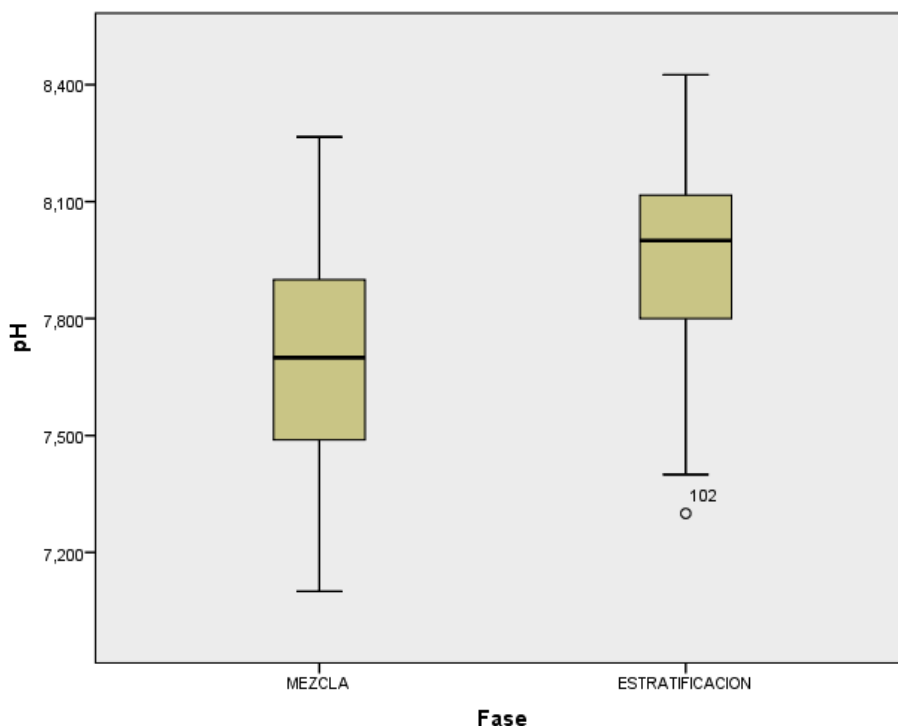
Ahora realizaremos un gráfico de cajas (Box Plot) para responder a la pregunta.

En el menú, Analizar -> Estadísticos descriptivos -> Explorar



Seleccionamos esta opción si solo queremos gráficos

Así obtenemos:

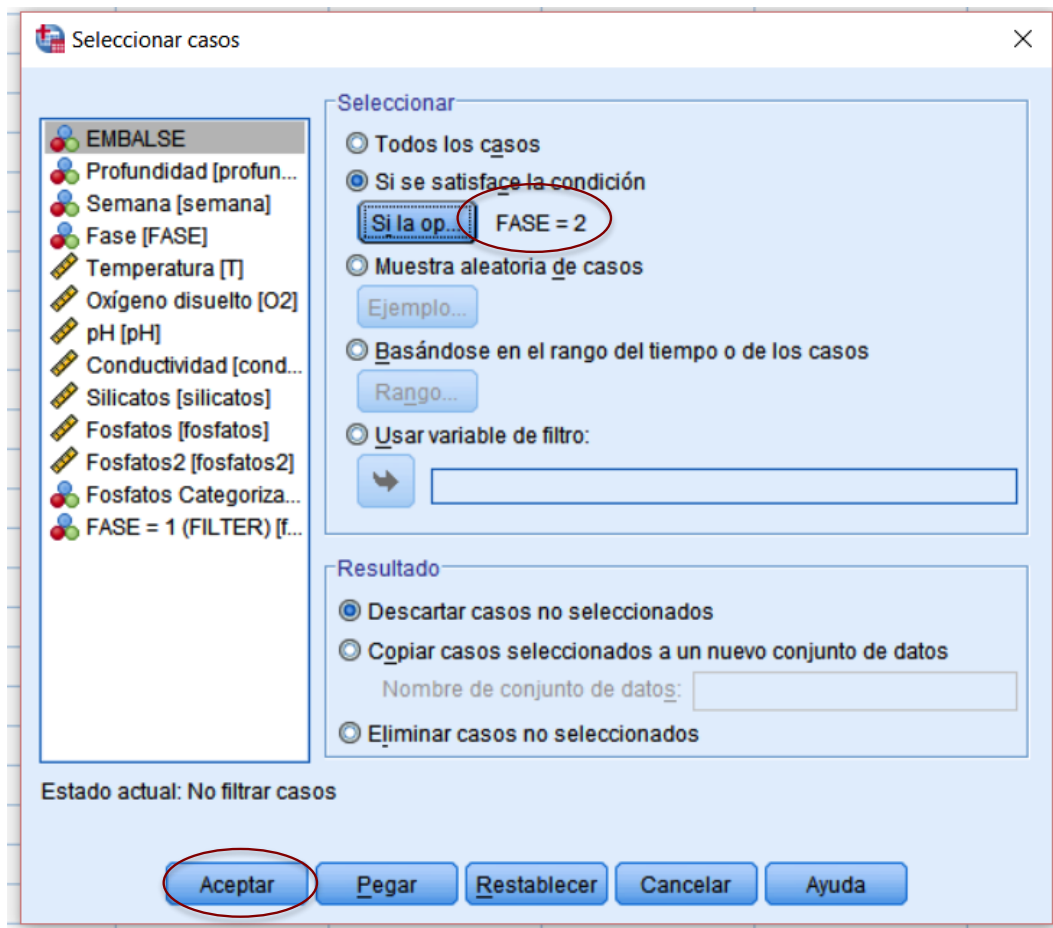


Podemos ver que, en la fase de Mezcla, la distribución del pH es simétrica, lo que no ocurre en la fase de Estratificación, donde la mediana tiene un desplazamiento que indica un pequeño sesgo o asimetría a la izquierda (negativa). Como información complementaria, podemos señalar que en la fase de Estratificación la variable pH presenta un punto atípico, específicamente la observación 102.

9. Realice la representación gráfica que considere más adecuada para analizar si la distribución de la Conductividad y de los Fosfatos es aproximadamente simétrica en la Fase de ESTRATIFICACIÓN. Justifique la respuesta. Comente los resultados gráficos.

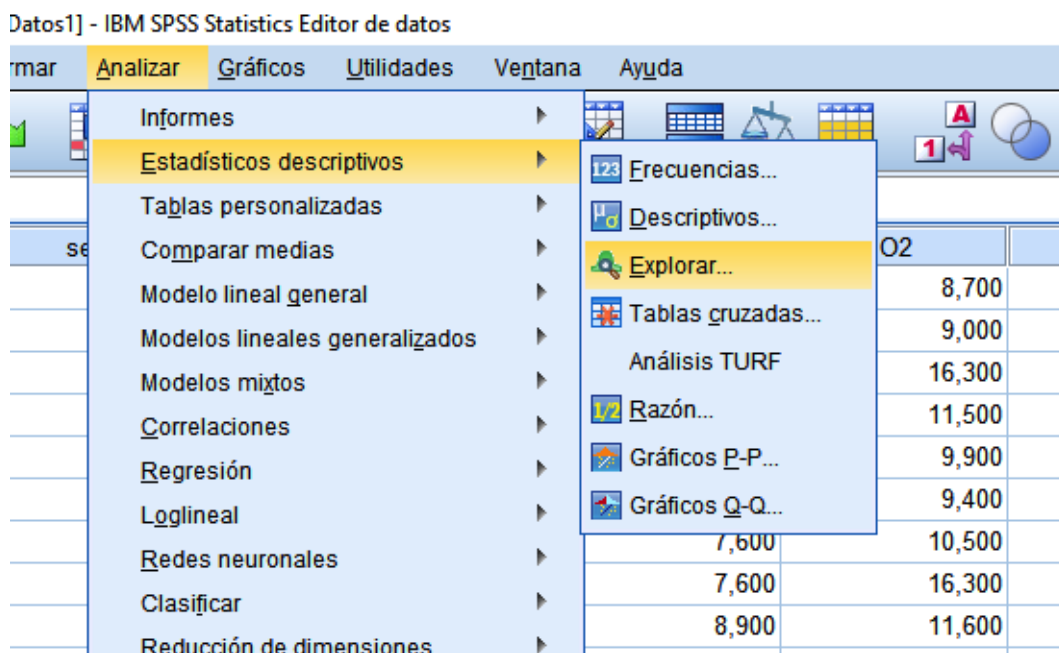
De forma análoga al ejercicio 7, no debemos trabajar con todos los datos, solo con aquellos de la fase de Estratificación (FASE=2).

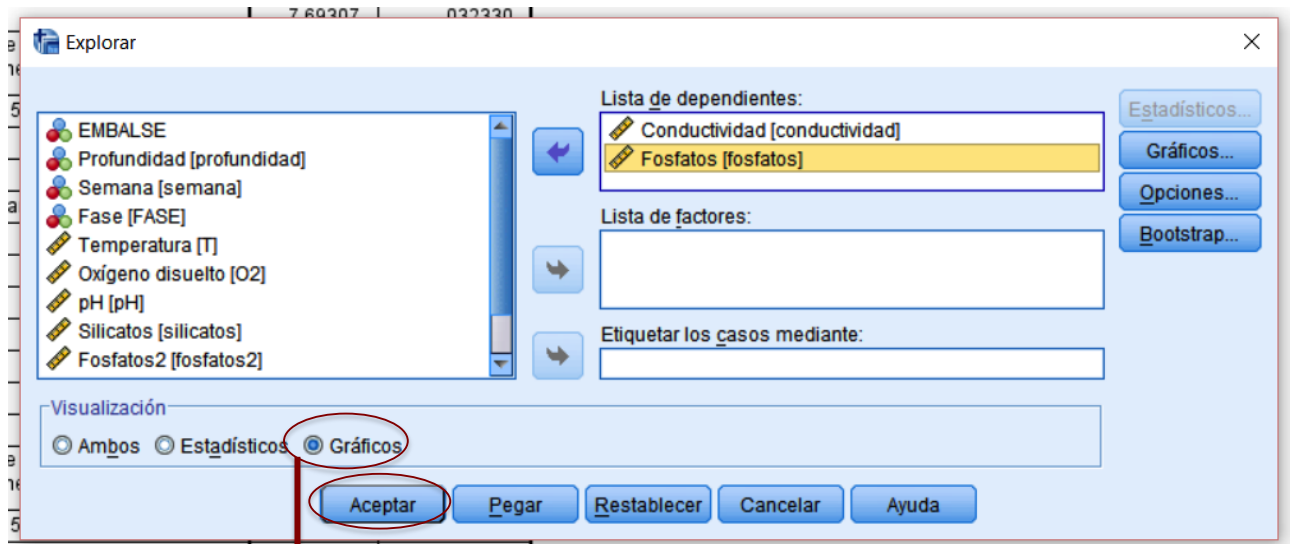
En el menú nos vamos a Datos -> Seleccionar casos, y nuestro criterio de selección será FASE = 2.



Una vez seleccionados los datos correctos, haremos un gráfico de caja para cada una de las variables.

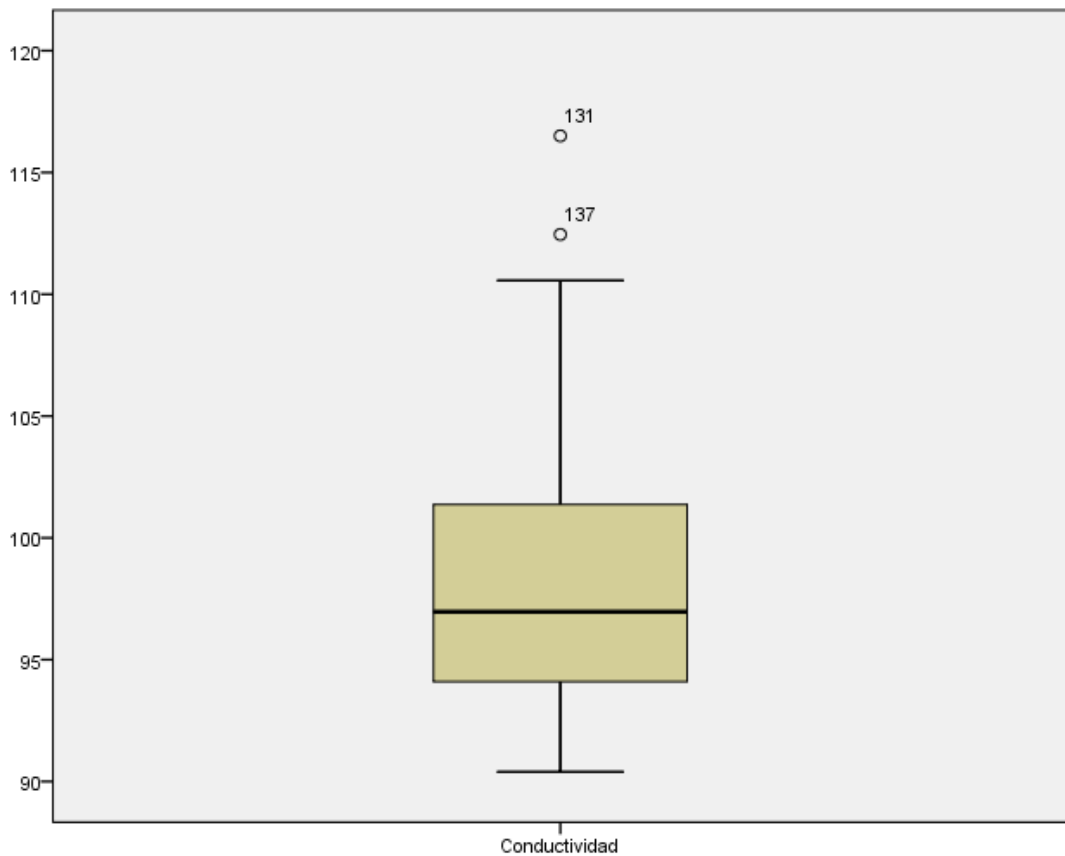
En el menú, Analizar -> Estadísticos descriptivos -> Explorar

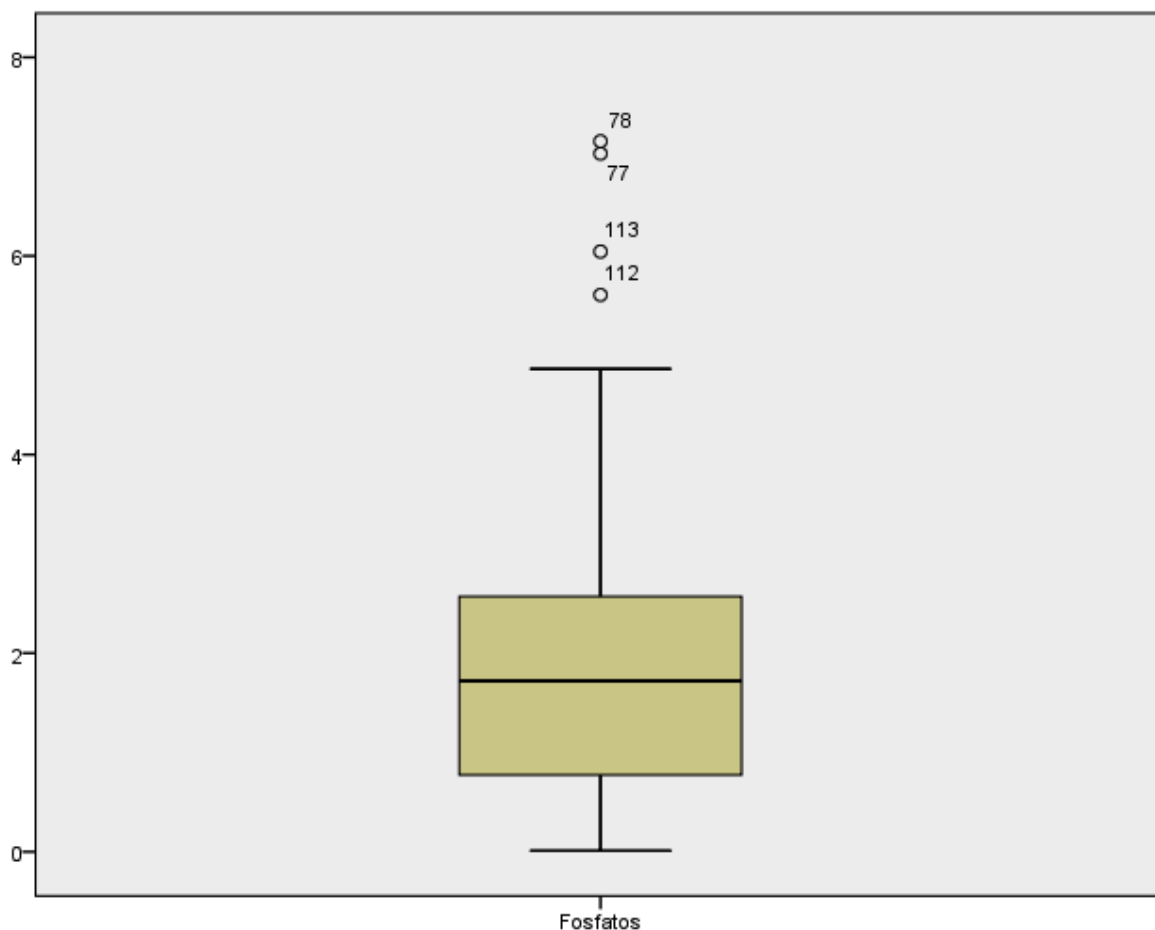




Seleccionamos esta opción si solo queremos gráficos

Así obtenemos:



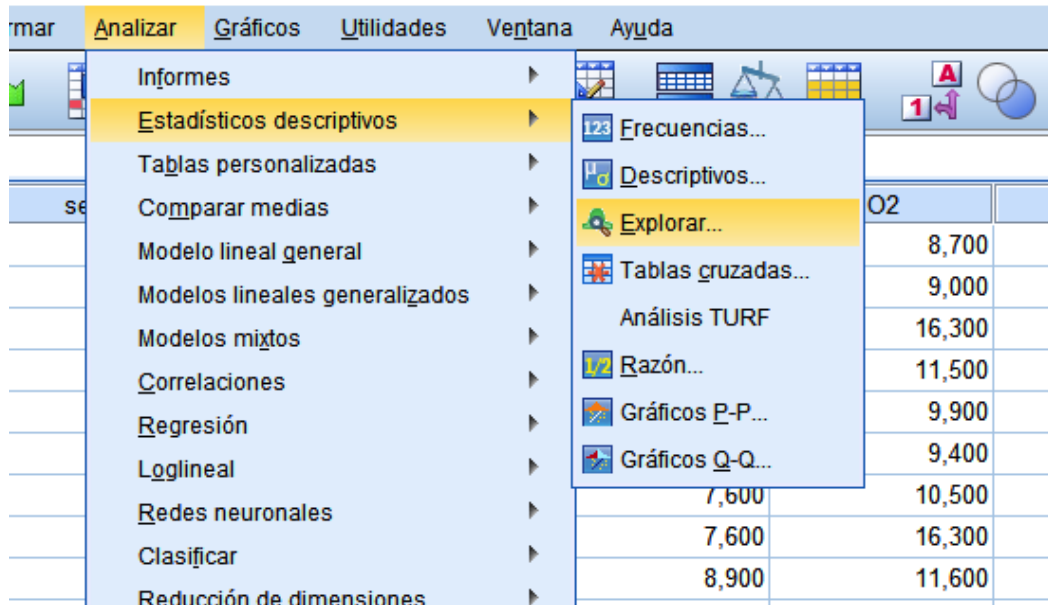


Podemos ver que la distribución de la variable conductividad presenta asimetría a la derecha (positiva) y que la distribución de la variable fosfatos presenta una pequeña asimetría a la izquierda (negativa). Como información complementaria, podemos señalar que en ambas variables tenemos observaciones atípicas y estas se encuentran en los valores superiores de la distribución.

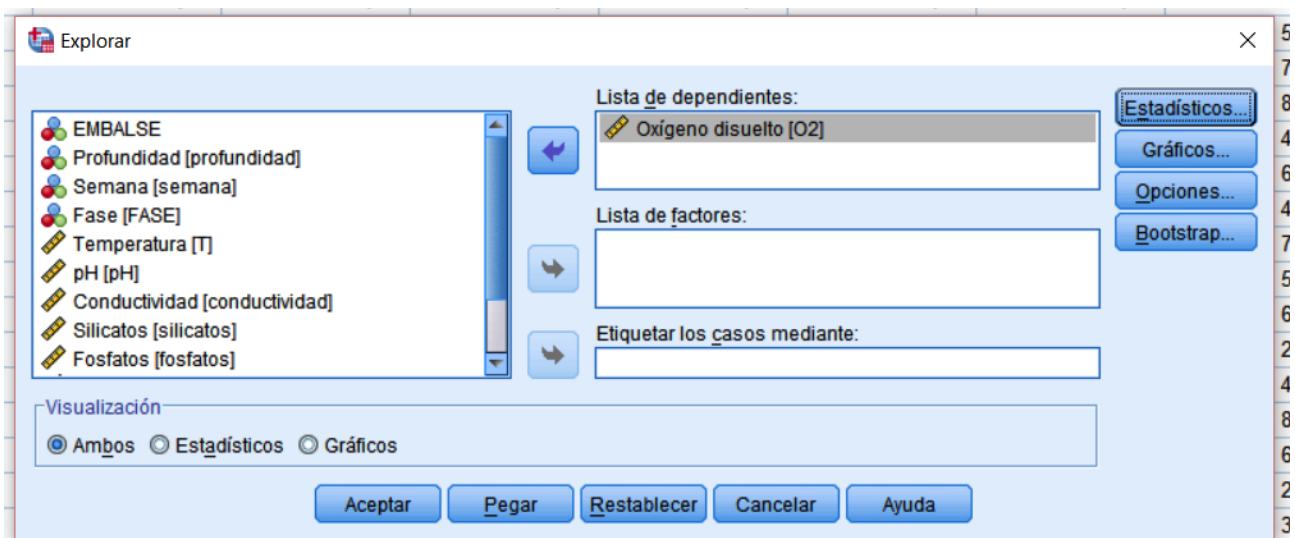
10. Escriba el valor de la media muestral y el intervalo de confianza al 95% para la media poblacional de la variable Oxígeno disuelto en la Fase de ESTRATIFICACIÓN (sponga normalidad). INTERPRETE LOS RESULTADOS.

Al igual que en el ejercicio anterior, tenemos que trabajar con los datos de la fase de Estratificación, por lo que la selección realizada con anterioridad sigue siendo adecuada.

Nuevamente nos vamos a Analizar -> Estadísticos descriptivos -> Explorar



Y consideramos para el análisis la variable oxígeno



Así obtenemos:

Descriptivos

		Estadístico	Error estándar
Oxígeno disuelto	Media	8,34006	,251140
	95% de intervalo de confianza para la media	Límite inferior Límite superior	7,83905 8,84107
	Media recortada al 5%	8,42715	
	Mediana	8,70000	
	Varianza	4,415	
	Desviación estándar	2,101185	
	Mínimo	2,854	
	Máximo	11,700	
	Rango	8,846	
	Rango intercuartil	3,176	
	Asimetría	-,550	,287
	Curtosis	-,416	,566

De esta forma tenemos $\bar{X} = 8,340$ y $I_{\mu}^{0,95} = [7,839; 8,841]$.

En la fase de estratificación, tenemos en promedio un Oxígeno disuelto de 8,34 unidades.

De los distintos intervalos

construidos a partir de sucesivos muestreos, un porcentaje del 95% contiene a la media poblacional de la variable Oxígeno disuelto (en la fase de estratificación). Tenemos una confianza del 95% de que el intervalo de 7,839 hasta 8,841 contenga a la media poblacional de la variable Oxígeno disuelto (en la fase de Estratificación).

Práctica 3

REGRESIÓN

OBJETIVOS y PREGUNTAS

Se piensa que existe una relación lineal entre la temperatura y los Silicatos, de forma que la Temperatura depende de los Silicatos del agua. Ajuste un Modelo de regresión lineal simple sólo para los datos correspondientes al embalse de LEÓN siguiendo los pasos que se piden a continuación:

1. Represente el diagrama de dispersión correspondiente, y comente cómo podría ser el tipo de relación entre ambas variables.
2. ¿Cuál es la magnitud de la relación. ¿Por qué? ¿Qué coeficiente utiliza para responder?
1. Ajuste un modelo de regresión lineal. Escriba el modelo.
2. Interpretación experimental de los coeficientes del modelo.
3. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.
4. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.
 - Calcule el valor pronosticado de para la Temperatura si el valor de los Silicatos es de 2.5. Comente si la predicción es adecuada y por qué. (Realice el cálculo con el ordenador utilizando 5 decimales). **No se olvide de poner, además, en el archivo de datos que ese dato que pretende predecir pertenece al embalse de LEÓN en la columna embalse.**
 - Realice ahora el ajuste también para el embalse de CÁCERES. ¿En qué embalse el ajuste lineal es mejor?. ¿Por qué?.

Se piensa que en la fase de ESTRATIFICACIÓN en el embalse de LEÓN, en aguas poco profundas, hasta los 10 m. de profundidad, la relación entre el Silicato (X) y la Temperatura (Y) es de tipo no lineal.

Ajuste, por tanto, para los datos de la fase de Estratificación¹, los modelos no lineales: Parabólico y Potencial. Compárelos y decida cuál de los dos es el mejor para explicar la relación entre las dos variables.

Trabaje con logaritmos neperianos, compare el poder predictivo, y el poder explicativo de cada modelo.

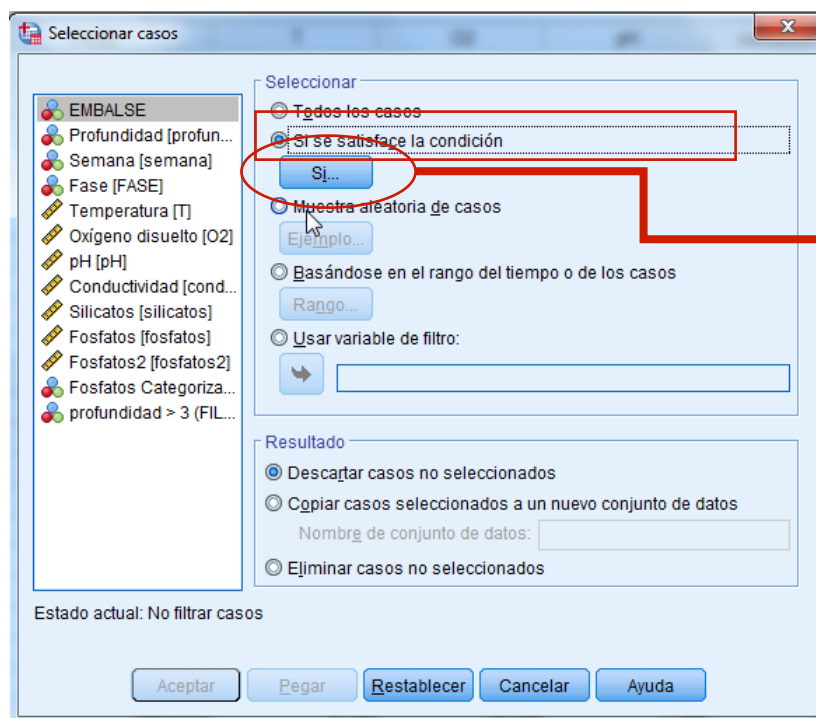
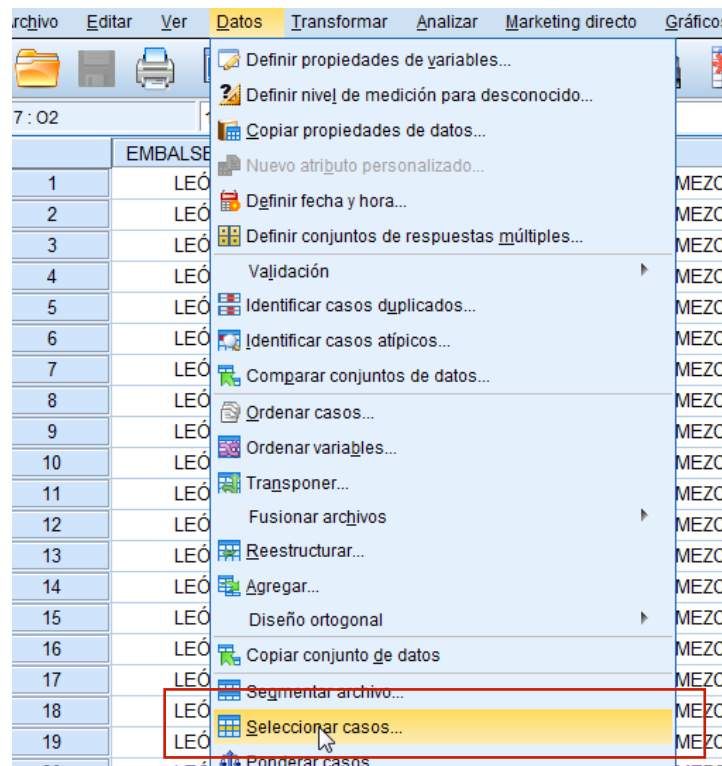
1. Represente el diagrama de dispersión correspondiente, y comente cómo podría ser el tipo de relación entre ambas variables.
 2. Ajuste un modelo de regresión parabólico. Escriba el modelo.
 3. ¿Cuál es la magnitud de la relación?. ¿Por qué? ¿Qué coeficiente utiliza para responder?.
 4. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.
 5. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.
 6. Ajuste un modelo de regresión potencial. Escriba el modelo.
 7. ¿Cuál es la magnitud de la relación?. ¿Por qué? ¿Qué coeficiente utiliza para responder?.
 8. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.
 9. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.
 10. ¿Cuál de los dos modelos es el mejor para explicar la relación entre las dos variables.
- Realice ahora el ajuste también para el embalse de CÁCERES. ¿En qué embalse el ajuste es mejor?. ¿Por qué?.

¹ Tenga en cuenta que en este apartado sólo debe trabajar con la fase de ESTRATIFICACIÓN y sólo en el embalse de LEÓN. Consulte la primera parte de las prácticas para ver cómo se seleccionan sólo algunos de los casos, y cómo segmentar (dividir) el archivo, pues necesitará realizar los dos procedimientos para el ajuste de la parábola. Tenga además en cuenta cómo están definidos en el SPSS los valores para cada categoría de cada variable.

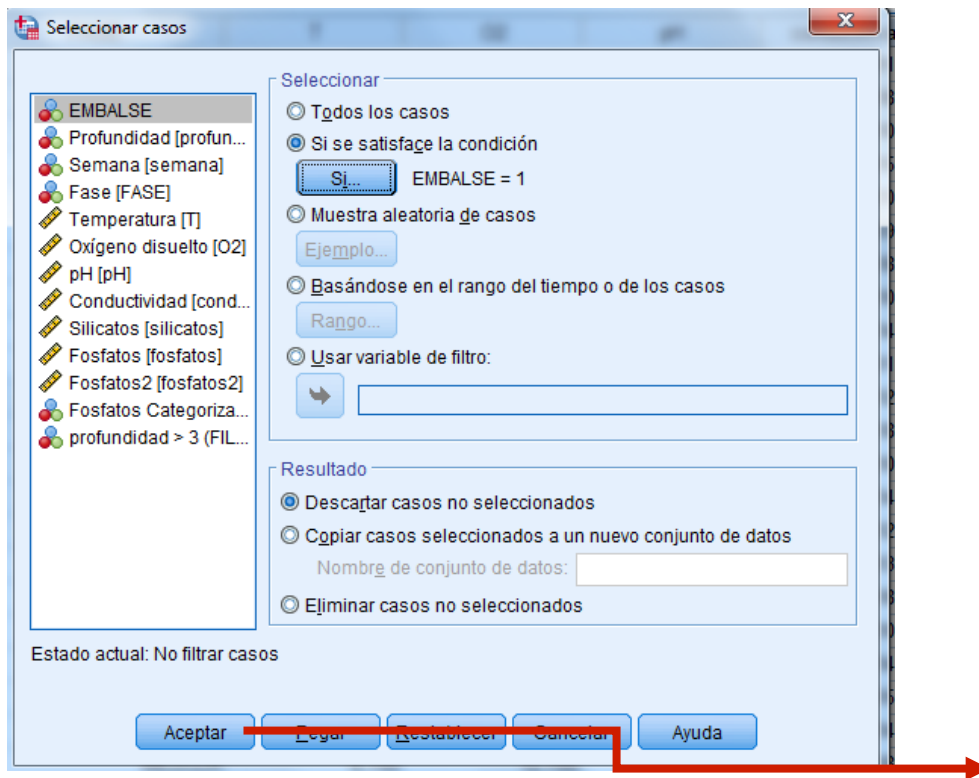
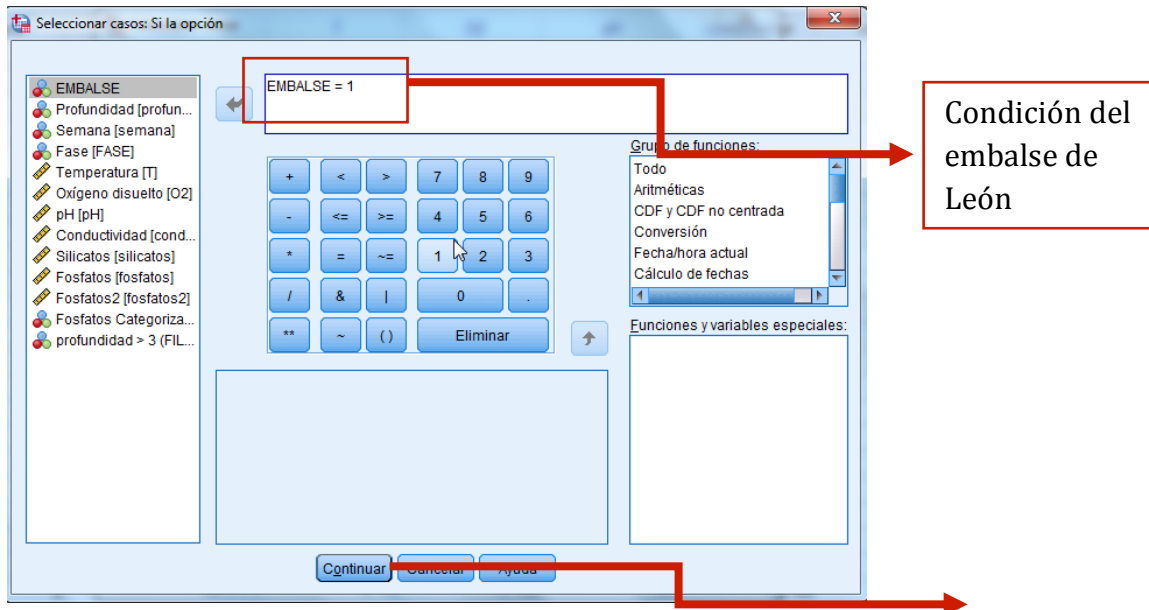
Práctica 3 REGRESIÓN

SOLUCIÓN

1. *Represente el diagrama de dispersión correspondiente, y comente cómo podría ser el tipo de relación entre ambas variables.*



En este botón se selecciona la condición a cumplir



Gráficos Utilidades Ventana Ayuda

Generador de gráficos...
Selector de plantillas de tablero...

Cuadros de diálogo antiguos

	rosstatosZ	rosstatos_recZ	
413	4,14		media
340	7,83		alta
946	5,81		alta
132	5,87		alta
973	3,18		baja
668	5,66		alta
402	7,35		alta
136	8,60		alta
522	4,98		media
875	6,58		alta
040	4,96		media

Barras...
Barras 3D...
Líneas...
Áreas...
Circular...
Máximos y mínimos...
Diagramas de cajas...
Barras de error...
Pirámide de población...
Dispersión/Puntos...
Histograma...

Dispersión/Puntos

Dispersión simple
Dispersión matricial
Puntos simples
Dispersión superpuesta
Dispersión 3D

Definir Aceptar Ayuda

Dispersión a seleccionar

Diagrama de dispersión simple

Eje Y: Temperatura [T]
Eje X: Silicatos [silicatos]

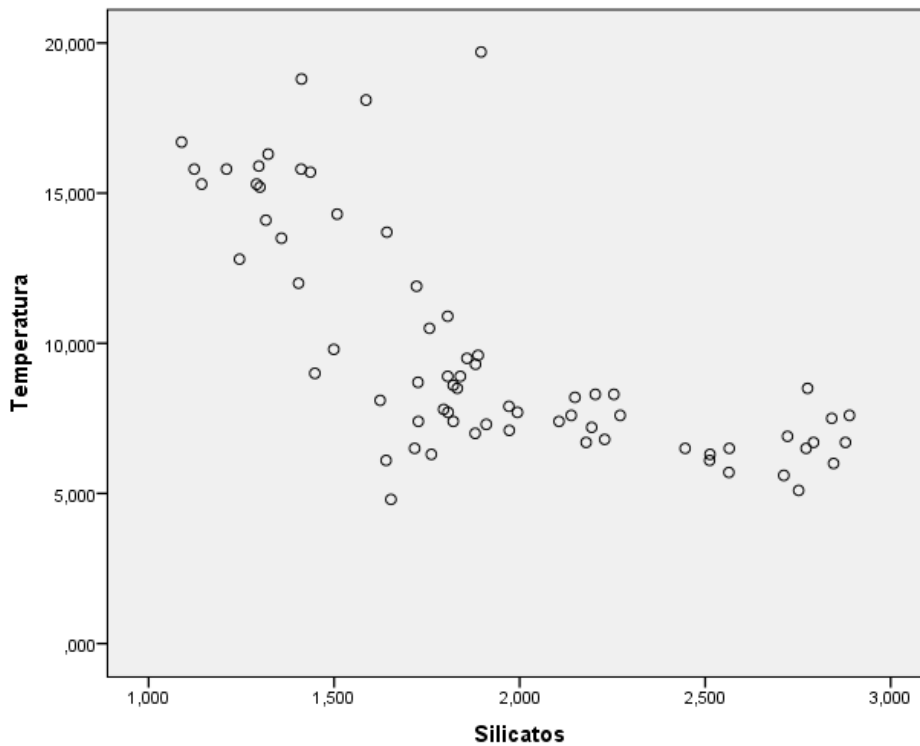
Establecer marcas por:
Etiquetar los casos mediante:

Panel mediante
Filas:
Anidar variables (sin filas vacías)
Columnas:
Anidar variables (sin columnas vacías)

Plantilla
 Usar las especificaciones gráficas de:
Archivo...

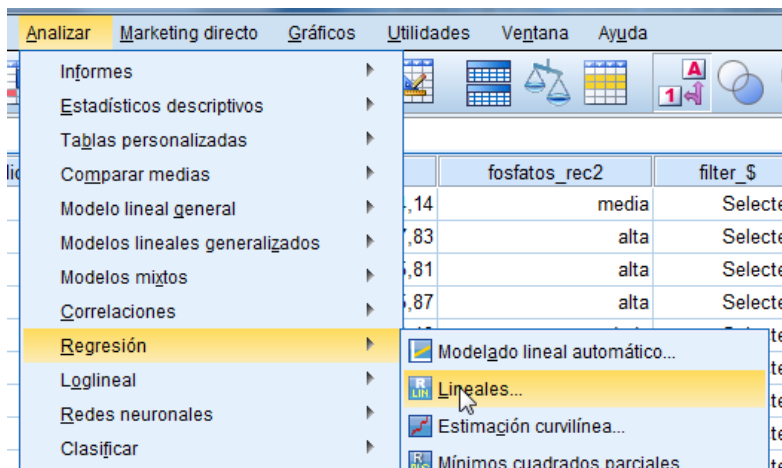
Aceptar Aceptar Establecer Cancelar Ayuda

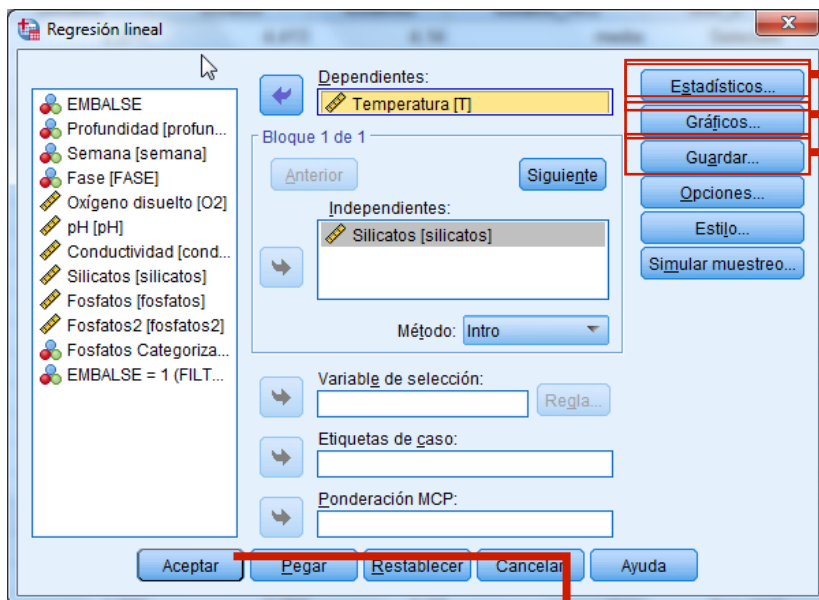
Variables seleccionadas del grupo de VARIABLES



En el gráfico se puede observar una cierta **relación lineal decreciente**, pero también podría acercarse mejor una **parábola**.

2. ¿Cuál es la magnitud de la relación. ¿Por qué? ¿Qué coeficiente utiliza para responder?





Pueden seleccionarse diferentes estadísticos

Pueden seleccionarse gráfico de residuales

Pueden seleccionarse diferentes estadísticos

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	.707 ^a	.500	.493	360367

Coefficiente de correlación lineal (grado de relación lineal entre las variables)

Coefficiente de determinación (multiplicado por 100 porcentaje explicado por el modelo)

a. Predictores: (Constante), Temperatura

3. Ajuste un modelo de regresión lineal. Escriba el modelo.

Coefficientes^a

Modelo		Coefficients no estandarizados		Coefficients estandarizados	t	Sig.
		B	Error estándar	Beta		
1	(Constante)	20,162	1,302		15,490	,000
	Silicatos	-5,408	,656	-,707	-8,248	,000

a. Variable dependiente: Temperatura

Coefficientes del modelo lineal

x

Por lo tanto, el modelo lineal se escribiría como: $y = 20,162 - 5,408x$
 Es decir: $Temperatura = 20,162 - 5,408 \cdot Silicatos$

4. Interpretación experimental de los coeficientes del modelo.

El término independiente (20,162) se interpreta como valor experimental de la y cuando la x es 0, es decir, el valor experimental de los **Silicatos** es de 20,162 cuando la **temperatura** es 0.

El coeficiente de la x (-5,408) se interpreta como incremento experimental de la y cuando la x se incremente en una unidad, es decir, el decrecimiento experimental de la **temperatura** es de 5,408 cuando los **Silicatos** se incremente en 1 unidad.

5. Considere que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.

Hemos visto que el modelo explica el 50%, por lo tanto este modelo lineal no sería muy bueno para explicar la relación entre las dos variables.

6. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.

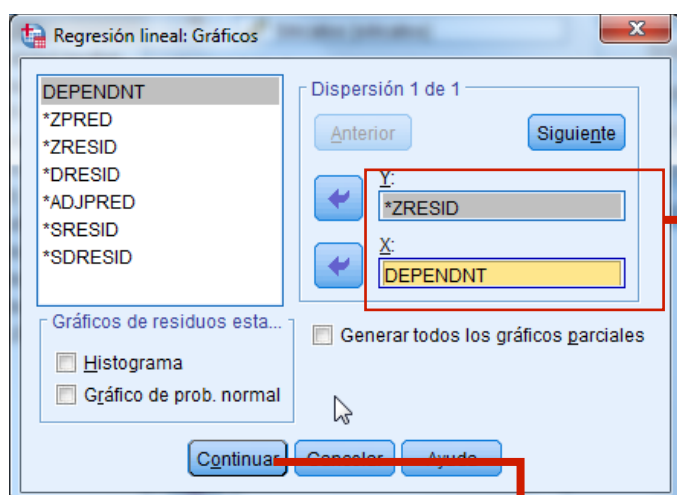
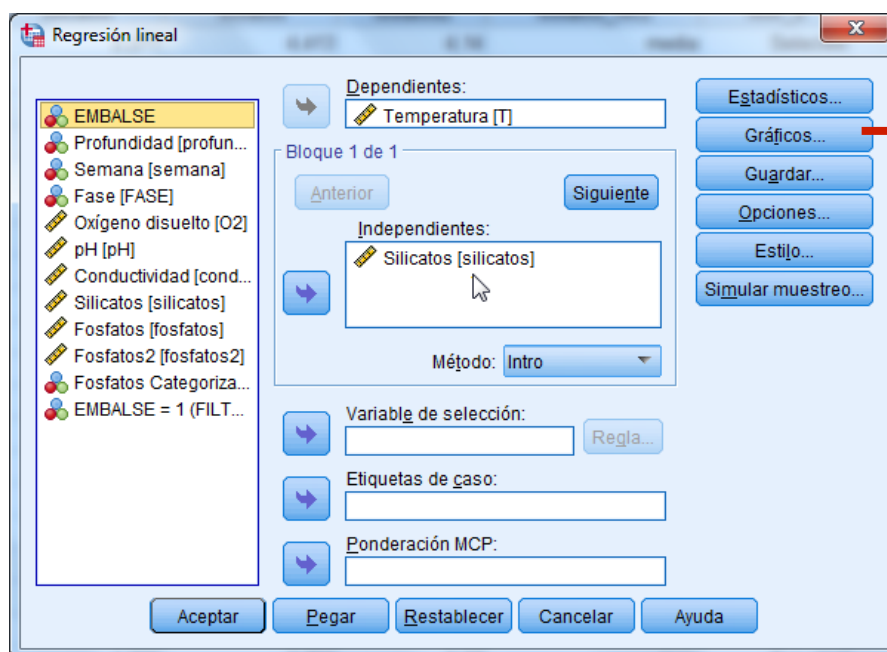


Gráfico de los residuos frente a la variable dependiente.

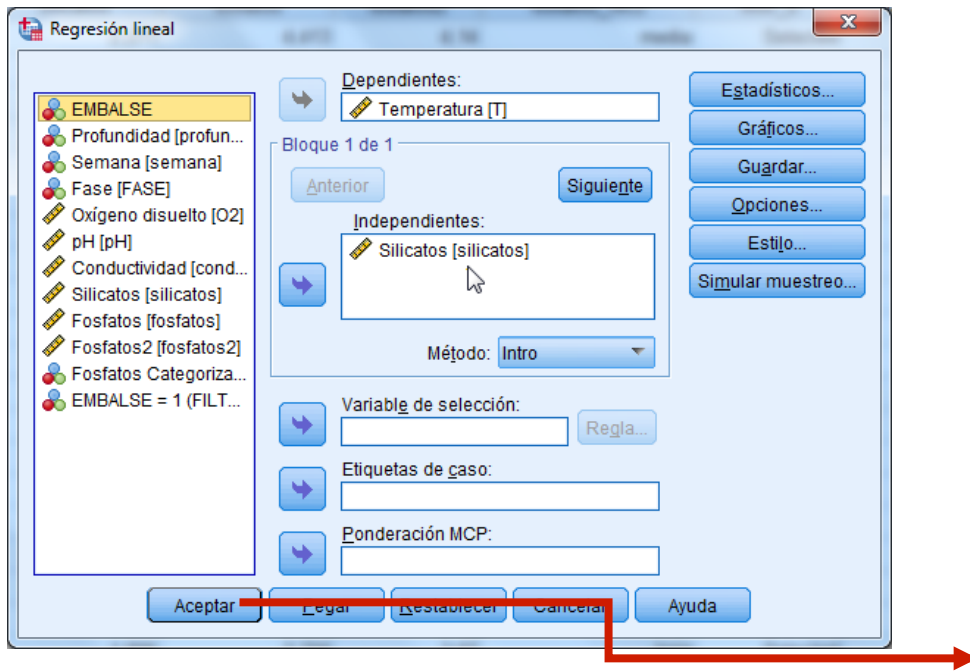
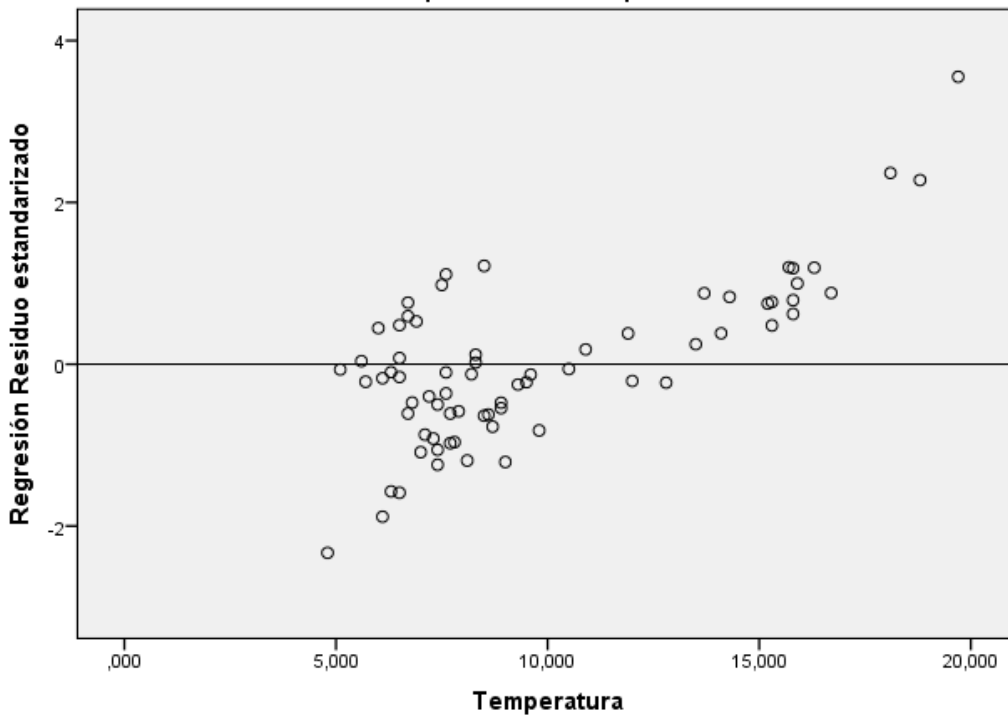


Diagrama de dispersión

Variable dependiente: Temperatura



Este gráfico de residuales tiene cierta forma, que junto con el bajo poder explicativo del apartado 7 nos llevan a deducir que este modelo lineal no es adecuado para explicar la relación de la temperatura en función de los silicatos.

- Calcule el valor pronosticado de para la Temperatura si el valor de los Silicatos es de 2.5. Comente si la predicción es adecuada y por qué. (Realice el cálculo con el ordenador utilizando 5 decimales). **No se olvide de poner, además, en el archivo de datos que ese dato que pretende predecir pertenece al embalse de LEÓN en la columna embalse.**

	EMBALSE	profundidad	semana	FASE	T	O2	pH	conductividad	silicatos	fos
136	CÁCERES	Fdo.	6	ESTRATIFICACION	9,700	5,486	7,524	110,240	2,475	
137	CÁCERES	Fdo.	7	ESTRATIFICACION	9,389	5,389	7,721	112,456	2,359	
138	CÁCERES	Fdo.	8	ESTRATIFICACION	8,045	4,105	7,538	110,569	2,475	
139	CÁCERES	Fdo.	9	ESTRATIFICACION	8,028	5,594	7,979	98,823	2,640	
140	CÁCERES	Fdo.	10	ESTRATIFICACION	6,611	4,759	8,159	95,209	2,194	
141	LEÓN								2,500	
142										

Regresión lineal

Dependientes: Temperatura [T]

Bloque 1 de 1

Independientes: Silicatos [silicatos]

Método: Intro

Variable de selección:

Etiquetas de caso:

Ponderación MCP:

Estadísticos...
Gráficos...
Guardar...
Opciones...
Estilo...
Simular muestreo...

Aceptar Pegar Restablecer Cancelar Ayuda

Regresión lineal: Guardar

Valores pronosticados

No estandarizados

Estandarizados

Corregidos

Error estándar de predicciones de media

Residuos

No estandarizados

Estandarizados

Método de Student

Eliminados

Eliminados estudentizados

Distancias

Mahalanobis

De Cook

Valores de influencia

Intervalos de predicción

Media Individuos

Intervalo de confianza: 95 %

Estadísticos de influencia

DfBetas

DfBetas estandarizadas

DfFit

DfFit estandarizado

Razón entre covarianzas

Estadísticos de los coeficientes

Crear estadísticos de los coeficientes

Crear un nuevo conjunto de datos

Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Archivo...

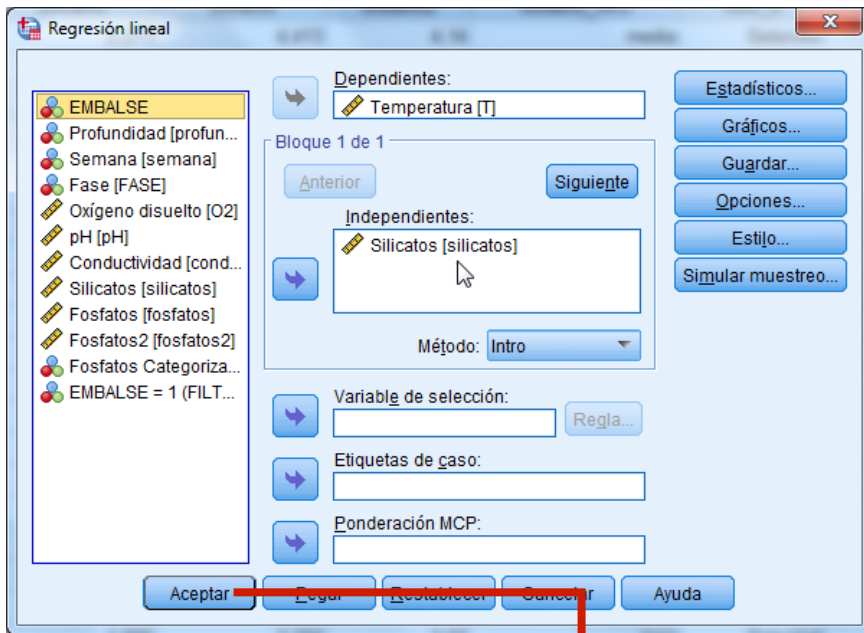
Exportar información del modelo a un archivo XML

Incluir la matriz de covarianzas

Examinar...

Continuar Cancelar Ayuda

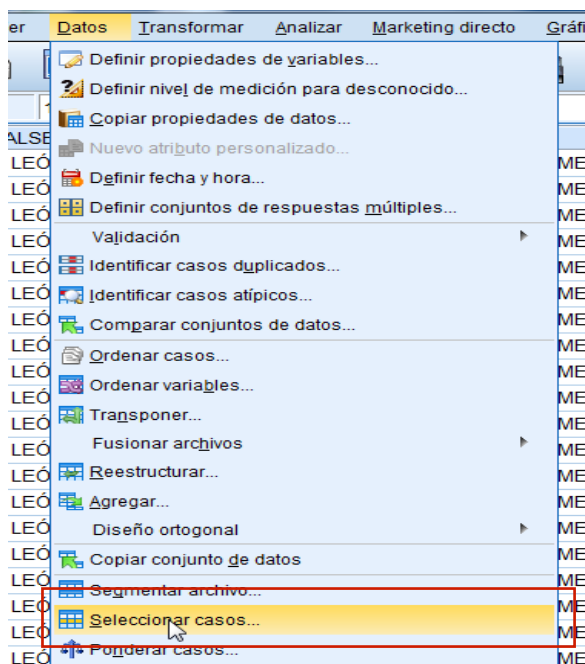
Seleccionar valores pronosticados no estandarizados

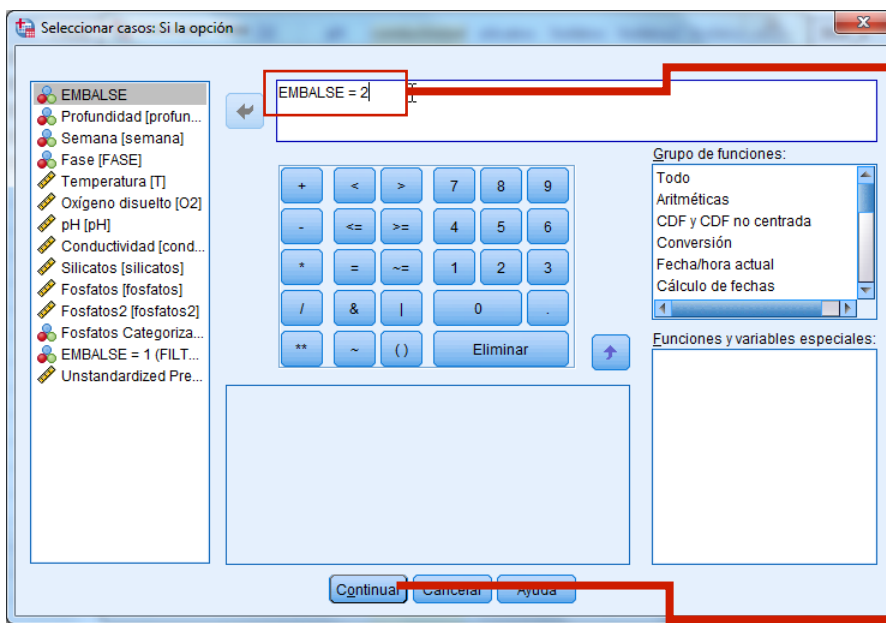
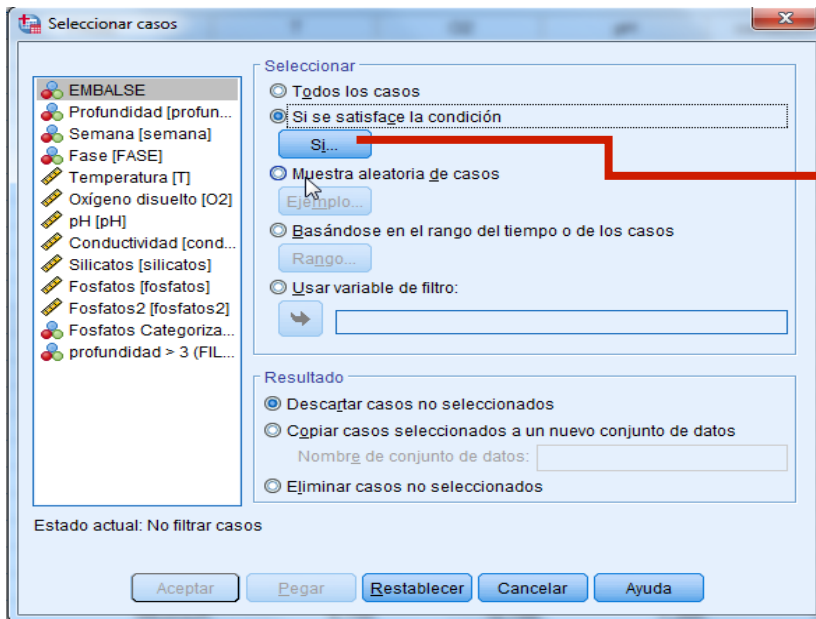


	EMBALSE	profundidad	semana	FASE	T	O2	pH	conductividad	silicatos	fosfatos	fosfatos2	fosfatos_rec2	filter_\$	PRE_1
136	CÁCERES	Fdo.	6	ESTRATIFICACION	9,700	5,486	7,524	110,240	2,475	,363	,80	baja	Not Selected	.
137	CÁCERES	Fdo.	7	ESTRATIFICACION	9,389	5,389	7,721	112,456	2,359	1,123	1,66	baja	Not Selected	.
138	CÁCERES	Fdo.	8	ESTRATIFICACION	8,045	4,105	7,538	110,569	2,475	,938	1,46	baja	Not Selected	.
139	CÁCERES	Fdo.	9	ESTRATIFICACION	8,028	5,594	7,979	98,823	2,640	,097	,63	baja	Not Selected	.
140	CÁCERES	Fdo.	10	ESTRATIFICACION	6,611	4,759	8,159	95,209	2,194	,301	,77	baja	Not Selected	.
141	LEÓN								2,500				Selected	6,64324

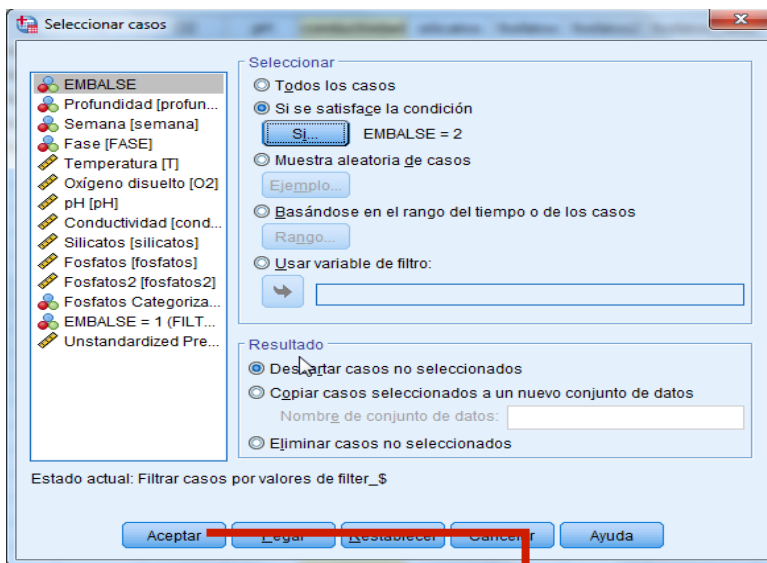
El valor previsto para la temperatura en el embalse de LEÓN cuando el valor de los silicatos sea de 2,5 es de 6,64324.

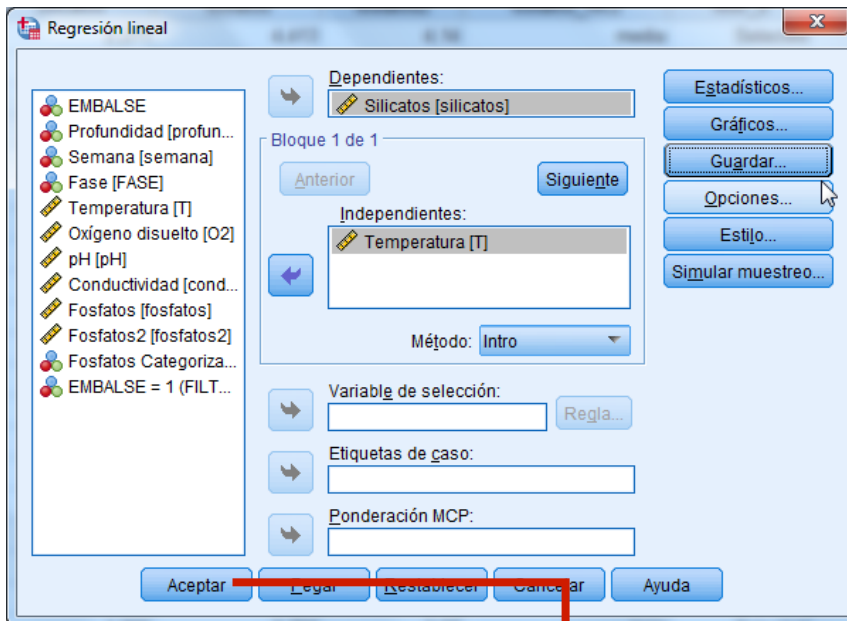
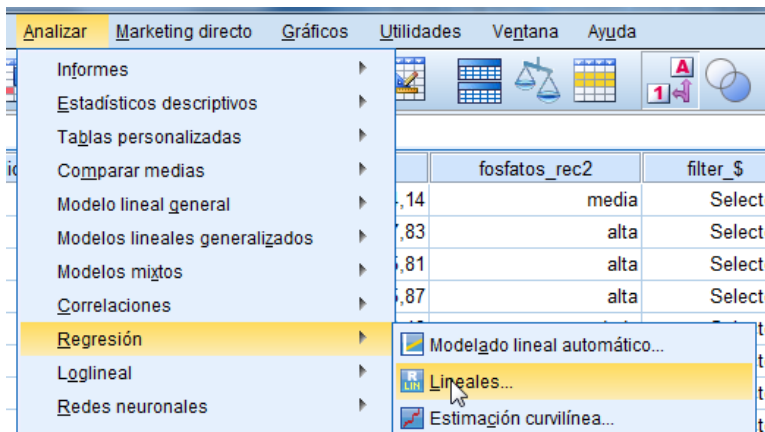
- Realice ahora el ajuste también para el embalse de CÁCERES. ¿En qué embalse el ajuste lineal es mejor. ¿Por qué?





Condición embalse de CÁCERES





Resumen del modelo^b

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,637 ^a	,405	,397	3,369606

a. Predictores: (Constante), Silicatos

b. Variable dependiente: Temperatura

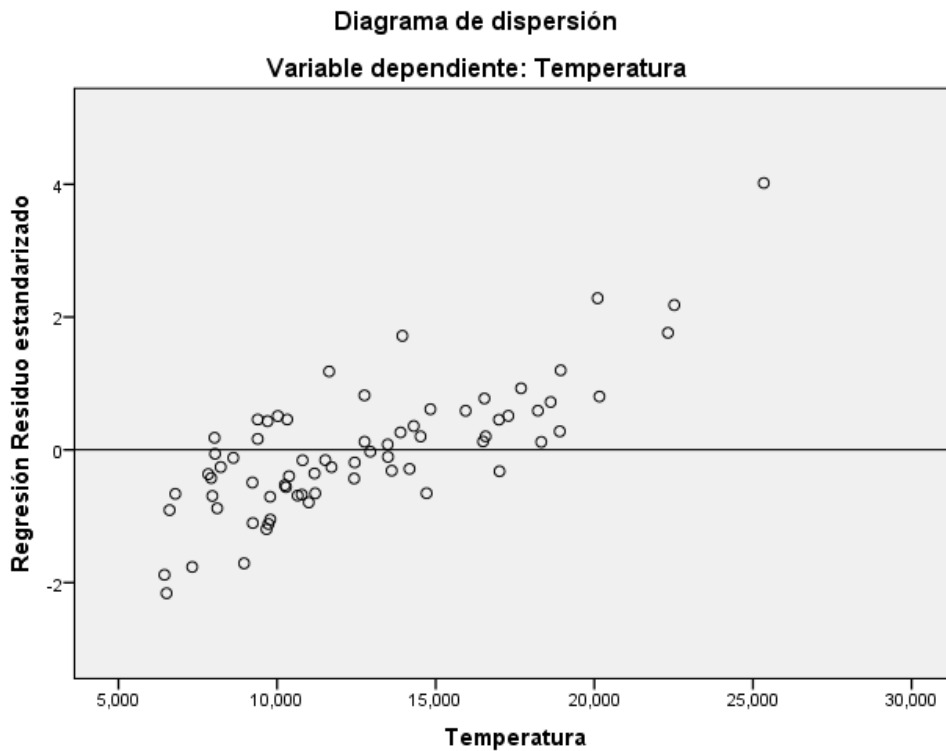
Coefficientes^a

Modelo		Coefficients no estandarizados		Coefficients estandarizados		t	Sig.
		B	Error estándar	Beta			
1	(Constante)	20,750	1,252			16,570	,000
	Silicatos	-5,052	,742	-,637		-6,809	,000

a. Variable dependiente: Temperatura

El valor de R^2 es menor, el modelo $Temperatura = 20,75 - 5,052 \cdot Silicatos$ solo explica el 40,5 % de las variaciones de la temperatura a partir de las variaciones de los silicatos.

Y el gráfico de residuales es el siguiente:



Que también tiene cierta forma.

Por lo tanto, el ajuste lineal es mejor en el embalse de LEÓN.

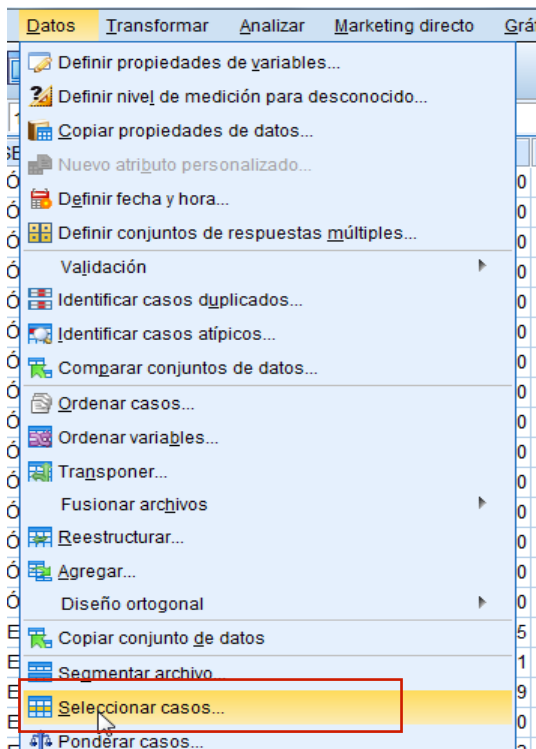
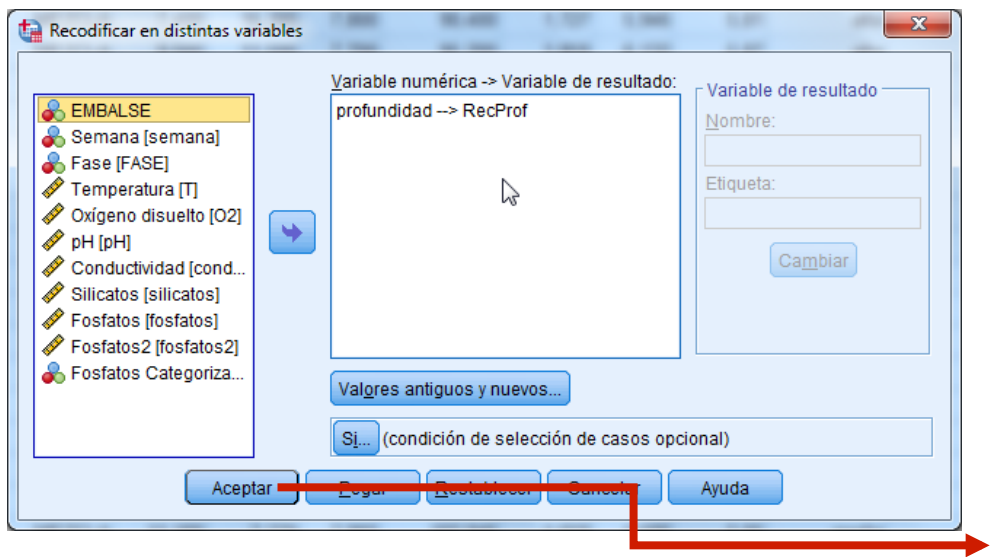
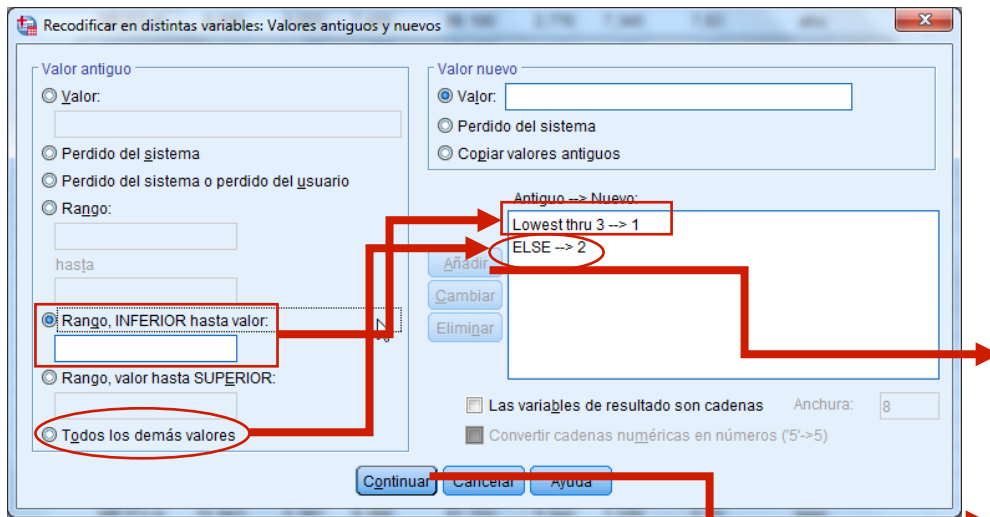
Se piensa que en la fase de ESTRATIFICACIÓN en el embalse de LEÓN, en aguas poco profundas, hasta los 10 m. de profundidad la relación entre el Silicato (X) y la Temperatura (Y) es de tipo no lineal.

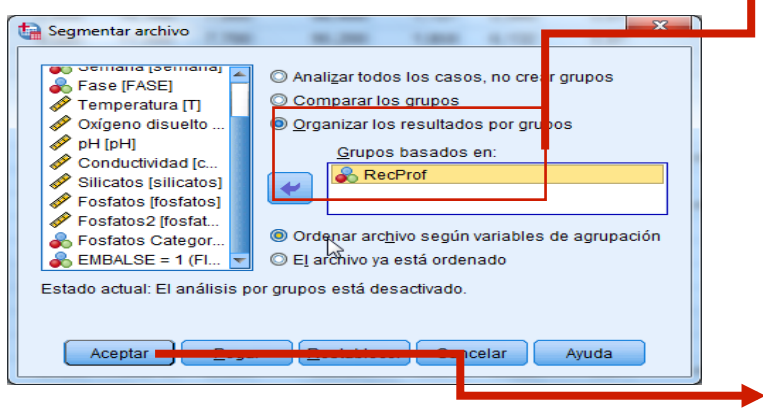
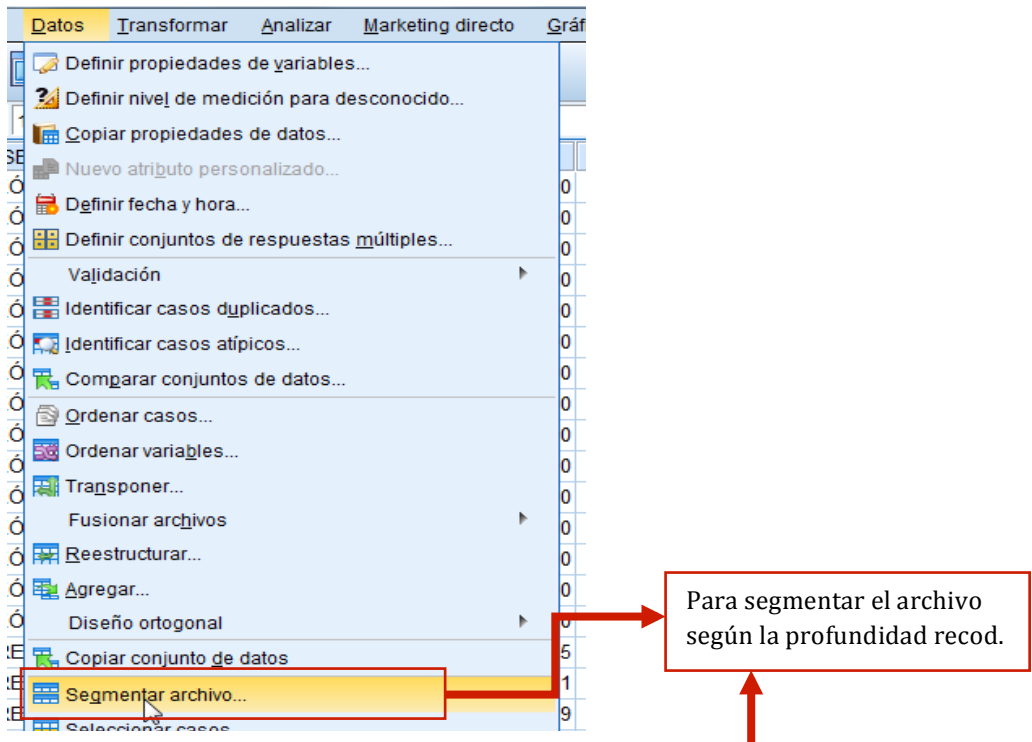
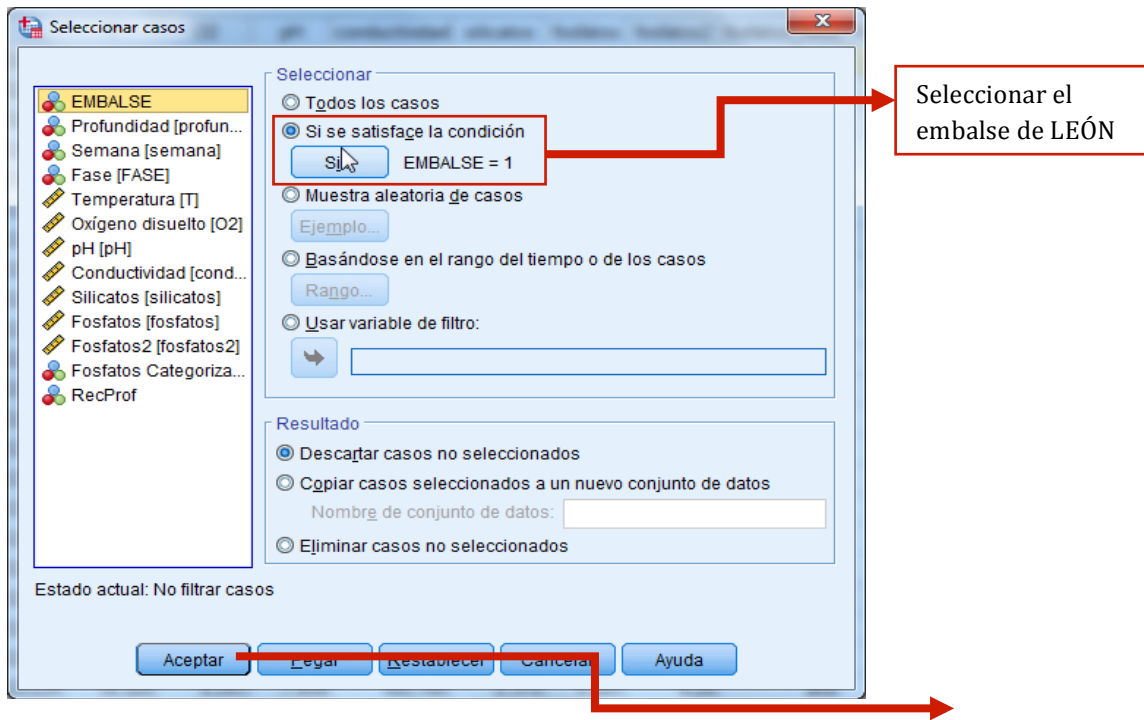
Ajuste, por tanto, para los datos de la fase de Estratificación², los modelos no lineales: Parabólico y Potencial. Compárelos y decida cuál de los dos es el mejor para explicar la relación entre las dos variables.

Trabaje con logaritmos neperianos, compare el poder predictivo, y el poder explicativo de cada modelo.

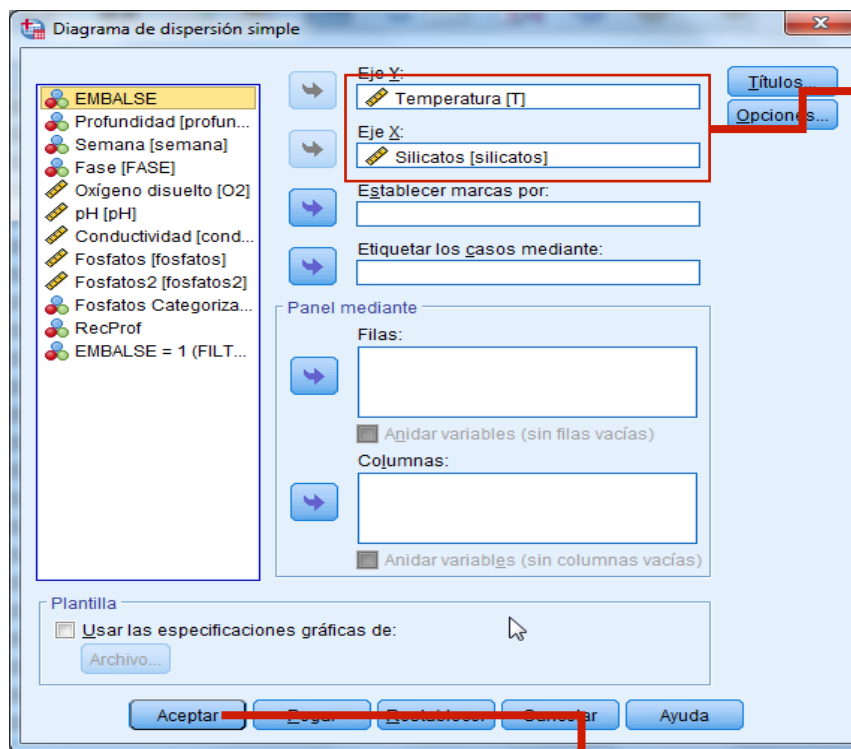
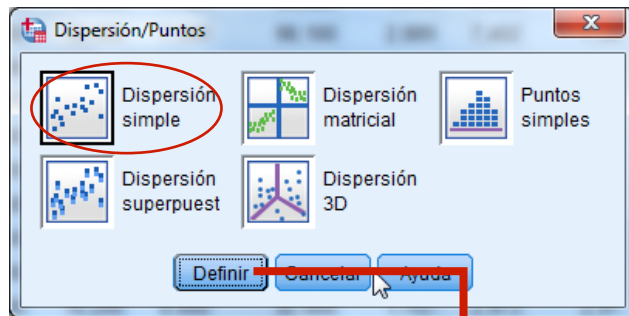
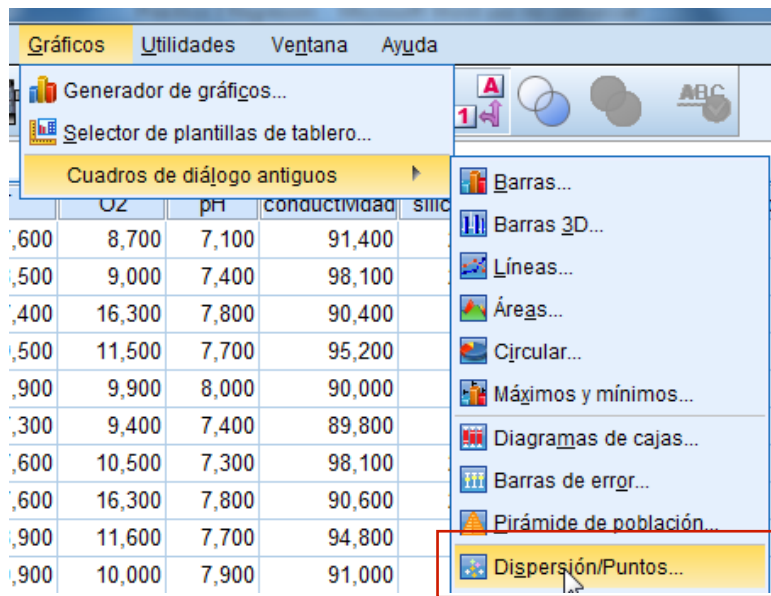
The image shows two screenshots from the SPSS software interface. The top screenshot displays the 'Transformar' (Transform) menu, with the option 'Recodificar en distintas variables...' (Recode into different variables...) highlighted in yellow. A red box and arrow point to this option with the text 'Para recodificar la variable ESTRATIFICACIÓN'. The bottom screenshot shows the 'Recodificar en distintas variables' dialog box. A red box highlights the 'Variable numérica -> Variable de resultado:' field, which contains 'profundidad --> RecProf', with an arrow pointing to it from the text 'Variables a recodificar y variable resultado'. Another red box highlights the 'Valores antiguos y nuevos...' button, with an arrow pointing to it from the text 'Para redefinir los valores'. The dialog box also shows a list of variables on the left, including 'SEMBALE', 'Semana [semana]', 'Fase [FASE]', 'Temperatura [T]', 'Oxígeno disuelto [O2]', 'pH [pH]', 'Conductividad [cond...]', 'Silicatos [silicatos]', 'Fosfatos [fosfatos]', 'Fosfatos2 [fosfatos2]', and 'Fosfatos Categoriza...'. At the bottom of the dialog are buttons for 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda'.

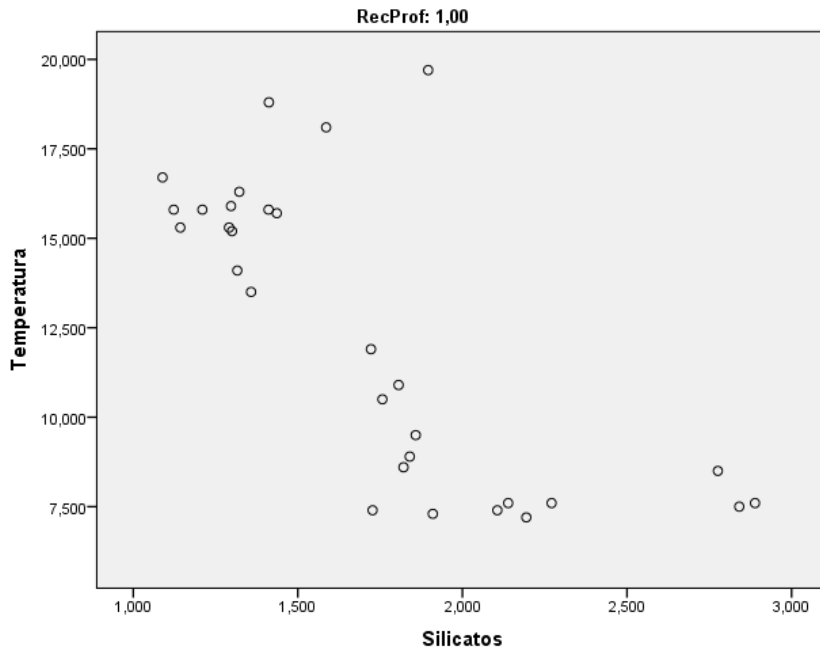
² Tenga en cuenta que en este apartado sólo debe trabajar con la fase de ESTRATIFICACIÓN y sólo en el embalse de LEÓN. Consulte la primera parte de las prácticas para ver cómo se seleccionan sólo algunos de los casos, y cómo segmentar (dividir) el archivo, pues necesitará realizar los dos procedimientos para el ajuste de la parábola. Tenga además en cuenta cómo están definidos en el SPSS los valores para cada categoría de cada variable.





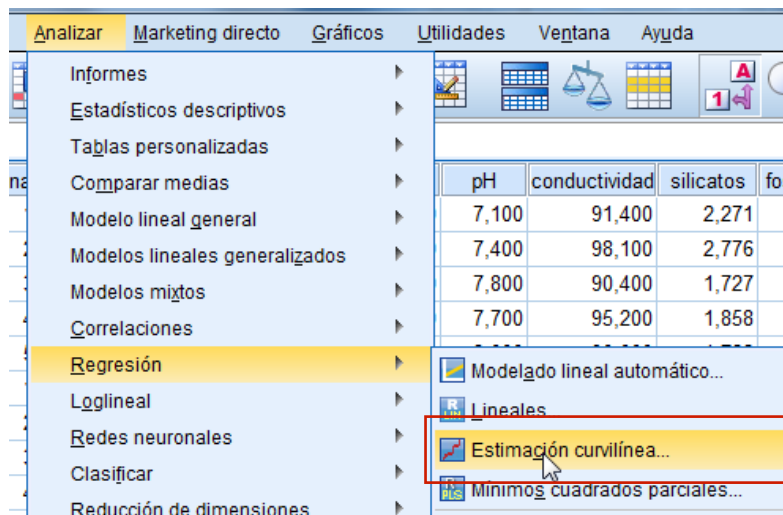
1. Represente el diagrama de dispersión correspondiente, y comente cómo podría ser el tipo de relación entre ambas variables.



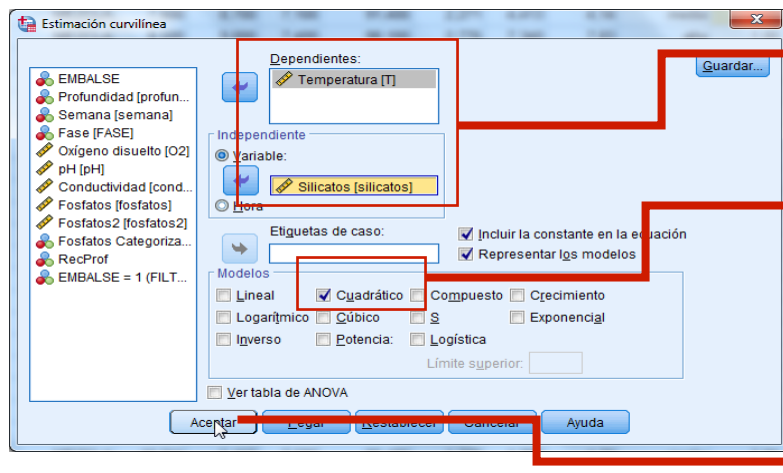


Un ajuste parabólico o potencial podría tener buenos resultados.

2. Ajuste un modelo de regresión parabólico. Escriba el modelo.



Para todo tipo de estimaciones no lineales



Variables

Modelo elegido

Resumen de modelo y estimaciones de parámetro^a

Variable dependiente: Temperatura

Ecuación	Resumen del modelo					Estimaciones de parámetro		
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2
Cuadrático	,599	20,207	2	27	,000	34,629	-19,189	3,367

La variable independiente es Silicatos.

a. RecProf = 1,00

Coeficiente de determinación

Coeficientes del modelo

La ecuación del modelo es:

$$\text{Silicato} = 34,629 - 19,189 \cdot \text{temperatura} + 3,367 \cdot \text{temperatura}^2$$

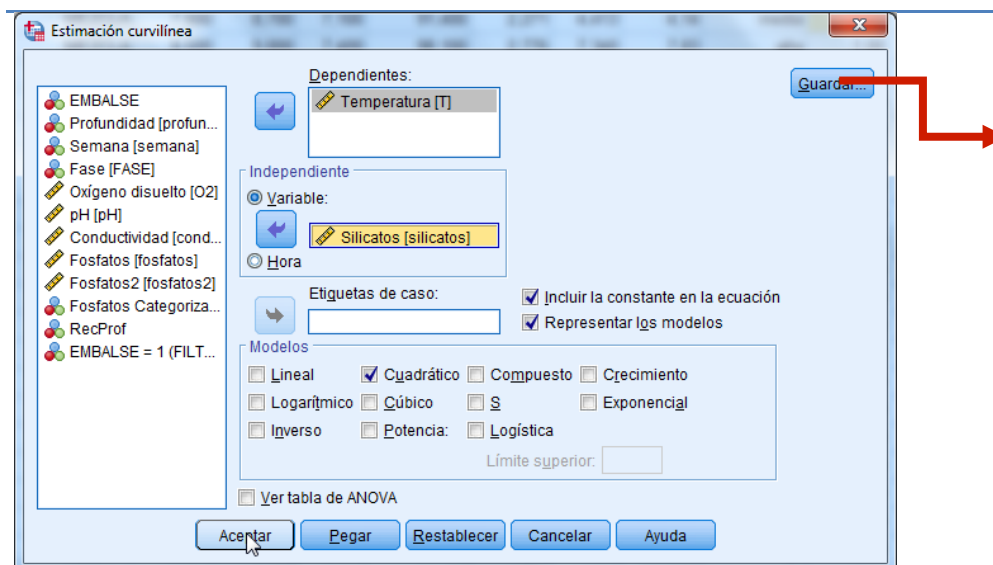
3. ¿Cuál es la magnitud de la relación?. ¿Por qué? ¿Qué coeficiente utiliza para responder?.

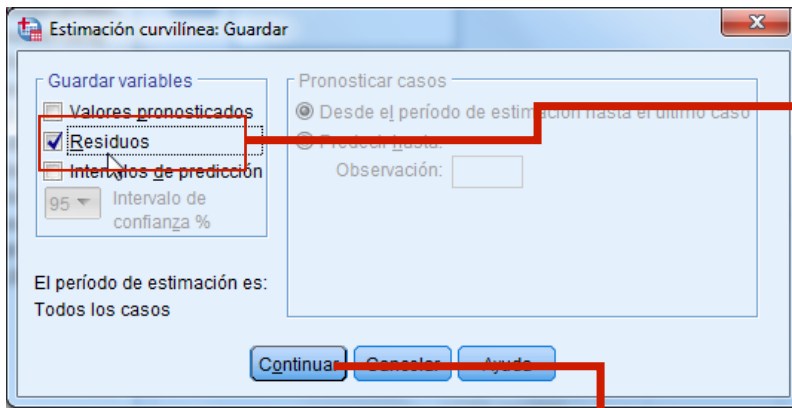
El grado de relación entre las variables es del 59,9%, valor del coeficiente de determinación, que nos indica el porcentaje de variabilidad de la y explicada por el modelo mediante las variaciones de la x.

4. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.

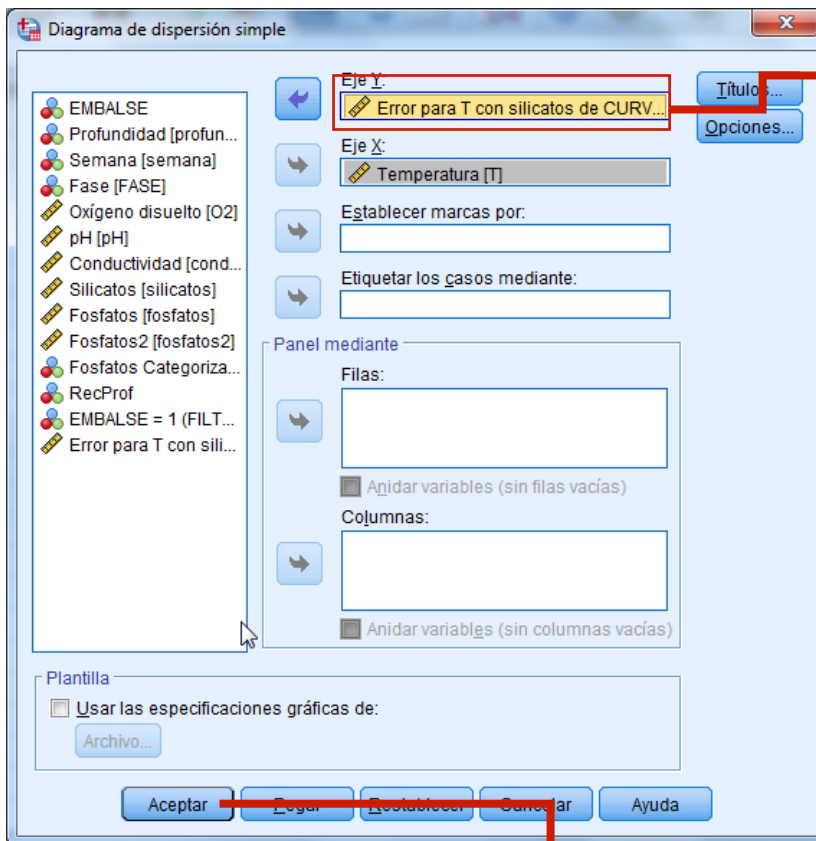
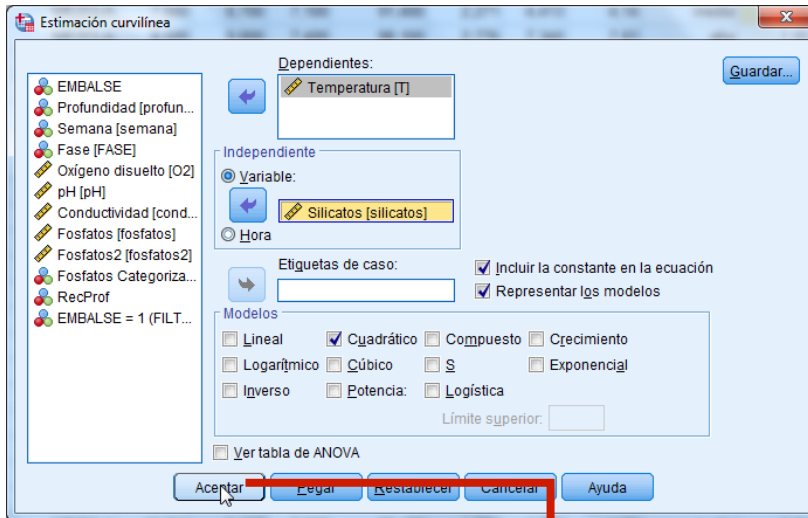
El modelo explica solo el 60% de las variaciones, por lo que no explica suficientemente los datos.

5. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.

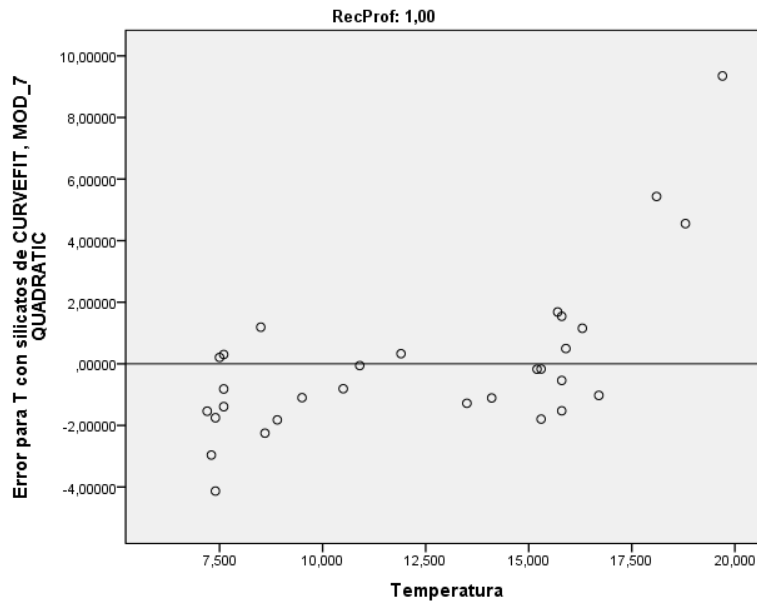




Permite guardar los residuos para el gráfico de residuales.

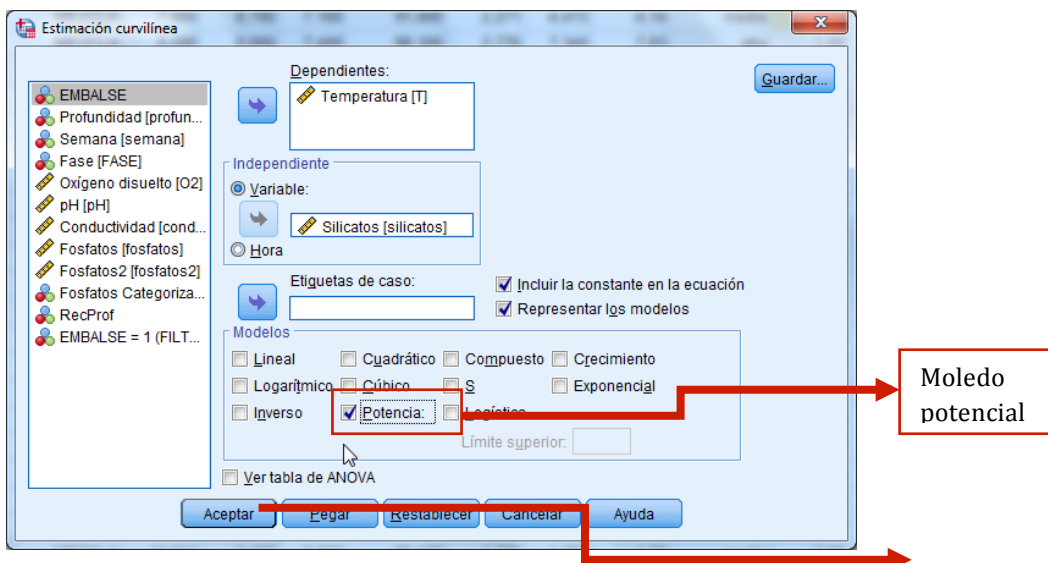
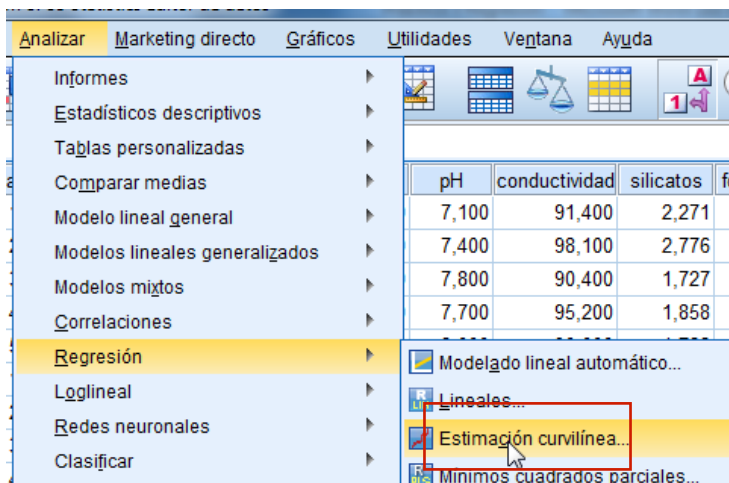


Variable de los residuos para el modelo cuadrático.



Este gráfico de residuales no presenta una forma determinada.

1. Ajuste un modelo de regresión potencial. Escriba el modelo.



Resumen de modelo y estimaciones de parámetro^a

Variable dependiente: Temperatura

Ecuación	Resumen del modelo					Estimaciones de parámetro	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Potencia	,630	47,594	1	28	,000	19,505	-1,013

La variable independiente es Silicatos.

a. RecProf = 1,00

Coefficiente de determinación

Coefficientes del modelo

La ecuación del modelo es: $Silicato = 19,505.Temperatura^{-1,013}$

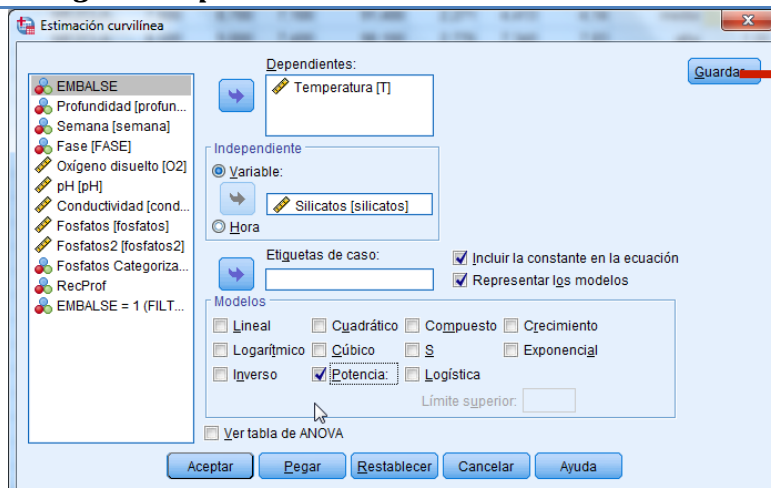
6. ¿Cuál es la magnitud de la relación?. ¿Por qué? ¿Qué coeficiente utiliza para responder?.

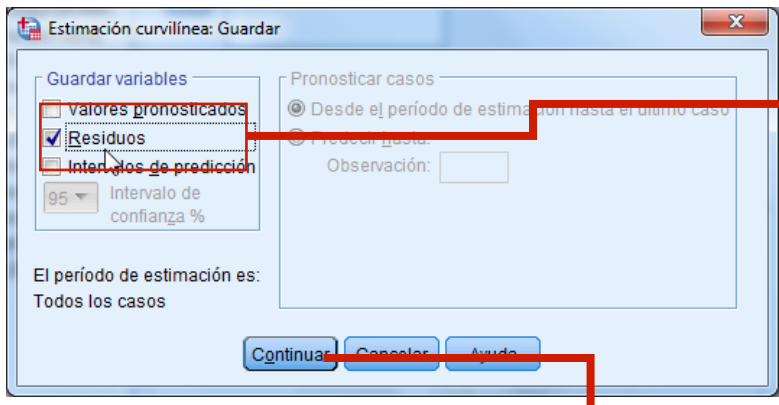
El grado de relación entre las variables es del 63%, valor del coeficiente de determinación, que nos indica el porcentaje de variabilidad de la y explicada por el modelo mediante las variaciones de la x.

7. Considera que el modelo es bueno desde el punto de vista explicativo. Justifique la respuesta.

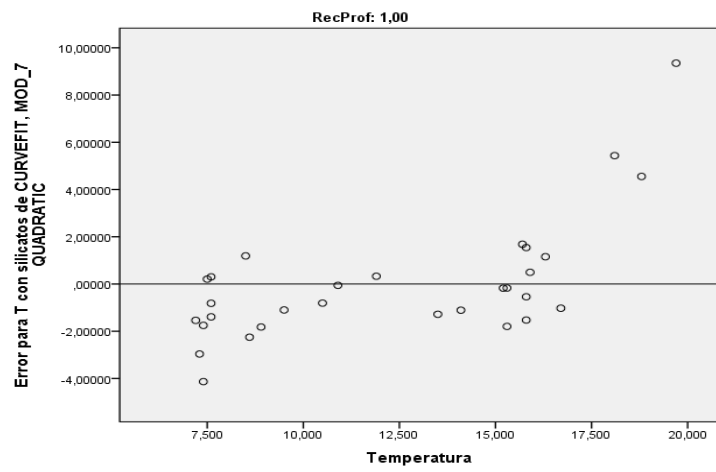
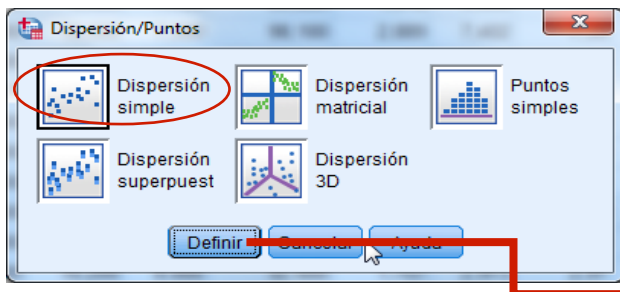
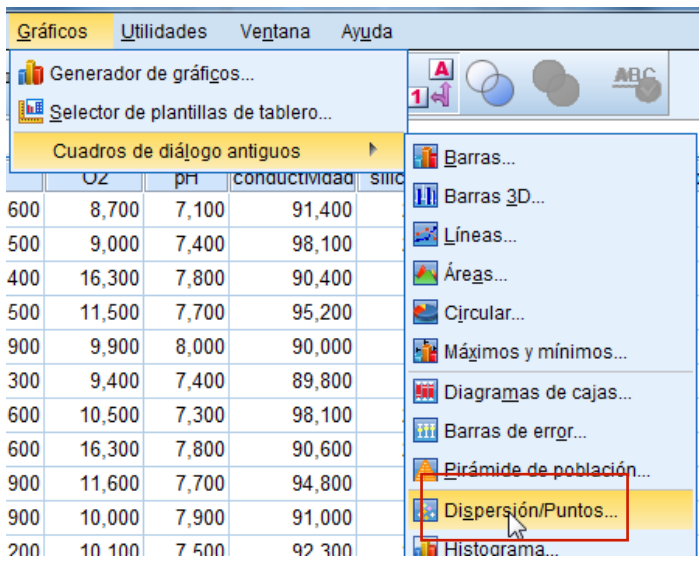
El poder explicativo del 63% es bajo, pero mejor que el de los anteriores modelos.

8. Represente el diagrama de residuos, representando los ZRES frente a la variable dependiente y analice los resultados. Realice una conclusión general del análisis de regresión que ha realizado.





Permite guardar los residuos para el gráfico de residuales.

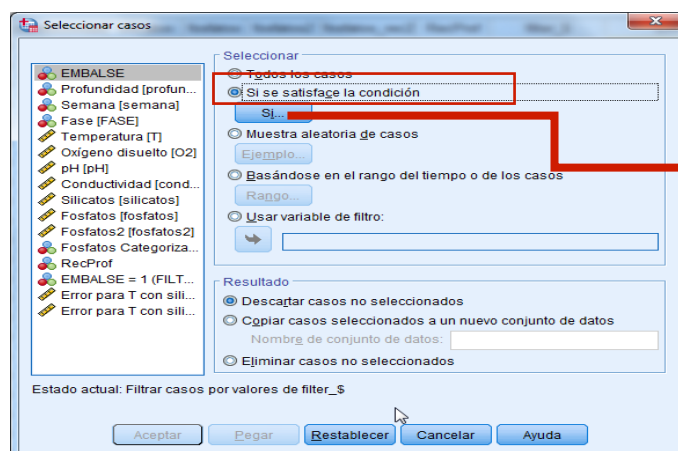
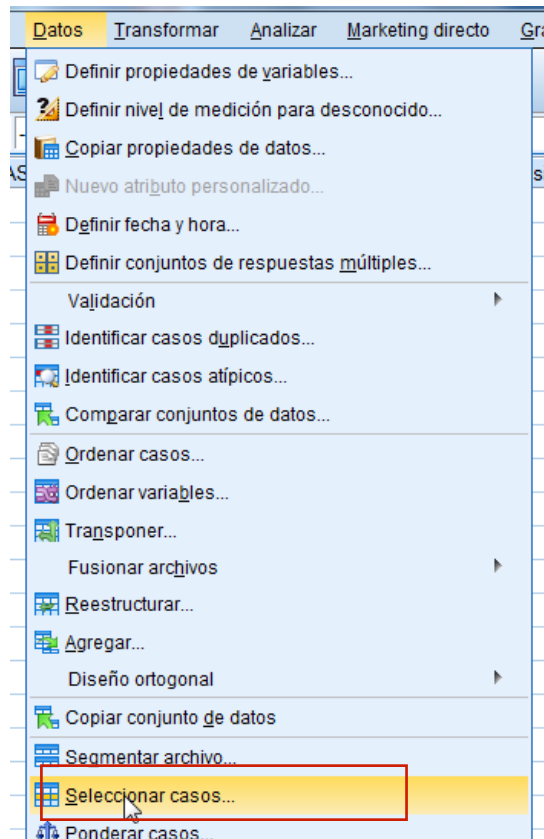


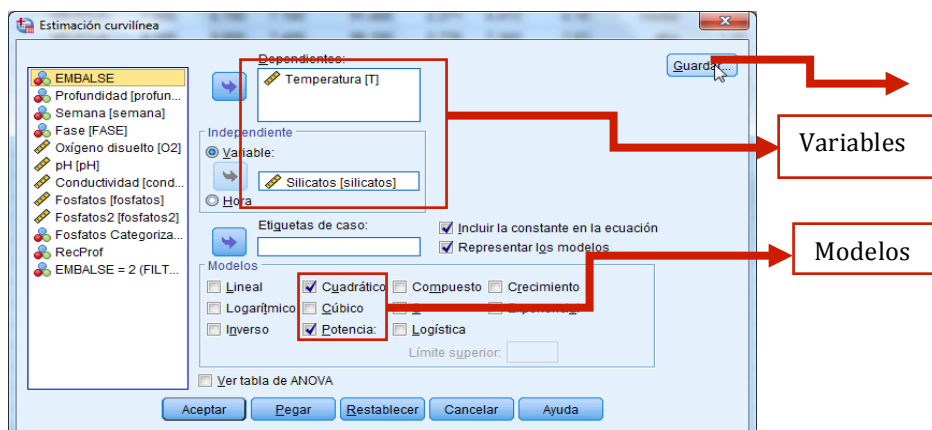
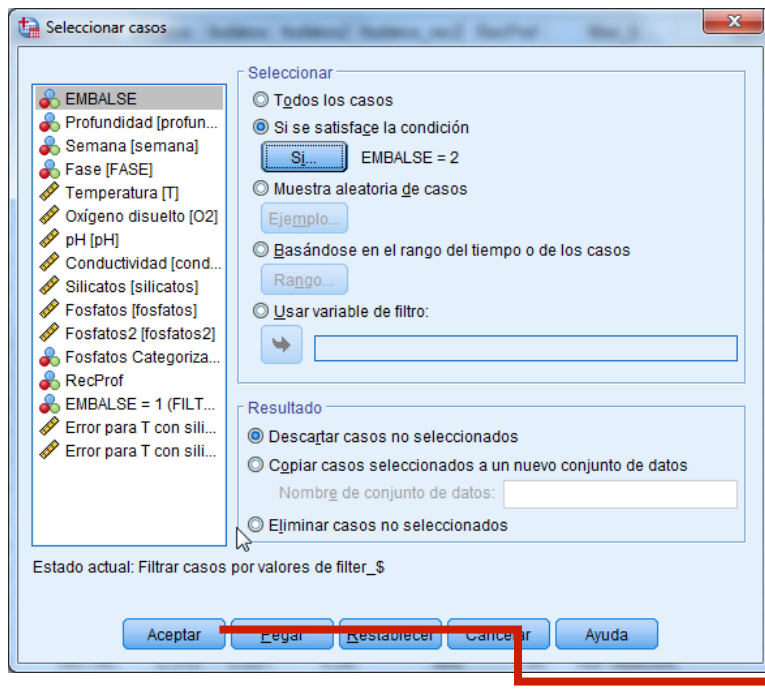
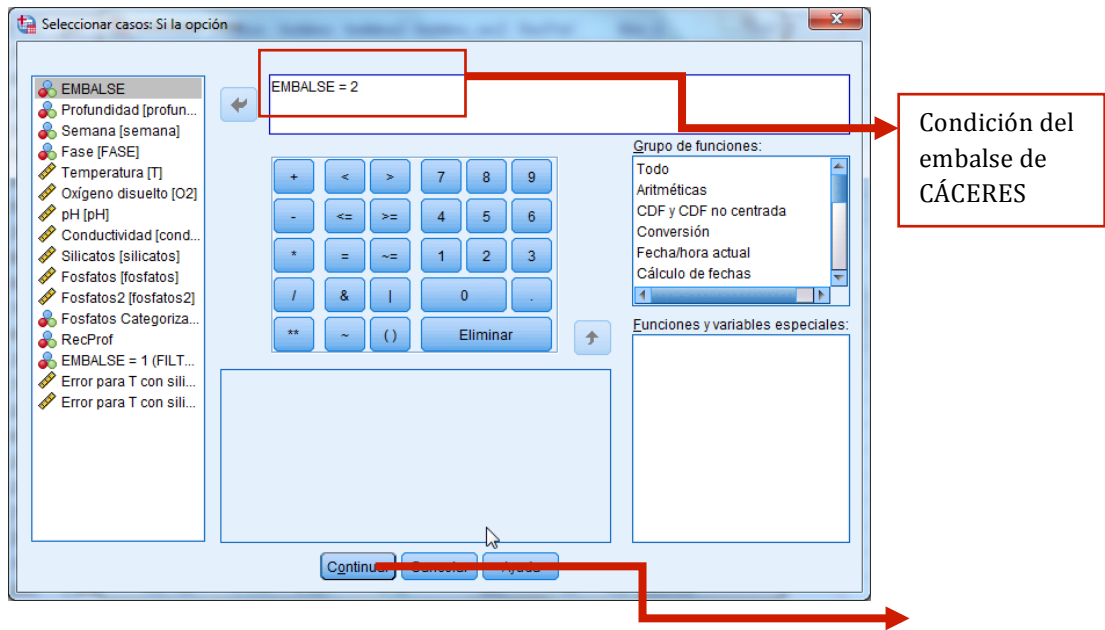
Este gráfico de residuales no tiene forma definida.

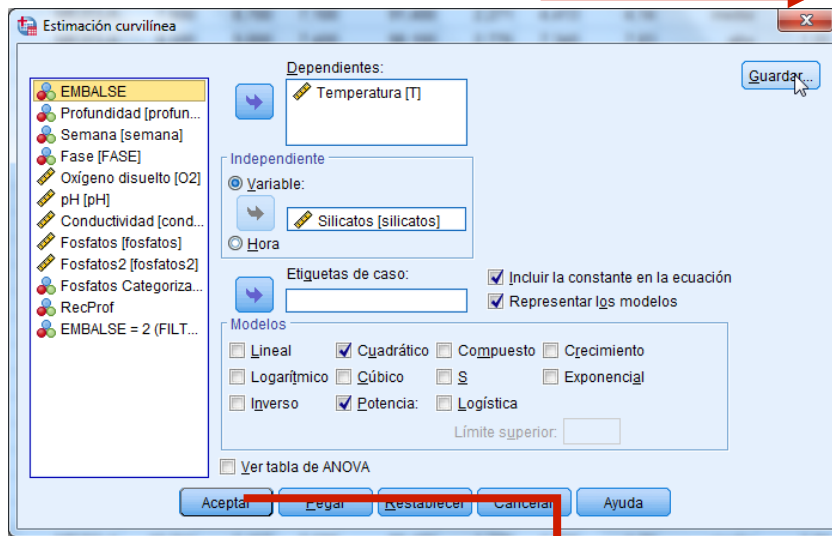
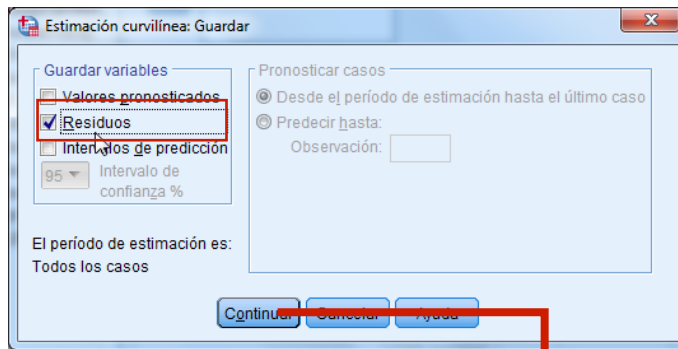
9. ¿Cuál de los dos modelos es el mejor para explicar la relación entre las dos variables.

El modelo potencial es el mejor porque es el que mayor poder explicativo tiene y sus gráficos de residuales no tienen forma definida.

- Realice ahora el ajuste también para el embalse de CÁCERES. ¿En qué embalse el ajuste es mejor. ¿Por qué?







Resumen de modelo y estimaciones de parámetro^a

Variable dependiente: Temperatura

Ecuación	Resumen del modelo					Estimaciones de parámetro		
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2
Cuadrático	,317	6,277	2	27	,006	23,304	-7,552	,960
Potencia	,357	15,550	1	28	,000	15,825	-,442	

La variable independiente es Silicatos.

a. RecProf = 1,00

Coefficientes de determinación de los modelos

Coefficientes de los modelos

A la vista de estos datos puede observarse que en el embalse de CÁCERES estos modelos tienen bastante bajo el poder explicativo, por lo que no procederemos a calcular los gráficos de residuales, pues no llega el poder explicativo ni a un 40%.

Práctica 4

INTERVALOS DE CONFIANZA Y CONTRASES DE HIPÓTESIS

OBJETIVOS y PREGUNTAS

1.- Construya el intervalo de confianza al 95% de confianza para la media poblacional variable Oxígeno disuelto.

1.1.- Analice en primer lugar si, efectivamente, se puede considerar que esta variable se distribuye según un modelo normal. Plantee las hipótesis adecuadas y concluya.

1.2.- Construya el intervalo de confianza e interprételo

1.3.- Escriba la expresión teórica del intervalo de confianza

1.4.- Construya el intervalo de confianza al 99% de confianza y compárelo con el anterior.

1.5.- Realice un gráfico en el que se representen los intervalos de confianza al 95% para la variable Temperatura a lo largo de las 10 semanas de estudio. Tenga en cuenta todos los datos.

**2.- Los investigadores afirman que, a nivel poblacional, la CONDUCTIVIDAD media en el embalse de CÁCERES es superior a 99,5 unidades.
Analice si dicha afirmación es cierta.**

2.1.- Analice si la variable CONDUCTIVIDAD en el embalse de CÁCERES se distribuye según un modelo normal. Utilice el test de Kolmogorov-Smirnov de la opción Estadísticos Descriptivos → Explorar

Para ello, plantee cuál es la hipótesis nula, la alternativa, y el p-valor correspondiente, así como la conclusión a la que llega.

2.2.- Realice el contraste de tendencia central adecuado.

Plantee la hipótesis nula y la hipótesis alternativa

Defina el tipo de contraste (unilateral/bilateral...)

Establezca el nivel de significación e interprete dicho valor.

Escriba la expresión del estadígrafo de contraste y cuál es su distribución teórica.

Escriba el valor experimental del estadígrafo de contraste

Escriba el p-valor, y represente gráficamente los resultados del contraste sobre la distribución teórica del estadígrafo

¿Cuál es la conclusión desde el punto de vista estadístico? .

¿Cuál es la conclusión experimental? .

3.- Analice si la variable OXÍGENO DISUELTO presenta un valor medio poblacional diferente en cada EMBALSE (seleccione solamente las profundidades de ≤ 15 m de la FASE DE ESTRATIFICACIÓN).

Explique, detalladamente todos los pasos de los tres contrastes que tiene que realizar para responder a esta pregunta:

4.- Teniendo en cuenta todos los datos del estudio ¿Se puede afirmar que la tendencia central de la variable Oxígeno disuelto es significativamente diferente entre los dos embalses?.

Explique, detalladamente todos los pasos de los tres contrastes que tiene que realizar para responder a esta pregunta: Contraste de Normalidad, de Homocedasticidad y de Tendencia Central.

5.- Analice si, a nivel poblacional, existen diferencias en tendencia central entre las mediciones de los fosfatos por ambos métodos (fosfatos 1 y fosfatos 2) (Trabaje sólo con los datos pertenecientes a la fase de MEZCLA).

5.1.- Realice los contrastes de normalidad adecuados:

5.2.-Realice el contraste de tendencia central adecuado

6.-Analice si, a nivel poblacional, existen diferencias en tendencia central entre las mediciones de los fosfatos por ambos métodos (fosfatos 1 y fosfatos 2) (Trabaje sólo con los datos pertenecientes a la fase de ESTRATIFICACIÓN).

6.1.- Realice los contrastes de normalidad adecuados:

6.2.-Realice el contraste de tendencia central adecuado

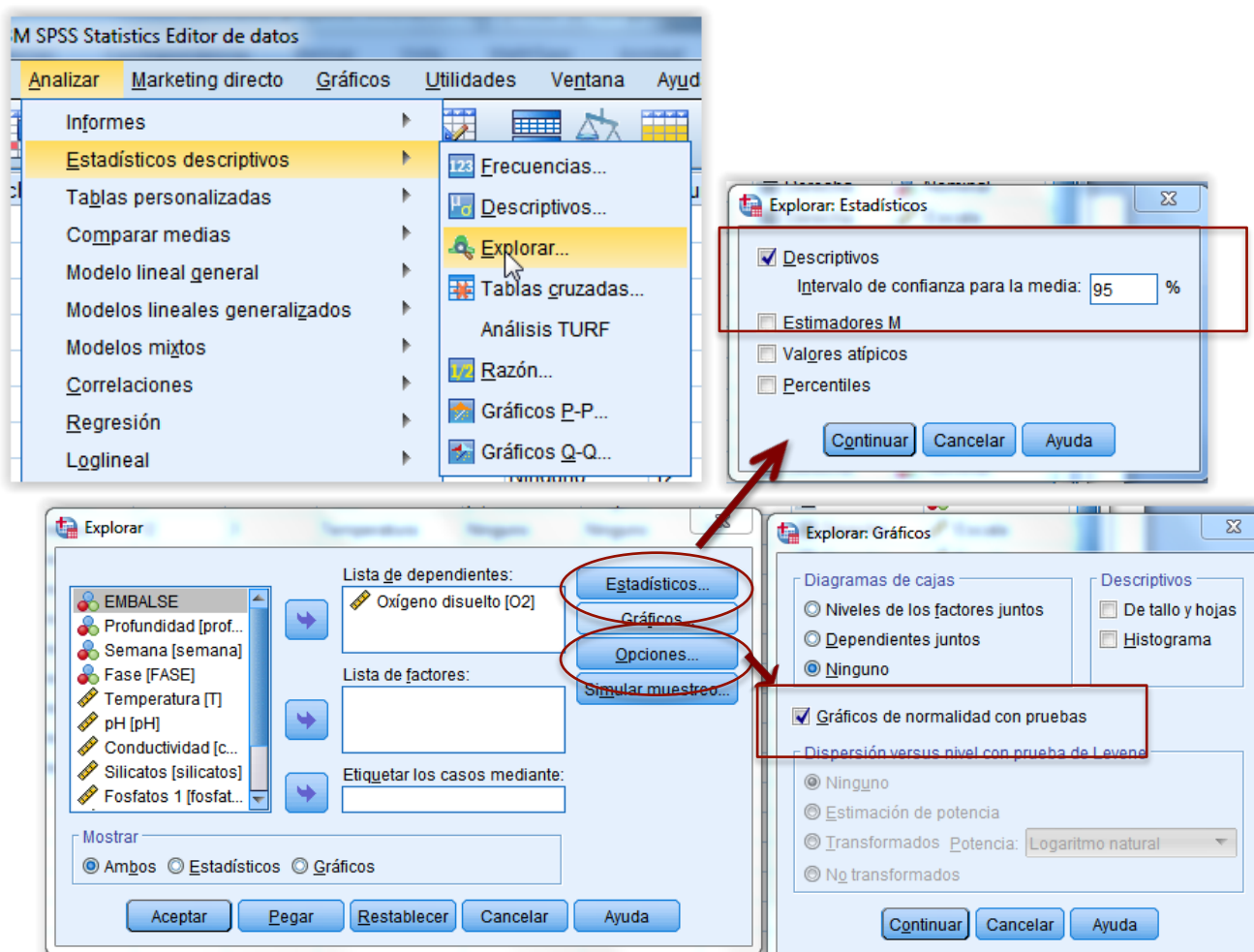
Práctica 5

INTERVALOS DE CONFIANZA Y CONTRASES DE HIPÓTESIS

SOLUCIÓN

1.- Construya el intervalo de confianza al 95% de confianza para la media poblacional variable Oxígeno disuelto.

1.1.- Analice en primer lugar si, efectivamente se puede considerar que esta variable se distribuye según un modelo normal. Plantee las hipótesis adecuadas y concluya.



H_0 : La variable Oxígeno Disuelto se distribuye según un modelo Normal

H_a : La variable Oxígeno Disuelto NO se distribuye según un modelo Normal

Nivel de significación: $\alpha = 0.05$ y $\alpha = 0.01$

Los resultados obtenidos con el SPSS, para este contraste, se muestran en la siguiente tabla, ya que se ha seleccionado en el botón Opciones, la opción Gráficos de normalidad con pruebas.

Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
Oxígeno disuelto	,068	140	,200*

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Como el p-valor es mayor que $\alpha = 0.05$ podemos concluir que no encontramos razones para rechazar la hipótesis nula, los resultados del contraste estadístico son no significativos; por tanto, podemos asumir que la distribución de la variable Oxígeno Disuelto se distribuye según un modelo Normal.

1.2.- Construya el intervalo de confianza e interprételo

Ahora se muestra la tabla de resultados en donde aparece el intervalo de confianza al 95%. (Obsérvese que a la hora de realizar el análisis en explorar, se ha dejado en el botón Estadísticos, la opción que viene por defecto: Intervalo de confianza para la media: 95%.

Descriptivos

		Estadístico	Error estándar
Oxígeno disuelto	Media	9,119	,226
	95% de intervalo de confianza para la media		
	Límite inferior	8,671	
	Límite superior	9,567	
	Media recortada al 5%	9,026	
	Mediana	9,259	
	Varianza	7,176	
	Desviación estándar	2,679	
	Mínimo	2,854	
	Máximo	16,300	
	Rango	13,446	
	Rango intercuartil	3,175	
	Asimetría	,400	,205
	Curtosis	,530	,407

Por tanto el Intervalo de confianza para la media poblacional de la variable Oxígeno disuelto será: $I = [8,671 ; 9,567]$. La interpretación sería: Tenemos una probabilidad del 95% de que este intervalo sea de los que si contiene a la media poblacional. O dicho de otra forma tenemos una confianza del 95% de que este intervalo sea de los que sí contiene a la media poblacional.

1.3.- Escriba la expresión teórica del intervalo de confianza

Teóricamente deberíamos utilizar la siguiente expresión: $I_{\mu}^{0,95} = \left[\bar{x} \pm z_{\alpha/2} \frac{\hat{S}}{\sqrt{n}} \right]$. Sin

embargo el SPSS utiliza: $I_{\mu}^{0,95} = \left[\bar{x} \pm t_{\alpha; n-1} \frac{\hat{S}}{\sqrt{n}} \right]$

1.4.- Construya el intervalo de confianza al 99% de confianza y compárelo con el anterior.

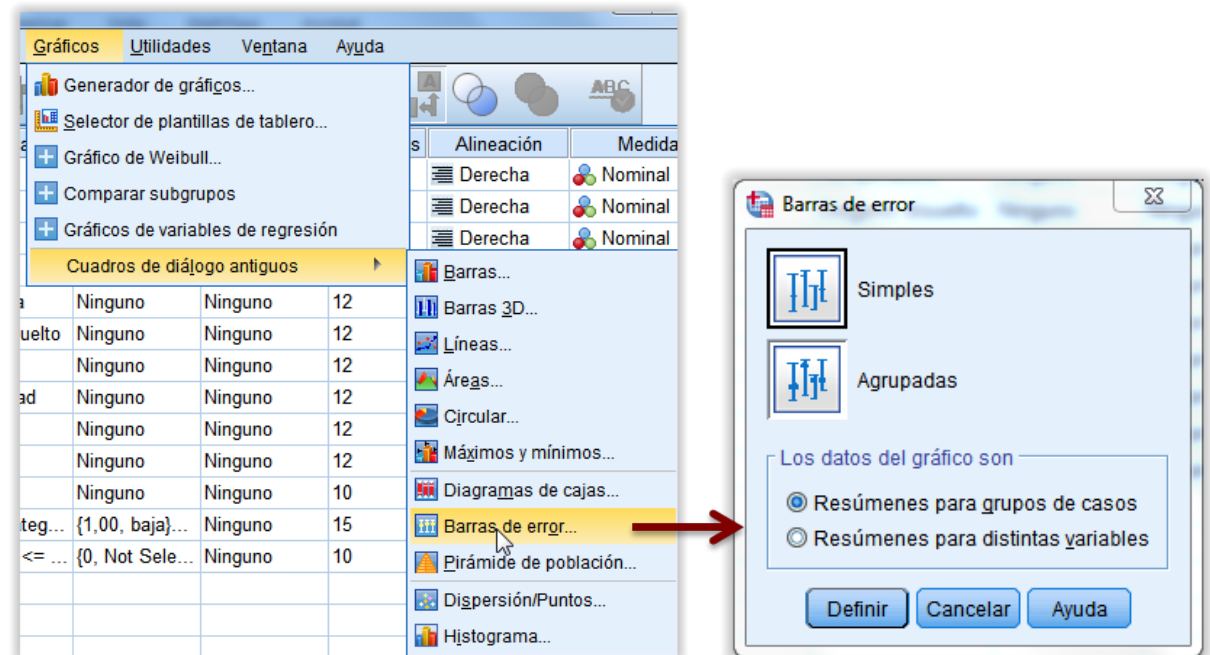
Se repite el análisis de la misma forma pero, en las ventanas de selección tenemos que cambiar dentro del botón Estadísticos, donde pone intervalo de confianza para la media y poner 99%

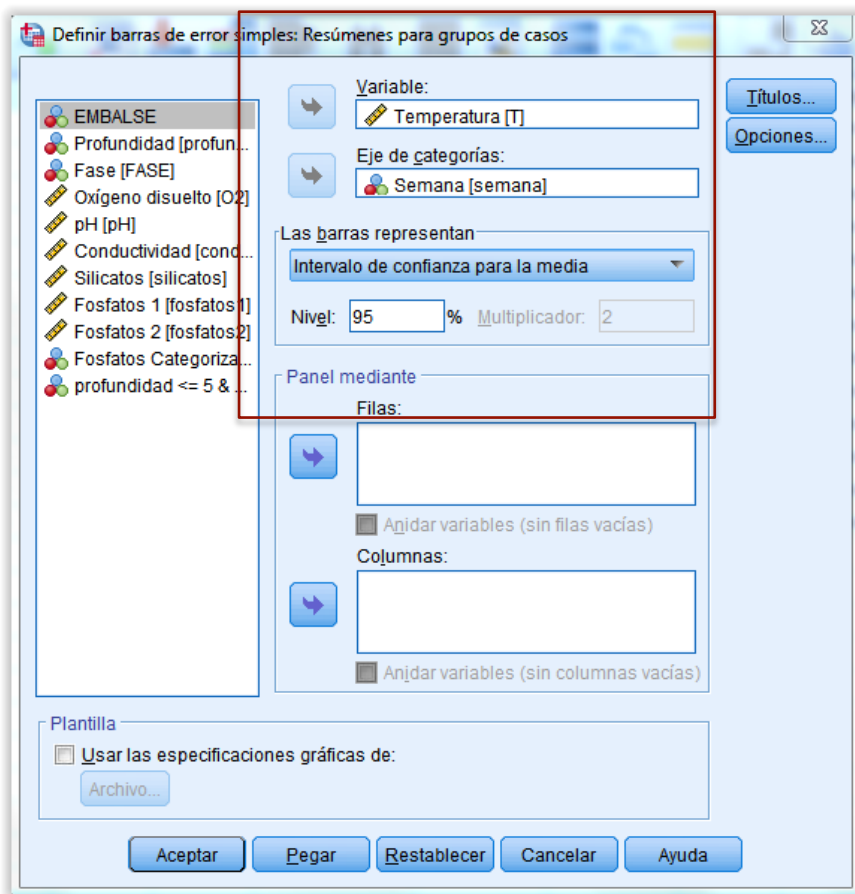
Descriptivos		Estadístico	Error estándar
Oxígeno disuelto	Media	9,119	,226
	99% de intervalo de confianza para la media	Límite inferior	8,528
		Límite superior	9,710

Por tanto tendremos: $I = [8.528 ; 9.710]$. Se observa que este intervalo presenta mayor amplitud que el anterior, pues se ha aumentado el nivel de confianza.

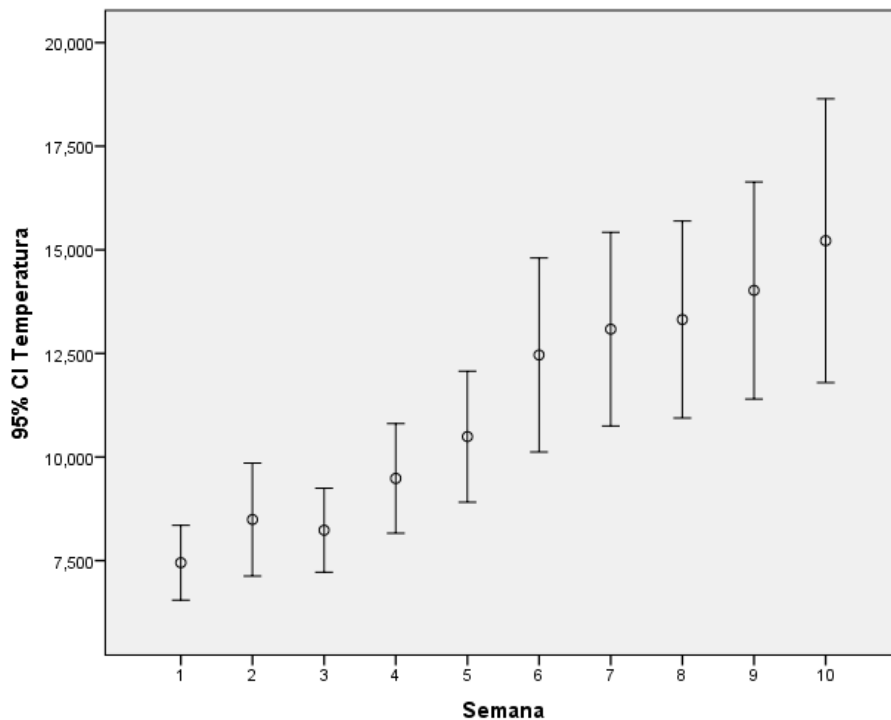
Tenemos una confianza del 99% de que este intervalo sea de los que contienen a la media poblacional de la variable Oxígeno Disuelto.

1.5.- Realice un gráfico en el que se representen los intervalos de confianza al 95% para la variable Temperatura a lo largo de las 10 semanas de estudio. Tenga en cuenta todos los datos.





El gráfico resultante se muestra a continuación. Vemos cómo con el paso del tiempo se va incrementando el valor medio de temperatura, y cómo hay intervalos de confianza que no se solapan y otros que sí.



Por ejemplo, los intervalos de las semanas 1,2,3 y 4 se solapan entre ellos por lo que se podría concluir que la temperatura media poblacional no difiere en esas semanas. Sin

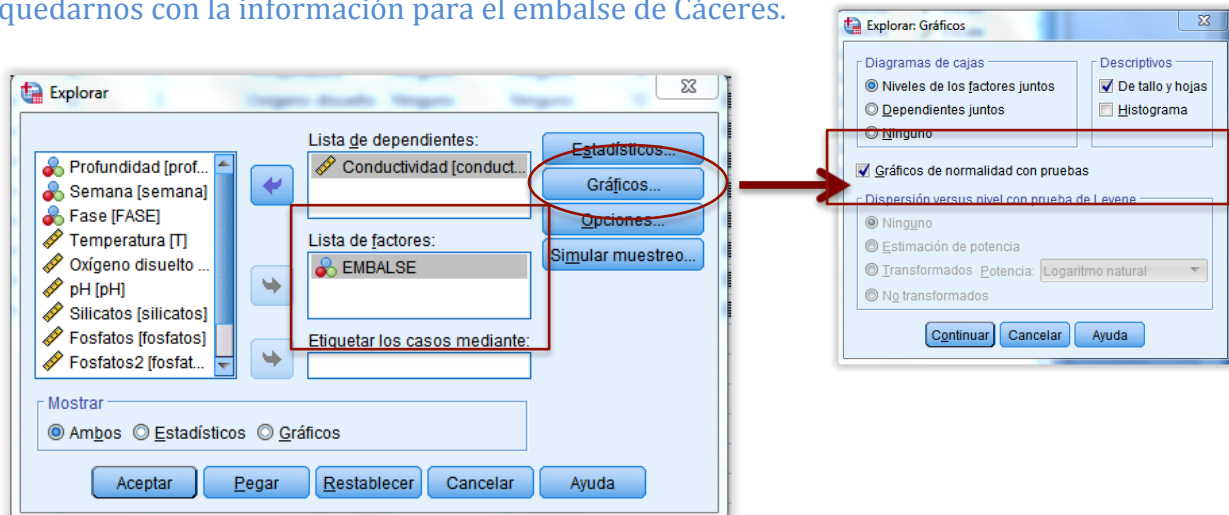
embargo, vemos cómo los intervalos de esas semanas no se solapan con la semana 10, con lo que podría afirmarse que la temperatura media en esta última semana es diferente (más alta) que en las 4 primeras semanas (a nivel poblacional).

2.- Los investigadores afirman que, a nivel poblacional, la CONDUCTIVIDAD media en el embalse de CÁCERES es superior a 99,5 unidades.

Analice si dicha afirmación es cierta.

2.1.-Analice si la variable CONDUCTIVIDAD en el embalse de CÁCERES se distribuye según un modelo normal. Utilice el test de Kolmogorov-Smirnov de la opción Estadísticos Descriptivos → Explorar

Para realizar este contraste tenemos que marcar las opciones que se indican en la siguiente figura, ya que queremos que nos haga el análisis por separado para cada embalse, y quedarnos con la información para el embalse de Cáceres.



Para ello, plantee cuál es la hipótesis nula, la alternativa, y el p-valor correspondiente, así como la conclusión a la que llega.

Ho: La variable conductividad se distribuye (en el embalse de Cáceres) según un modelo Normal

Ha: La variable conductividad NO se distribuye (en el embalse de Cáceres) según un modelo Normal

Los resultados obtenidos con el SPSS se muestran en la siguiente tabla:

Pruebas de normalidad				
		Kolmogorov-Smirnov ^a		
	EMBALSE	Estadístico	gl	Sig.
Conductividad	LEÓN	,135	70	,003
	CÁCERES	,063	70	,200*

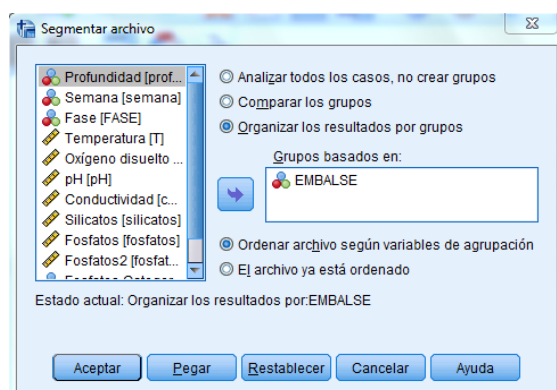
*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Como el p-valor es mayor que $\alpha = 0.05$ podemos concluir que no encontramos razones para rechazar la hipótesis nula, los resultados del contraste estadístico son no significativos; por tanto, podemos asumir que, en el embalse de Cáceres, la distribución de la variable Conductividad se distribuye según un modelo Normal

2.2.- Realice el contraste de tendencia central adecuado.

NOTA: Como tenemos que trabajar sólo con los datos correspondientes al embalse de Cáceres, deberemos ir a segmentar archivo, o seleccionar casos. En este caso mostraremos la opción Dividir archivo, teniendo en cuenta que sólo nos fijaremos en los resultados correspondientes al embalse de Cáceres:



Plantee la hipótesis nula y la hipótesis alternativa

$$H_0: \mu = 99.5$$

$$H_a: \mu > 99.5$$

Defina el tipo de contraste (unilateral/bilateral...)

Contraste Unilateral superior

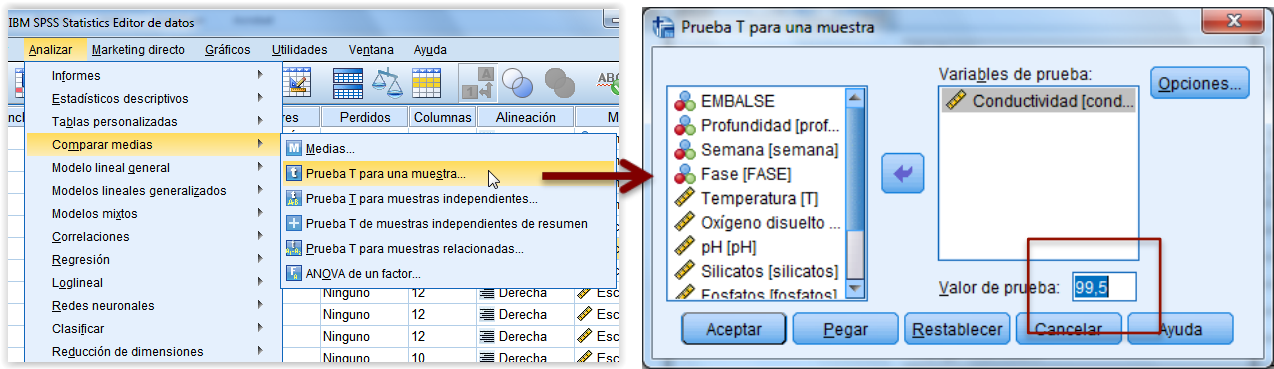
Establezca el nivel de significación e interprete dicho valor.

$$\alpha = 0.05 \text{ y } \alpha = 0.01$$

El nivel de significación es la probabilidad de cometer un error tipo I, es decir, la probabilidad máxima que estamos dispuestos a asumir al equivocarnos rechazando indebidamente la hipótesis nula.

Una vez divido el archivo para que nos haga el análisis correspondiente para cada embalse por separado, vamos a la opción: Comparar medias → Prueba T para una muestra.

Deberemos hacer la selección del análisis tal como aparece en la siguiente captura de pantalla. Se selecciona la variable de la que queremos hacer el contraste, y es MUY IMPORTANTE no olvidar de poner en valor de prueba el **valor de la hipótesis nula 99,5**.



Escriba la expresión del estadígrafo de contraste y cuál es su distribución teórica.

El tamaño de la muestra es de 70, por tanto aunque se desconozca la variabilidad poblacional (σ) el estadístico de contraste se distribuye según un modelo Normal Estándar. El SPSS siempre utiliza el estadígrafo t-Student.

Si lo hiciésemos a mano	Con el SPSS
$Z = \frac{\bar{x} - 99,5}{\hat{s}/\sqrt{n}}$	$t = \frac{\bar{x} - 99,5}{\hat{s}/\sqrt{n}}$

Tablas con los resultados del contraste de hipótesis:

Prueba de muestra única^a

Valor de prueba = 99.5						
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Conductividad	2,247	69	,028	1,508981	,16909	2,84888

a. EMBALSE = CÁCERES

El SPSS también proporciona en los resultados una tabla con los estadísticos descriptivos de la variable conductividad para el embalse de Cáceres: aparecen el tamaño de muestra (N), la media muestral, la desviación típica y el error estándar de la media.

Estadísticas de muestra única^a

	N	Media	Desviación estándar	Error estándar de la media
Conductividad	70	101,00898	5,619386	,671645

a. EMBALSE = CÁCERES

Observar, cómo en la parte inferior de las tablas nos recuerda que estos resultados son para el embalse de Cáceres.

Escriba el valor experimental del estadígrafo de contraste

El valor experimental es: $t_{\text{exp.}} = 2,247$

En la tabla anterior, nos muestra: el valor t-experimental (2,247), los grados de libertad ($69 = n-1$), el p-valor si el contraste fuese bilateral, y donde pone diferencias de medias es la diferencia entre el valor de la media muestral (101,00898) menos el valor de la hipótesis nula (99,5).

El intervalo de confianza al 95% que se presenta en esta tabla, hace referencia a un intervalo de confianza para la diferencia entre μ y μ_0 . Como podemos observar dicho intervalo de confianza no contiene al 0 y son valores positivos.

Escriba el p-valor, y represente gráficamente los resultados del contraste sobre la distribución teórica del estadígrafo

El SPSS proporciona el p-valor para un contraste bilateral. Por ello, para este caso, en el que el contraste es unilateral, deberemos dividir el p-valor que resulta en el análisis (sig. (bilateral) 0.028) entre 2, para obtener el verdadero p-valor del contraste, este será por tanto:

$$\mathbf{p - valor = 0.028/2 = 0.014}$$

¿Cuál es la conclusión desde el punto de vista estadístico? .

Desde el punto de vista estadístico los resultados son probablemente significativos, ya que el p-valor está comprendido entre 0,05 y 0,01. Eso es equivalente a que el valor experimental del estadígrafo de contraste, como puede verse en la figura del punto 6, se encuentra en la región crítica a un nivel de significación del 5%, y en la región de aceptación a un nivel de significación del 1%.

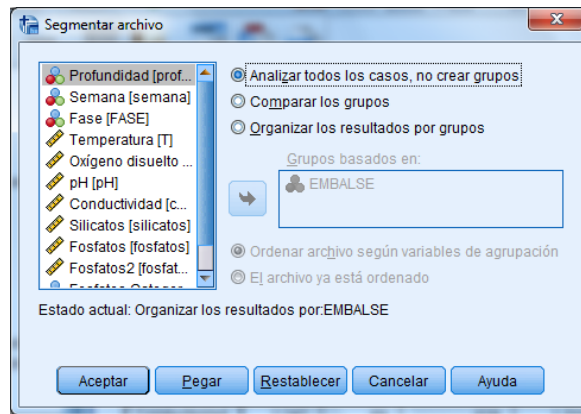
¿Cuál es la conclusión experimental?.

La conclusión experimental es que, efectivamente, ocurre lo que piensan los investigadores, que a nivel poblacional en el embalse de Cáceres, la media de la variable conductividad es mayor de 99,5 unidades.

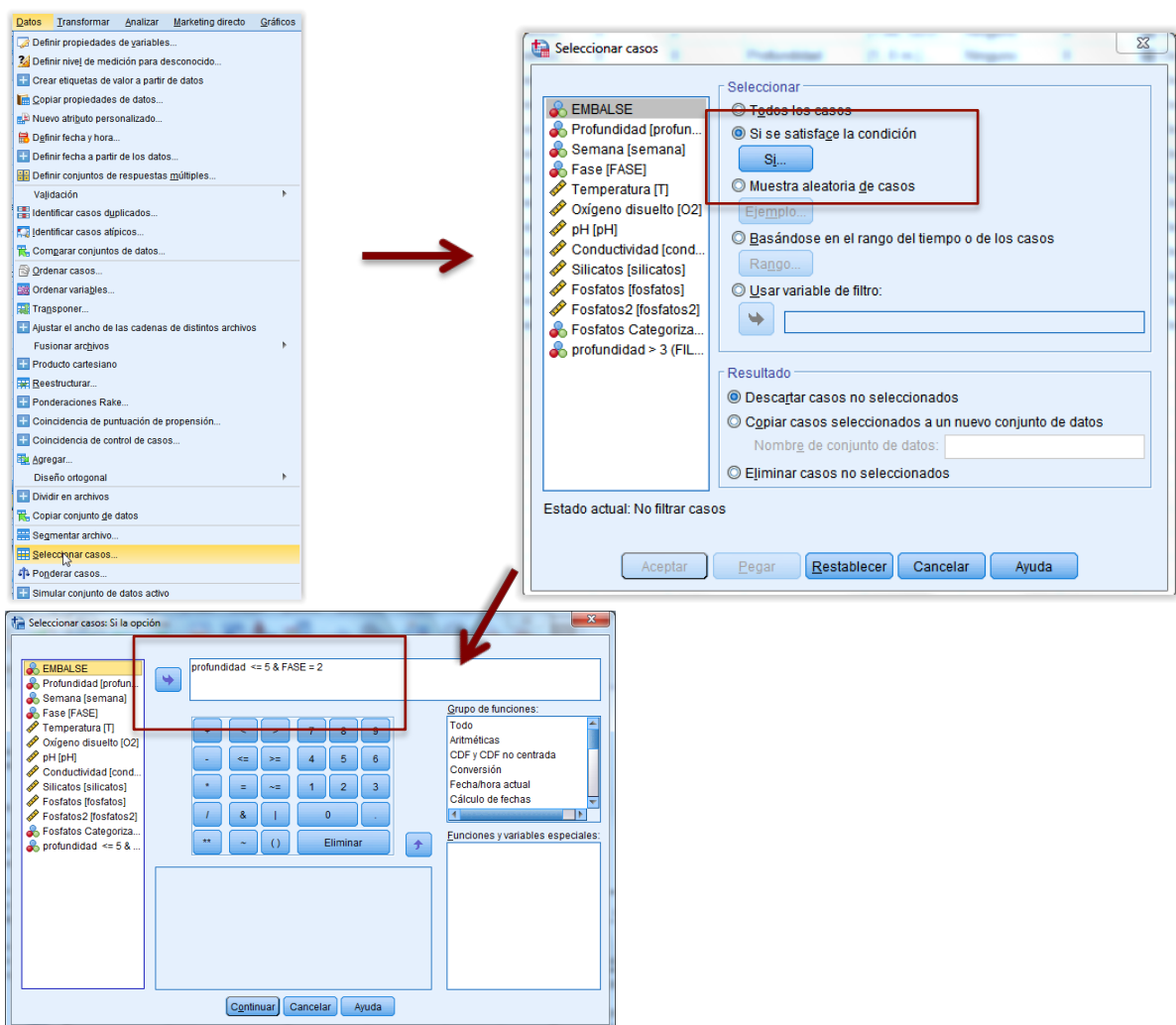
3.- Analice si la variable OXÍGENO DISUELTO presenta un valor medio poblacional diferente en cada EMBALSE (seleccione solamente las profundidades de ≤ 15 m de la FASE DE ESTRATIFICACIÓN.

Se trata de un contraste de tendencia central para dos **poblaciones independientes**.

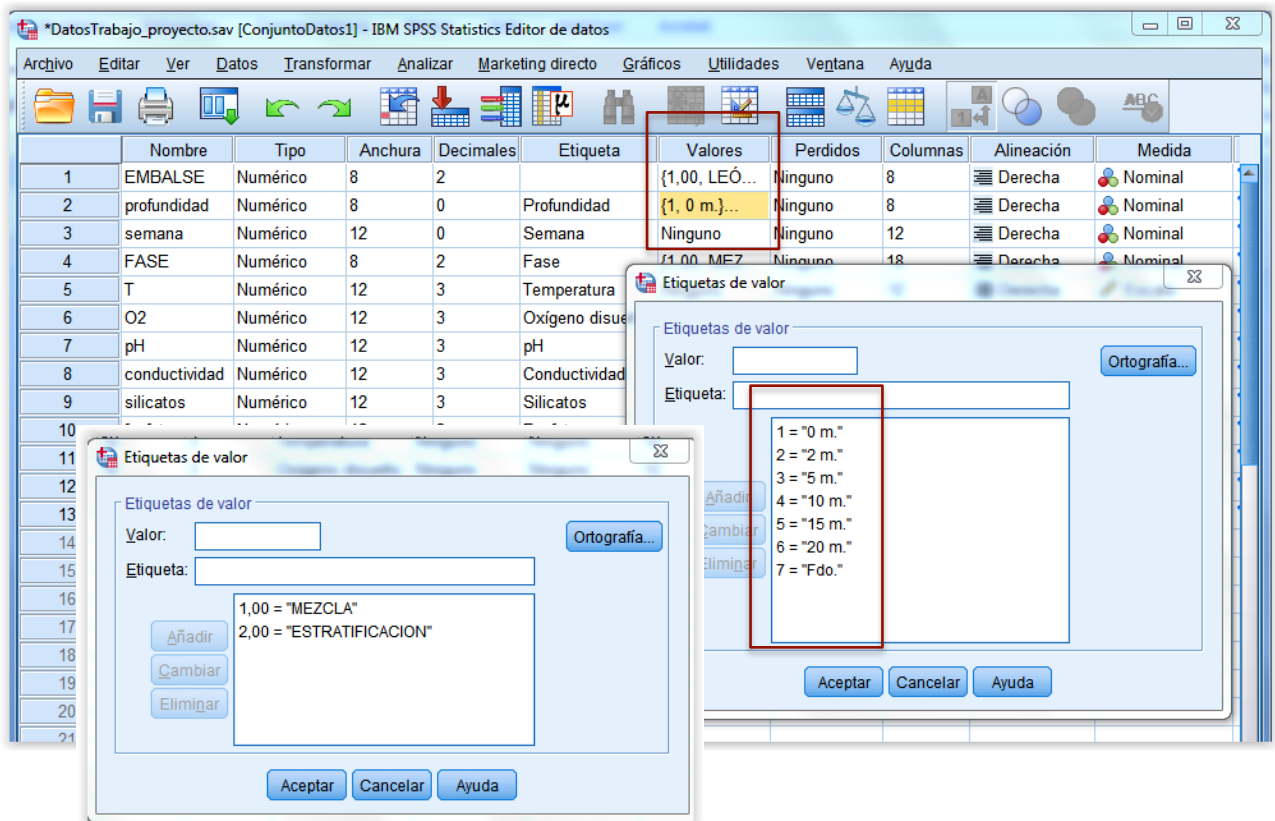
Antes de comenzar con este apartado, debemos volver a tener todo el archivo de datos junto. En el apartado anterior lo habíamos dividido por embalses para ver los resultados de cada uno por separado. Por tanto, volvemos a segmentar archivo y decimos que queremos trabajar con todos los datos:



Ahora deberemos ir a la opción de seleccionar casos, para quedarnos sólo con aquellas profundidades de ≤ 15 m. de la fase de ESTRATIFICACIÓN:



Como puede observarse en la captura de pantalla anterior se ha escrito **profundidad ≤ 5 & FASE = 2**. Si observamos en la ventana de características de variables, cuando observamos las categorías de la variable **profundidad**, de ≤ 15 metros tienen asignados **valores** en la tabla de datos del SPSS ≤ 5 ; y que la categoría de la variable FASE correspondiente a la estratificación, le corresponde el **valor = 2**



Explique, detalladamente todos los pasos de los tres contrastes que tiene que realizar para responder a esta pregunta:

Para responder a esta pregunta necesitamos hacer tres contrastes, pues necesitamos decidir cuál es el estadígrafo de contraste, y la elección dependerá de los resultados de los dos primeros contrastes:

- 1.- Un contraste de normalidad para la variable Óxígeno disuelto en cada embalse.
- 2.- Un contraste de Homocedasticidad.
- 3.- El contraste de tendencia central correspondiente.

1.- CONTRASTES de NORMALIDAD (utilice el test de Kolmogorov-Smirnov de la opción Explorar)

Este contraste se realiza en la opción del menú explorar tal como se hizo en el primer apartado de la práctica.

Los resultados para cada embalse se muestran en la siguiente tabla:

Pruebas de normalidad				
	EMBALSE	Kolmogorov-Smirnov ^a		
		Estadístico	gl	Sig.
Oxígeno disuelto	LEÓN	,131	25	,200*
	CÁCERES	,148	25	,161

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Ho: La variable Oxígeno disuelto en el embalse de León, se distribuye según un modelo Normal

Ha: La variable Oxígeno disuelto en el embalse de León, NO se distribuye según un modelo Normal

Valor experimental del estadígrafo de contraste: 0,131

p-valor: 0,200

Conclusión estadística y experimental.

Los resultados son No Significativos, por tanto no podemos rechazar la Ho. La variable Oxígeno disuelto en el embalse de León, se distribuye según un modelo Normal

Ho: La variable Oxígeno disuelto en el embalse de Cáceres, se distribuye según un modelo Normal

Ha: La variable Oxígeno disuelto en el embalse de Cáceres, NO se distribuye según un modelo Normal

Valor experimental del estadígrafo de contraste 0.148

p-valor: 0.161

Conclusión estadística y experimental

Los resultados son No Significativos, por tanto no podemos rechazar la Ho. La variable Oxígeno disuelto en el embalse de León, se distribuye según un modelo Normal

2.- CONTRASTE de HOMOCEASTICIDAD

Ho: Existe homocedasticidad (la varianza poblacional de la variable Ox. Disuelto en el embalse de León es la misma que la de dicha variable en el embalse de Cáceres)

Poner fórmula

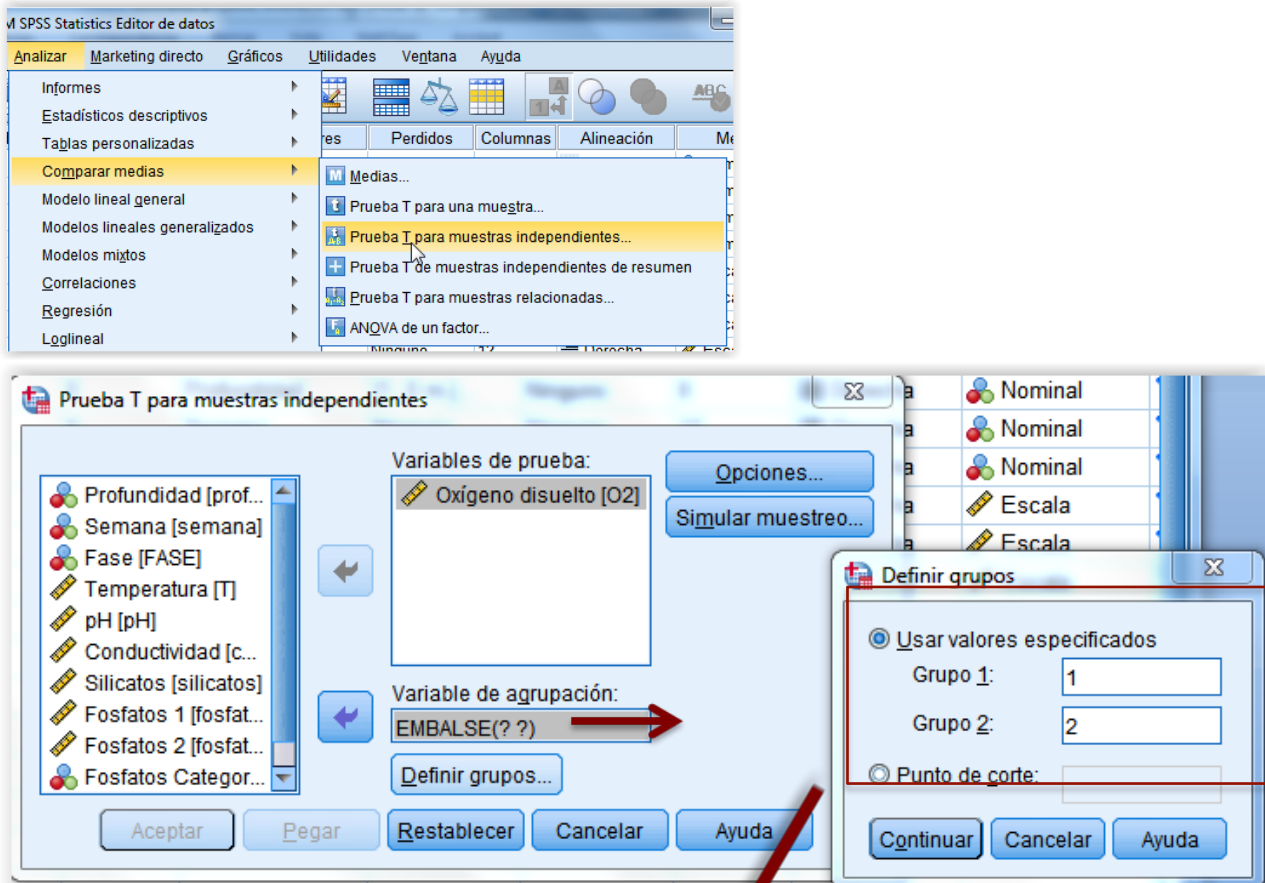
Ha: Existe heterocedasticidad (la varianza poblacional de la variable Ox. Disuelto en el embalse de León NO es la misma que la de dicha variable en el embalse de Cáceres)

Valor experimental del estadígrafo de contraste: 6,696

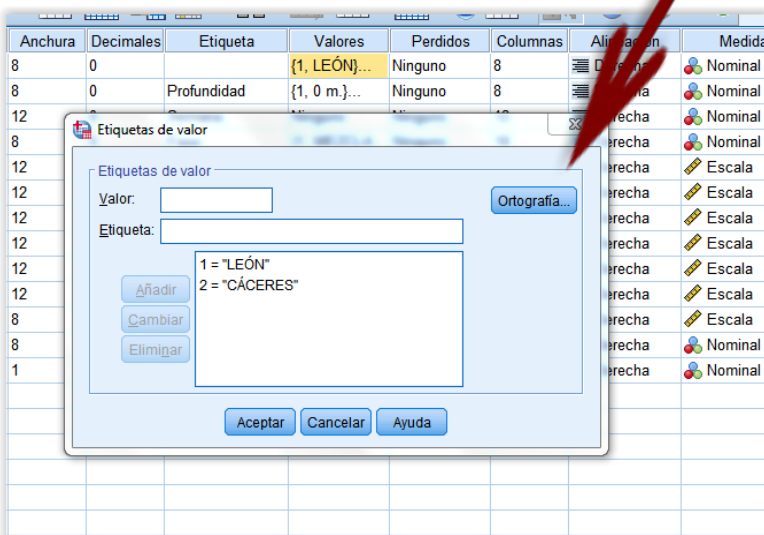
p-valor: 0,013

Conclusión estadística y experimental. Se rechaza la hipótesis nula, por lo que no podemos asumir homocedasticidad. (Hay heterocedasticidad), las varianzas poblacionales son distintas.

Los resultados de este contraste de Homocedasticidad, tal como se muestran a continuación, se obtienen cuando se realiza el contraste paramétrico de tendencia central, en las siguientes opciones:



Ponemos 1 y 2, cuando se presiona el botón Definir Grupos, porque el 1 se corresponde con el embalse de LEÓN y el 2 con el de CÁCERES



La parte de la tabla de resultados correspondiente al test de homocedasticidad, aparece a continuación:

		Prueba de Levene de igualdad de varianzas	
		F	Sig.
Oxígeno disuelto	Se asumen varianzas iguales	6,696	,013
	No se asumen varianzas iguales		

El SPSS realiza el test de Levene para contrastar la homocedasticidad. El valor experimental del estadígrafo de contraste vale 6,696, y el p-valor asociado 0,013, que

El resultado de este contraste, nos lleva a tener que decidir cuál es el estadígrafo de contraste.

Si el ejercicio lo hiciésemos a mano, utilizando el árbol de contrastes para dos poblaciones normales e independientes, con tamaño de muestra pequeño ($n_1=25$), y heterocedasticidad.

En el SPSS, tenemos en una misma tabla los resultados del test de homocedasticidad, que hemos escrito más arriba y continúa con la siguiente parte sobre el contraste para las medias de las dos poblaciones, que aparece a continuación, así como un resumen descriptivo de los datos analizados en este apartado. El programa utiliza el test de Levene para contrastar la homocedasticidad, y ya hemos visto que el valor experimental es 6,696 y el p-valor 0,013 (lo que implica que existe heterocedasticidad) por lo que deberemos realizar la lectura del resultado del contraste para igualdad de medias en la fila inferior de la tabla (señalada en rojo)

prueba t para la igualdad de medias						
t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
8,708	48	,000	3,215495	,369265	2,473037	3,957952
8,708	33,247	,000	3,215495	,369265	2,464431	3,966558

El programa también nos muestra una tabla con resultados descriptivos de la variable Oxígeno disuelto para cada embalse.

Estadísticas de grupo					
	EMBALSE	N	Media	Desviación estándar	Media de error estándar
Oxígeno disuelto	LEÓN	25	10,16400	,754365	,150873
	CÁCERES	25	6,94851	1,685186	,337037

3.- CONTRASTE DE TENDENCIA CENTRAL

Ho: $\mu_{\text{León}} = \mu_{\text{Cáceres}}$

Ha: $\mu_{\text{León}} \neq \mu_{\text{Cáceres}}$

Tipo de contraste: **bilateral**

Nivel de significación $\alpha=0.05$; $\alpha=0.01$

Expresión del estadígrafo de contraste y distribución teórica

$$t_{\text{Welch}} = \frac{\bar{x}_{\text{León}} - \bar{x}_{\text{Cáceres}}}{\sqrt{\frac{s_{\text{León}}^2}{n_{\text{León}}} + \frac{s_{\text{Cáceres}}^2}{n_{\text{Cáceres}}}}} \rightarrow t_f$$

Valor experimental del estadígrafo de contraste: **8.708**

p-valor: **0,000**

Conclusión estadística:

Se rechaza la Ho, con resultados altamente significativos. El p-valor es menor que el nivel de significación al 0.01; por tanto, el valor experimental se encuentra en la región de rechazo de la Ho a dicho nivel de significación, como puede verse en la siguiente figura:

Conclusión experimental:

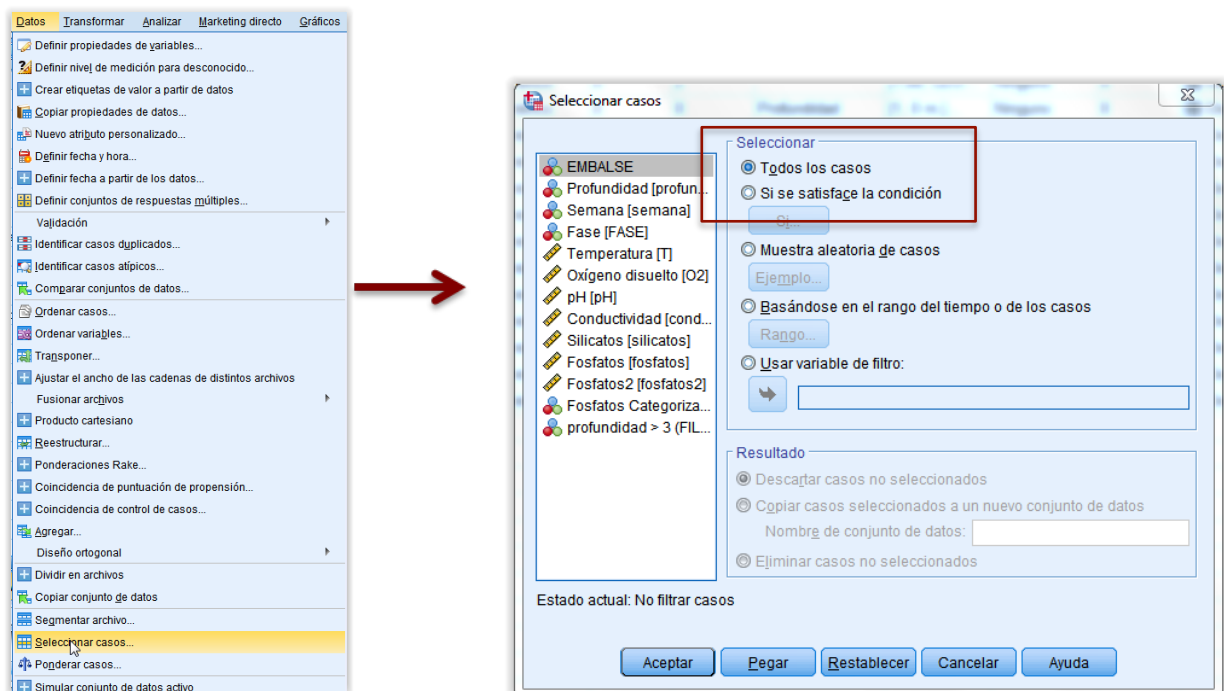
Podemos afirmar que la media poblacional de Oxígeno disuelto difiere entre los dos embalses en aquellas profundidades ≤ 15 m. de la fase de estratificación.

4.- Teniendo en cuenta todos los datos del estudio ¿Se puede afirmar que la tendencia central de la variable Oxígeno disuelto es significativamente diferente entre los dos embalses?.

Explique, detalladamente todos los pasos de los tres contrastes que tiene que realizar para responder a esta pregunta:

En primer lugar tenemos que verificar si existe normalidad o no de la variable Oxígeno disuelto en cada embalse.

Como anteriormente habíamos seleccionado los casos correspondientes a las profundidades de 15 metros o menos de la fase de estratificación, tenemos que eliminar dicha selección: tal como se muestra a continuación:



Una vez incluidos todos los datos otra vez en el estudio, realizamos el contraste de normalidad como en el apartado anterior:

CONTRASTES de NORMALIDAD (utilice el test de Kolmogorov-Smirnov de la opción Explorar)

H₀: La variable Oxígeno disuelto se distribuye según un modelo Normal en el embalse de León

H_a: La variable Oxígeno disuelto NO se distribuye según un modelo Normal en el embalse de León

Valor experimental del estadígrafo de contraste: 0.174

p-valor: 0.000

Conclusión estadística y experimental.

Como p-valor < 0.01, se rechaza H_0 con resultados altamente significativos.

Por tanto, la variable Oxígeno disuelto **NO se distribuye según un modelo Normal** en el embalse de León

H_0 : La variable Oxígeno disuelto se distribuye según un modelo Normal en el embalse de Cáceres

H_a : La variable Oxígeno disuelto **NO se distribuye según un modelo Normal** en el embalse de Cáceres

Valor experimental del estadígrafo de contraste: 0.108

p-valor: 0.041

Conclusión estadística y experimental

Como p-valor < 0.05, se encuentran razones para rechazar la H_0 (Resultados No Significativos)

Por tanto, se asume que la variable Oxígeno disuelto **NO se distribuye según un modelo Normal** en el embalse de León

A Continuación se presenta la tabla con los resultados de los contrastes de Normalidad con el test de Kolmogorov-Smirnov

Pruebas de normalidad				
	EMBALSE	Kolmogorov-Smirnov ^a		
		Estadístico	gl	Sig.
Oxígeno disuelto	LEÓN	,174	70	,000
	CÁCERES	,108	70	,041

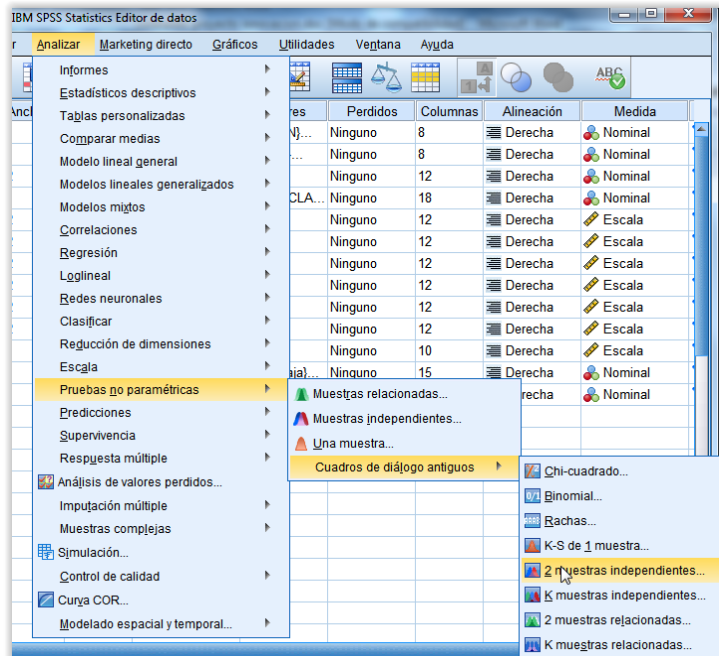
a. Corrección de significación de Lilliefors

Por tanto tendremos que realizar un contraste para la igualdad de medianas, UN CONTRASTE NO PARAMÉTRICO. Como se trata de datos independientes, el test que tenemos que utilizar es el test de la U de Mann-Whitney.

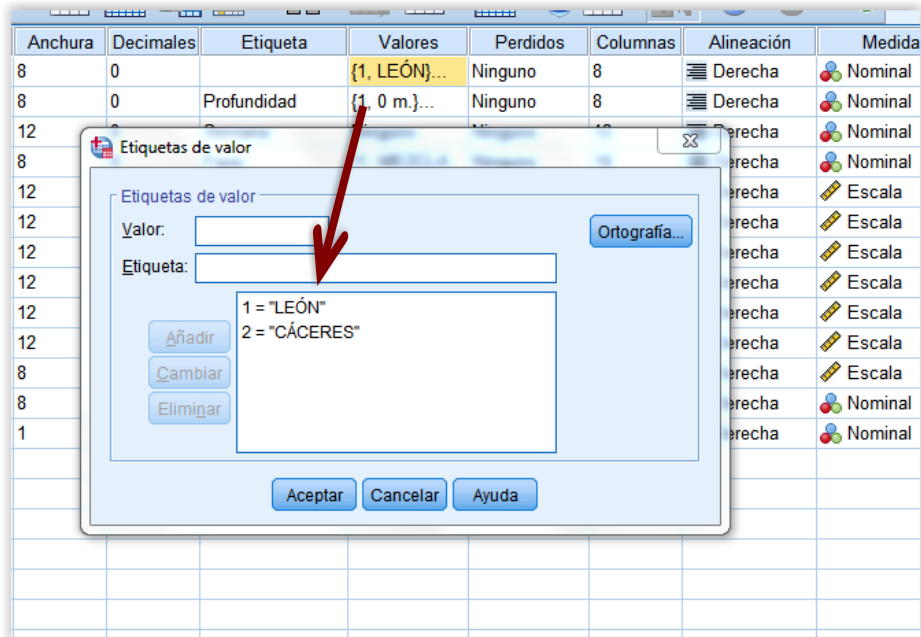
a. Variable de agrupación: EMBALSE

CONTRASTE DE TENDENCIA CENTRAL

Para realizar este contraste de tendencia central, tenemos que seleccionar la opción de test no paramétricos → cuadros de diálogo antiguos → 2 muestras independientes (pues no son los mismos datos: tenemos los datos del embalse de León y los datos del embalse de Cáceres).



En esta última opción de selección, vamos a hacer el test de la U de Mann-Whitney, (datos independientes). Seleccionamos la variable Oxígeno disuelto, y como variable de agrupación ponemos el EMBALSE. En el botón Definir grupos, se nos abre una nueva ventana para que decidamos entre qué categorías de la variable Embalse queremos hacer el análisis. En este caso sólo tenemos 2 (1 = León y 2 = Cáceres). Conviene ver en la tabla de datos cuáles son los valores de las etiquetas para hacer correctamente la selección (como se muestra abajo).



Resultados del análisis

Rangos

	EMBALSE	N	Rango promedio	Suma de rangos
Oxígeno disuelto	LEÓN	70	98,84	6919,00
	CÁCERES	70	42,16	2951,00
	Total	140		

Estadísticos de prueba^a

	Oxígeno disuelto
U de Mann-Whitney	466,000
W de Wilcoxon	2951,000
Z	-8,269
Sig. asintótica (bilateral)	,000

Ho: $Me_{León} = Me_{Cáceres}$

Ha: $Me_{Cáceres} \neq Me_{Cáceres}$

Nivel de significación $\alpha = 0.0$; $\alpha = 0.01$

Expresión del estadígrafo de contraste y distribución teórica

$$U = \min(U_1, U_2)$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$
$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$
$$R_1 = \sum \text{rangos grupo 1}$$
$$R_2 = \sum \text{rangos grupo 2}$$

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Valor experimental del estadígrafo de contraste: $Z_{\text{exp}} = -8.269$

p-valor: 0.000

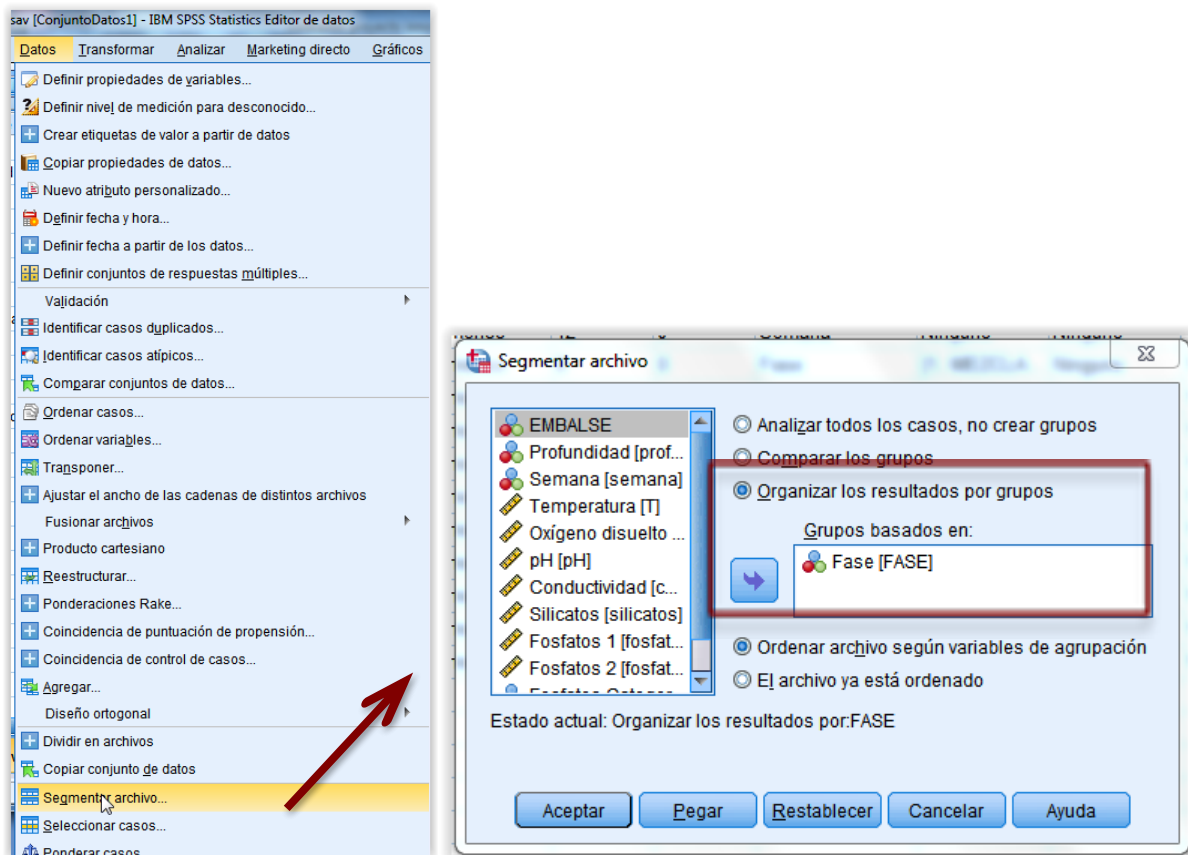
Conclusión estadística y experimental.

Desde el punto de vista estadístico se rechaza la hipótesis nula con resultados altamente significativos, ya que el p-valor < 0.01; de modo que experimentalmente podremos concluir que la Mediana de Oxígeno disuelto no es la misma, a nivel poblacional, en los dos embalses.

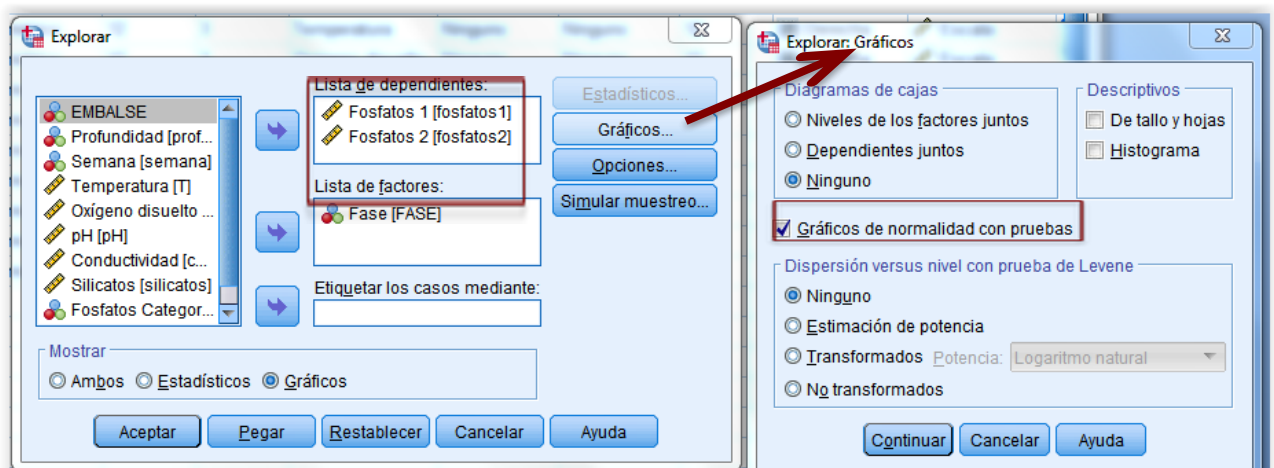
5.- Analice si, a nivel poblacional, existen diferencias en tendencia central entre las mediciones de los fosfatos por ambos métodos (fosfatos 1 y fosfatos 2) (Trabaje sólo con los datos pertenecientes a la fase de MEZCLA).

5.1.- Realice los contrastes de normalidad adecuados:

Primero segmentamos el archivo, para luego fijarnos únicamente en los resultados para la fase de MEZCLA:



Ahora realizamos el contraste de normalidad en Analizar→Estadísticos descriptivos→Explorar (tal como hemos hecho en secciones anteriores).



Los resultados se muestran a continuación:

Pruebas de normalidad				
	Fase	Kolmogorov-Smirnov ^a		
		Estadístico	gl	Sig.
Fosfatos 1	MEZCLA	,069	70	,200*
	ESTRATIFICACION	,138	70	,002
Fosfatos 2	MEZCLA	,085	70	,200*
	ESTRATIFICACION	,161	70	,000

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Ho: La variable Fosfatos 1 es Normal en la fase de Mezcla

Ha: La variable Fosfatos 1 no es Normal en la fase de Mezcla

p-valor: 0.200

Conclusión Estadística: No encontramos razones para rechazar la Ho; resultados N.S.
Por tanto podemos asumir una distribución Normal de la variable Fosfatos 1 en la fase de Mezcla

Ho: La variable Fosfatos 2 es Normal en la fase de Mezcla

Ha: La variable Fosfatos 2 no es Normal en la fase de Mezcla

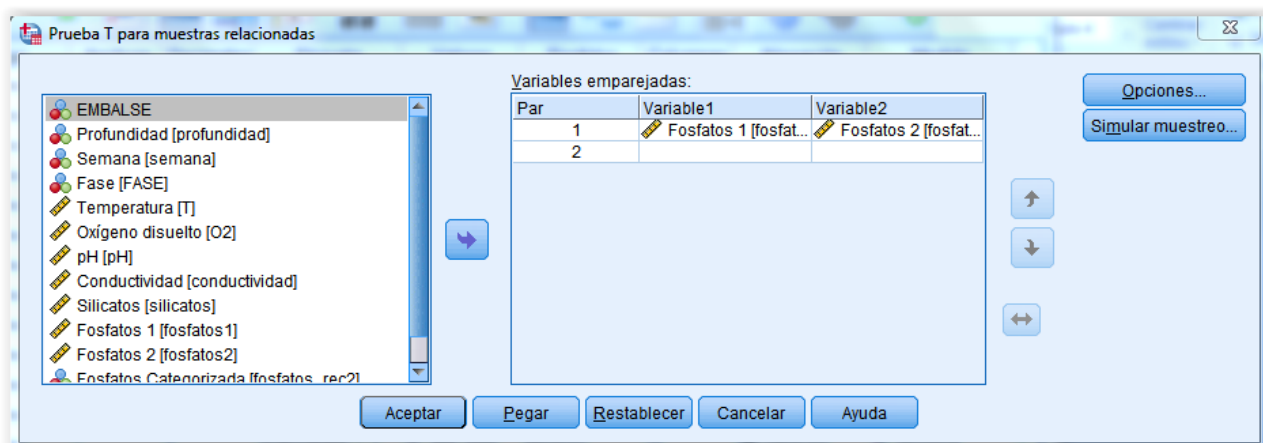
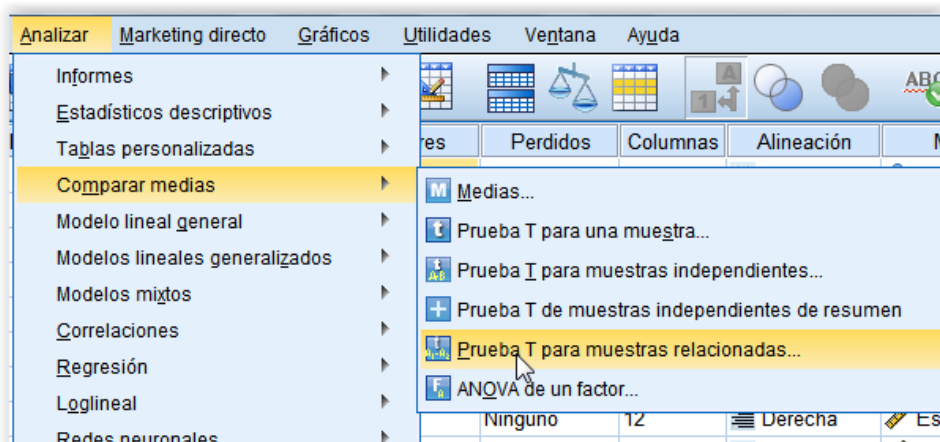
p-valor: 0.200

Conclusión Estadística: No encontramos razones para rechazar la Ho; resultados N.S.
Por tanto podemos asumir una distribución Normal de la variable Fosfatos 2 en la fase de Mezcla

5.2.-Realice el contraste de tendencia central adecuado

Ahora pasamos a realizar el contraste de tendencia central. Hay que tener en cuenta que se trata de DATOS APAREADOS, pues son los mismos puntos de muestreo medidos por dos procedimientos diferentes.

Por tanto, tenemos dos poblaciones normales y datos apareados. La opción en el menú del SPSS donde se realiza el análisis de tendencia central adecuado a estas condiciones es:



Tenemos que seleccionar las variables Fosfatos 1 y Fosfatos 2, teniendo en cuenta que el programa realizará las diferencias en ese sentido, es decir: Fosfatos 1 – Fosfatos 2.

Recordemos que para realizar un test paramétrico de tendencia central para datos apareados hay que trabajar con la muestra de diferencias (conseguida haciendo la diferencia entre los datos de cada muestra).

Los resultados del SPSS se muestran a continuación:

Estadísticas de muestras emparejadas ^a					
		Media	N	Desviación estándar	Media de error estándar
Par 1	Fosfatos 1	4,605	70	2,206	,264
	Fosfatos 2	4,805	70	2,150	,257

a. Fase = MEZCLA

Prueba de muestras emparejadas									
	Diferencias emparejadas					t	gl	Sig. (bilateral)	
	Media	Desv. estándar	Error estándar	95% de intervalo de confianza de la diferencia					
				Inferior	Superior				
Par 1	Fosfatos 1 - Fosfatos 2	-,200	,341	,041	-,282	-,200	-4,910	69	,000

Ho: $\mu_{\text{diferencias}} = 0 \iff \mu_{\text{Fosfatos 1}} = \mu_{\text{Fosfatos 2}}$

Ha: $\mu_{\text{diferencias}} \neq 0 \iff \mu_{\text{Fosfatos 1}} \neq \mu_{\text{Fosfatos 2}}$

Tipo de contraste: **bilateral**

Nivel de significación $\alpha=0.05$; $\alpha=0.01$

Expresión del estadígrafo de contraste y distribución teórica

Si lo hiciésemos a mano (ya que $n>30$)	Con el SPSS
$Z_{\text{apareados}} = \frac{\bar{x}_d}{\frac{\hat{s}_d}{\sqrt{n}}} \equiv N(0,1)$	$t_{\text{apareados}} = \frac{\bar{x}_d}{\frac{\hat{s}_d}{\sqrt{n}}} \equiv t_{n-1}$

Valor experimental del estadígrafo de contraste: **-4.910**

p-valor: **0,000**

Conclusión estadística:

Se rechaza la Ho, con resultados altamente significativos. El p-valor es menor que el nivel de significación al 0.01; por tanto, el valor experimental se encuentra en la región de rechazo de la Ho a dicho nivel de significación. La media de la población de diferencias es distinta de 0.

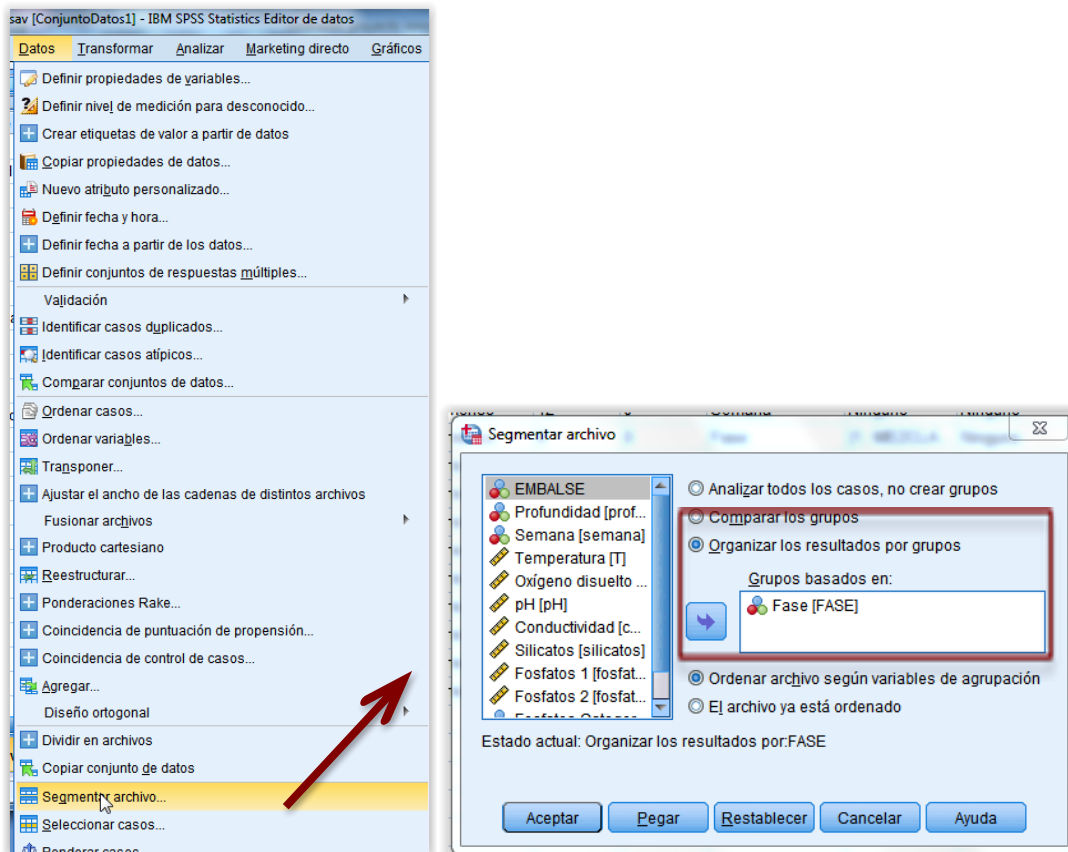
Conclusión experimental:

Podemos afirmar que en la FASE DE MEZCLA la media poblacional de la variable Fosfatos 1 difiere de la media poblacional de la variable Fosfatos 2.

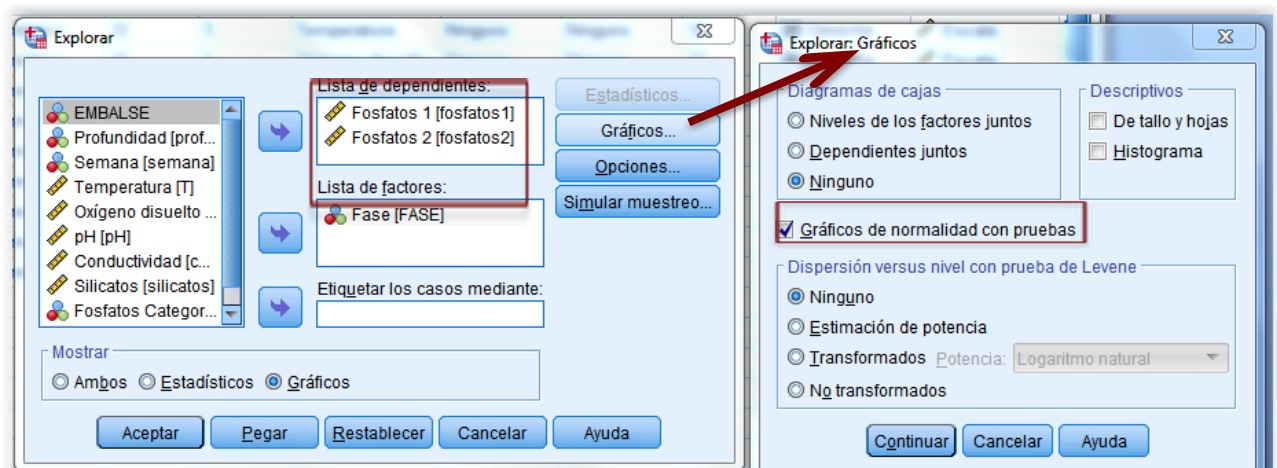
6.-Analice si, a nivel poblacional, existen diferencias en tendencia central entre las mediciones de los fosfatos por ambos métodos (fosfatos 1 y fosfatos 2) (Trabaje sólo con los datos pertenecientes a la fase de ESTRATIFICACIÓN).

6.1.- Realice los contrastes de normalidad adecuados:

Primero segmentamos el archivo, para luego fijarnos únicamente en los resultados para la fase de ESTRATIFICACIÓN:



Ahora realizamos el contraste de normalidad en Analizar→Estadísticos descriptivos→Explorar (tal como hemos hecho en secciones anteriores).



Los resultados se muestran a continuación:

Pruebas de normalidad				
	Fase	Kolmogorov-Smirnov ^a		
		Estadístico	gl	Sig.
Fosfatos 1	MEZCLA	,069	70	,200*
	ESTRATIFICACION	,138	70	,002
Fosfatos 2	MEZCLA	,085	70	,200*
	ESTRATIFICACION	,161	70	,000

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Ho: La variable Fosfatos 1 es Normal en la fase de Estratificación

Ha: La variable Fosfatos 1 no es Normal en la fase de Estratificación

p-valor: 0.002

Conclusión Estadística: Se rechaza la Ho con resultados Altamente Significativos. El p-valor es < 0.01

Por tanto podemos asumir que la variable **Fosfatos 1** no se distribuye según un modelo Normal en la fase de Estratificación

Ho: La variable Fosfatos 2 es Normal en la fase de Estratificación

Ha: La variable Fosfatos 2 no es Normal en a fase de Estratificación

p-valor: 0.000

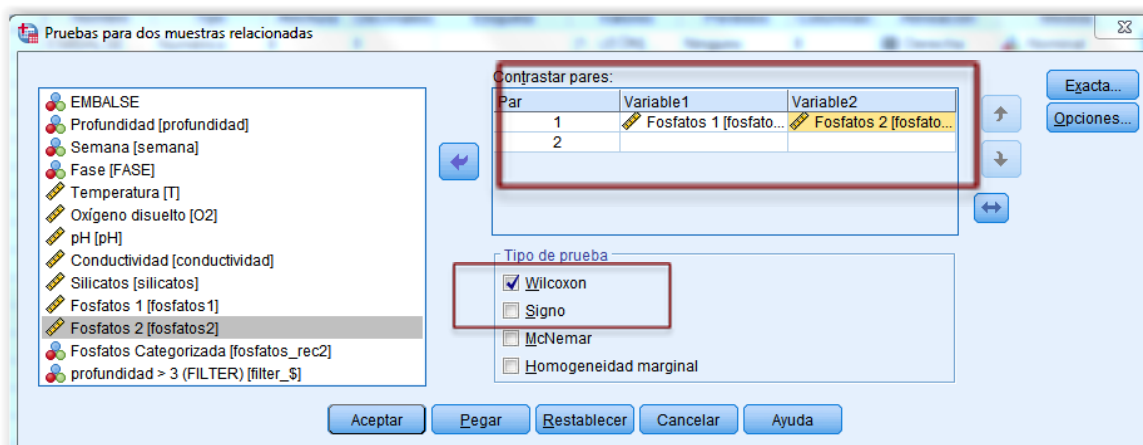
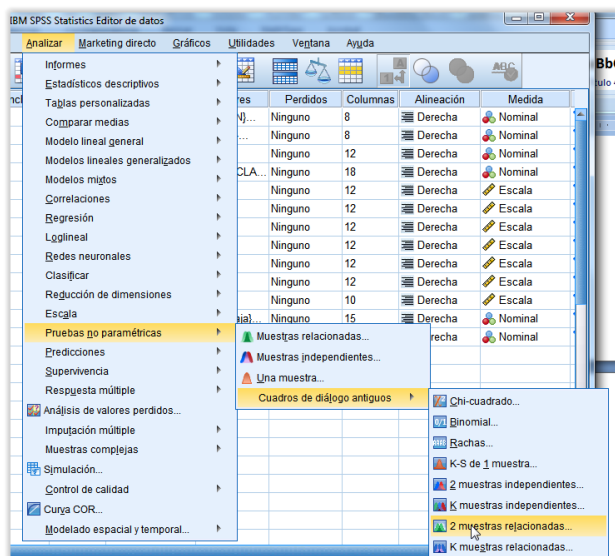
Conclusión Estadística: Se rechaza la Ho con resultados Altamente Significativos. El p-valor es < 0.01

Por tanto podemos asumir que la variable **Fosfatos 2** no se distribuye según un modelo Normal en la fase de Estratificación

6.2.-Realice el contraste de tendencia central adecuado

Ahora pasamos a realizar el contraste de tendencia central. Hay que tener en cuenta que se trata de DATOS APAREADOS, pues son los mismos puntos de muestreo medidos por dos procedimientos diferentes.

Por tanto, tenemos dos poblaciones NO NORMALES y datos apareados. La opción en el menú del SPSS donde se realiza el análisis de tendencia central adecuado a estas condiciones es:



Ya que tenemos que realizar el test de Wilcoxon (test no paramétrico para contraste de tendencia central en datos apareados).

Los resultados del SPSS se muestran a continuación:

Prueba de rangos con signo de Wilcoxon

		Rangos ^a		
		N	Rango promedio	Suma de rangos
Fosfatos 2 - Fosfatos 1	Rangos negativos	15 ^b	23,80	357,00
	Rangos positivos	55 ^c	38,69	2128,00
	Empates	0 ^d		
	Total	70		

a. Fase = **ESTRATIFICACION**

b. Fosfatos 2 < Fosfatos 1

c. Fosfatos 2 > Fosfatos 1

d. Fosfatos 2 = Fosfatos 1

Estadísticos de prueba ^{a,b}	
	Fosfatos 2 - Fosfatos 1
Z	-5,182^c
Sig. asintótica (bilateral)	,000

a. Fase = **ESTRATIFICACION**

b. Prueba de rangos con signo de Wilcoxon

c. Se basa en rangos negativos.

El test de Wilcoxon trabaja sobre medinas, por tanto:

Ho: Mediana Población de diferencias = 0 \leftrightarrow Mediana Fosfatos 1 = Mediana Fosfatos 2

Ha: Mediana Población de diferencias \neq 0 \leftrightarrow Mediana Fosfatos 1 \neq Mediana Fosfatos 2

Tipo de contraste: **bilateral**

Nivel de significación $\alpha=0.05$; $\alpha=0.01$

Expresión del estadígrafo de contraste y distribución teórica

$$T = \min(T_{(+)}, T_{(-)})$$

$$T_{(+)} = \sum \text{rangos de diferencias (+)}$$

$$T_{(-)} = \sum \text{rangos de diferencias (-)}$$

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Valor experimental del estadígrafo de contraste: **-5.182**

p-valor: **0,000**

Conclusión estadística:

Se rechaza la Ho, con resultados altamente significativos. El p-valor es menor que el nivel de significación al 0.01; por tanto, el valor experimental se encuentra en la región de rechazo de la Ho a dicho nivel de significación. La MEDIANA de la población de diferencias es distinta de 0.

Conclusión experimental:

Podemos afirmar que en la FASE DE ESTRATIFICACIÓN la MEDIANA poblacional de la variable Fosfatos 1 difiere de la MEDIANA poblacional de la variable Fosfatos 2.

Práctica 5

ANÁLISIS DE LA VARIANZA

OBJETIVOS y PREGUNTAS

1.- Analice si la variable pH difiere en media entre las profundidades 10 , 15, 20 metros y Fondo teniendo en cuenta todos los datos de los 2 embalses pero solamente los correspondientes a la FASE DE ESTRATIFICACIÓN

1. Analice si se verifican las hipótesis para poder realizar un ANOVA: Normalidad y Homocedasticidad. Para ello establezca la hipótesis nula y alternativa de cada contraste.
2. Teniendo en cuenta los resultados de la pregunta anterior que contraste debemos realizar para responder a la pregunta experimental planteada.
3. ¿Cuáles son la hipótesis nula y alternativa?.
4. Escriba el nivel o niveles de significación con los cuales quiere trabajar.
5. Cuál es el estadígrafo de contraste y su distribución de probabilidad bajo el supuesto de hipótesis nula cierta.
6. Cuál es valor del estadígrafo experimental de nuestros datos. Escriba la tabla de resultados típica de este contraste.
7. Cuál es el p-valor del contraste. Interpretelo y escriba las conclusiones estadísticas y experimentales.
8. A partir de los resultados anteriores contraste entre que profundidades se ha detectado diferencias en las profundidades.
9. ¿Qué test ha utilizado para responder la pregunta anterior y porqué ha elegido ese test.
10. Escriba los pasos del contraste anterior en términos generales.

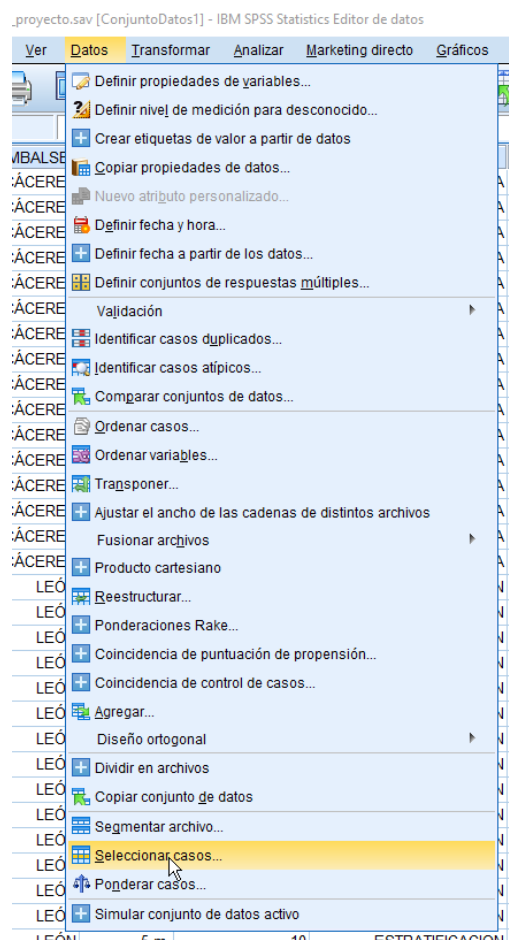
Práctica 5

ANÁLISIS DE LA VARIANZA

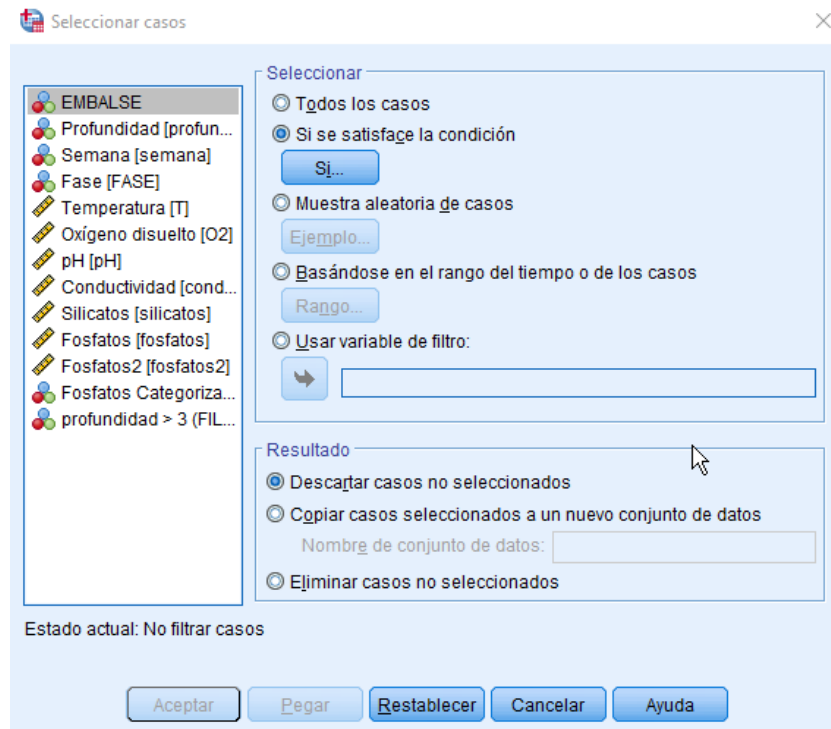
SOLUCIÓN


Lo primero de todo antes de realizar el análisis debemos seleccionar los datos de la FASE DE ESTRATIFICACION ya que el enunciado nos dice que debemos hacer la comparación para este conjunto de datos. Para ello, realizaremos los siguientes pasos:

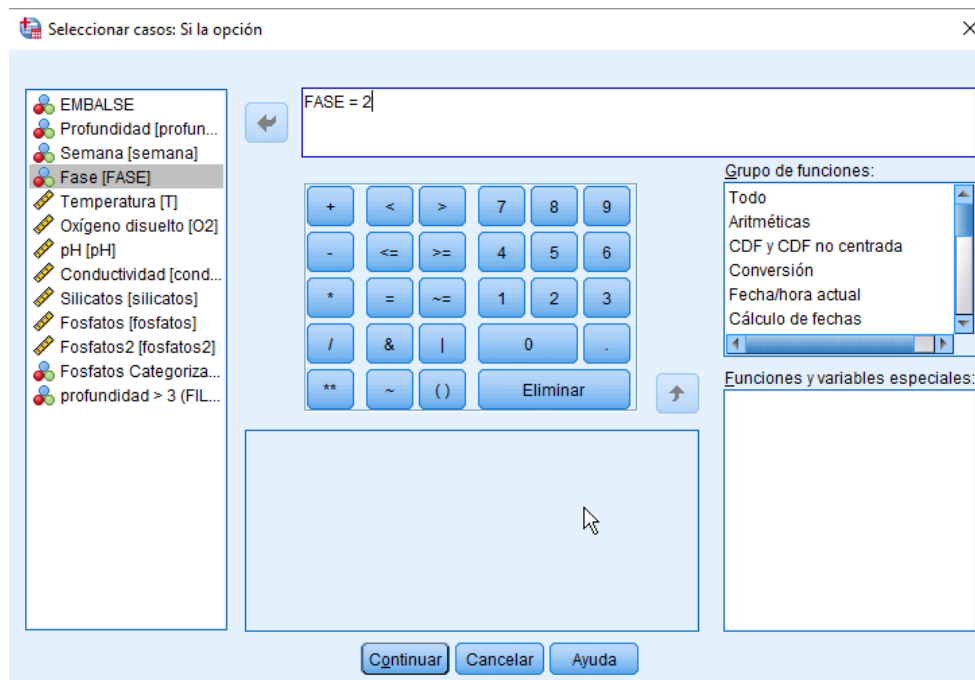
- Menú **Datos** → **Seleccionar casos...**



- En el menú emergente seleccionar la opción **“Si se satisface la condición”** y presionar el botón **SI**.



- En la nueva ventana cuando se presiona el SI, debemos seleccionar la variable Fase del cuadro izquierdo y pasarla al cuadro derecho con el botón . Después marcar el igual en la calculadora inferior y escribir el valor que representa a la Fase de estratificación. En este caso el valor es el 2.



- En la hoja de los valores se puede ver como los valores de fase mezcla están tachados y los de estratificación no.

	EMBALSE	profundidad	semana	FASE	T
52	CÁCERES	10 m.	2	MEZCLA	9,389
53	CÁCERES	10 m.	3	MEZCLA	11,520
54	CÁCERES	10 m.	4	MEZCLA	10,778
55	CÁCERES	10 m.	5	MEZCLA	13,487
56	CÁCERES	15 m.	1	MEZCLA	6,516
57	CÁCERES	15 m.	2	MEZCLA	7,959
58	CÁCERES	15 m.	3	MEZCLA	10,803
59	CÁCERES	15 m.	4	MEZCLA	12,445
60	CÁCERES	15 m.	5	MEZCLA	9,223
61	CÁCERES	20 m.	1	MEZCLA	6,449
62	CÁCERES	20 m.	2	MEZCLA	11,641
63	CÁCERES	20 m.	3	MEZCLA	10,376
64	CÁCERES	20 m.	4	MEZCLA	9,779
65	CÁCERES	20 m.	5	MEZCLA	10,279
66	CÁCERES	Fdo.	1	MEZCLA	7,324
67	CÁCERES	Fdo.	2	MEZCLA	6,792
68	CÁCERES	Fdo.	3	MEZCLA	8,621
69	CÁCERES	Fdo.	4	MEZCLA	8,222
70	CÁCERES	Fdo.	5	MEZCLA	7,831
71	LEÓN	0 m.	6	ESTRATIFICACION	15,800
72	LEÓN	0 m.	7	ESTRATIFICACION	15,900
73	LEÓN	0 m.	8	ESTRATIFICACION	16,700
74	LEÓN	0 m.	9	ESTRATIFICACION	16,300
75	LEÓN	0 m.	10	ESTRATIFICACION	19,700
76	LEÓN	2 m.	6	ESTRATIFICACION	15,200
77	LEÓN	2 m.	7	ESTRATIFICACION	15,300
78	LEÓN	2 m.	8	ESTRATIFICACION	15,800
79	LEÓN	2 m.	9	ESTRATIFICACION	15,800
80	LEÓN	2 m.	10	ESTRATIFICACION	18,800
81	LEÓN	5 m.	6	ESTRATIFICACION	13,500
82	LEÓN	5 m.	7	ESTRATIFICACION	14,100

- A partir de ahora las preguntas se realizarán con los datos seleccionados de Estratificación.

1. Analice si se verifican las hipótesis para poder realizar un ANOVA: Normalidad y Homocedasticidad. Para ello establezca la hipótesis nula y alternativa de cada contraste.

Antes de realizar un Análisis de la Varianza se deben comprobar si los datos cumplen los supuestos para que los resultados del análisis sean fiables.

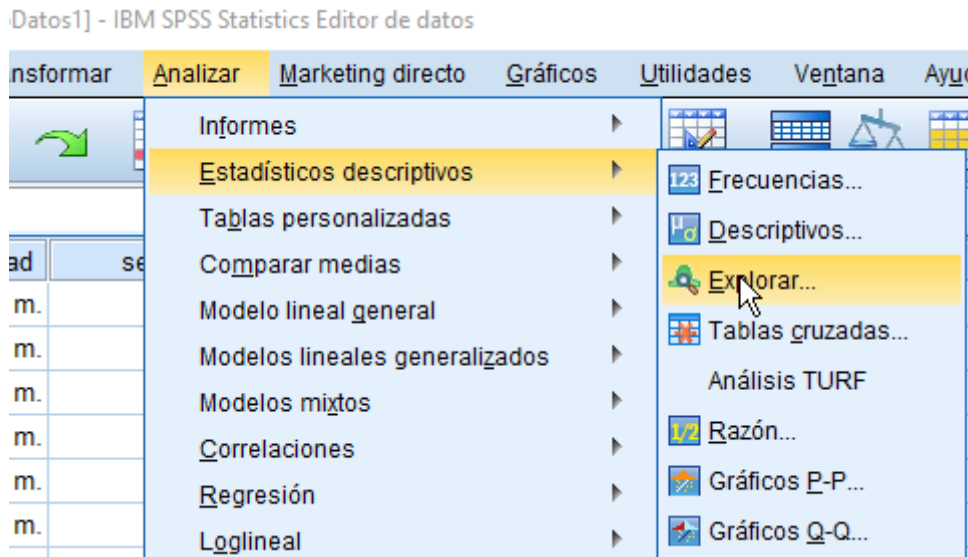
1.- El primer supuesto a comprobar para realizar un ANOVA es la Normalidad, El ANOVA es un test robusto para desviaciones moderadas de este supuesto, por lo tanto, los resultados pueden ser fiables para ligeras desviaciones de Normalidad. Para contrastar este supuesto inicial partimos de las siguientes hipótesis para cada grupo de comparación:

H₀: El pH sigue una distribución Normal.

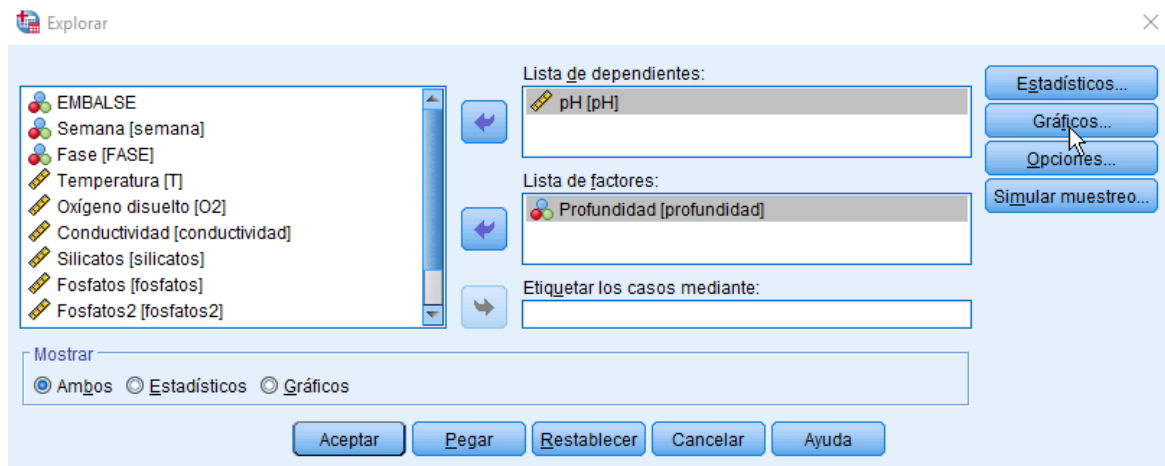
H_a: El pH no sigue una distribución Normal.

En Spss se pueden hacer dos test de Normalidad: El test de Kolmogorov-Smirnov y el Shapiro-Wilk. Generalmente se pueden utilizar los dos, si bien, Shapiro-Wilks sería más potente para muestras con tamaños más pequeños.

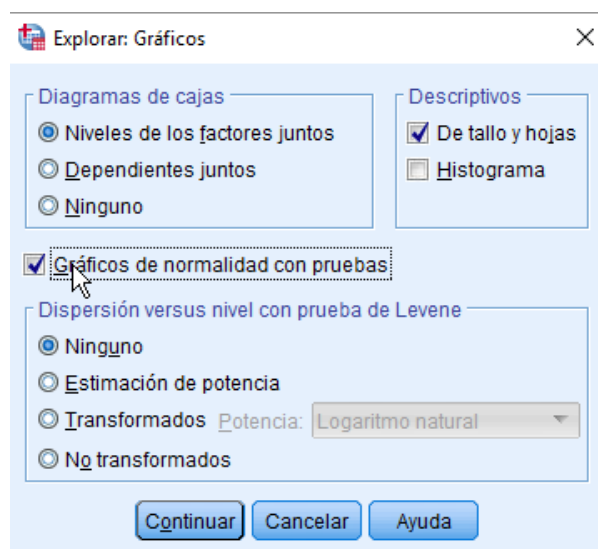
Como se ha visto en otras prácticas para realizar este contraste debemos ir a la opción de **Explorar** en el submenú **Estadística descriptiva** del menú **Analizar**



Dentro de explorar debemos marcar el botón de gráficos.



En la ventana emergente se marca el cuadro de **Gráficos de normalidad con pruebas...**



Se presiona el botón de continuar y en la ventana de resultados se busca la tabla de los contrastes de normalidad:

Pruebas de normalidad

Profundidad	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
pH 0 m.	,170	10	,200 [*]	,923	10	,382
2 m.	,188	10	,200 [*]	,886	10	,154
5 m.	,167	10	,200 [*]	,957	10	,746
10 m.	,202	10	,200 [*]	,940	10	,551
15 m.	,178	10	,200 [*]	,968	10	,872
20 m.	,250	10	,078	,866	10	,090
Fdo.	,200	10	,200 [*]	,961	10	,801

P-valores

Se observa como los p-valores para todos los grupos son superiores al nivel de significación usualmente utilizado del 0,05. Por lo tanto, no tenemos motivos para rechazar la H_0 , es decir, **no** tenemos motivos para considerar que la variable pH no sigue una distribución Normal en cada uno de las profundidades.

2.- El segundo supuesto, y el más sensible a desviaciones del mismo, es que las varianzas poblacionales de las 7 profundidades sean iguales. Es decir, que los grupos sean homocedásticos. Las hipótesis a contrastar son:

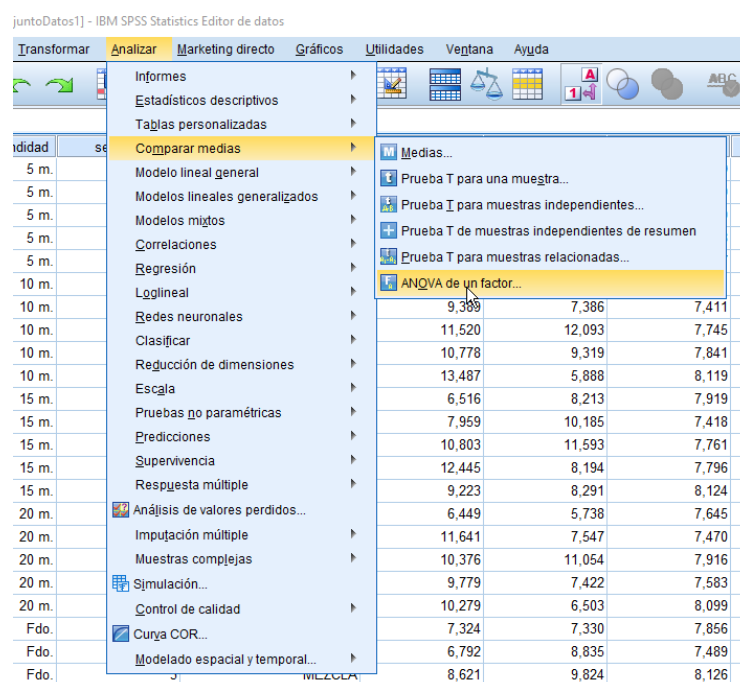
$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma_7^2$$

$$H_a : \text{Alguna } \sigma_i^2 \text{ es diferente, } \forall i$$

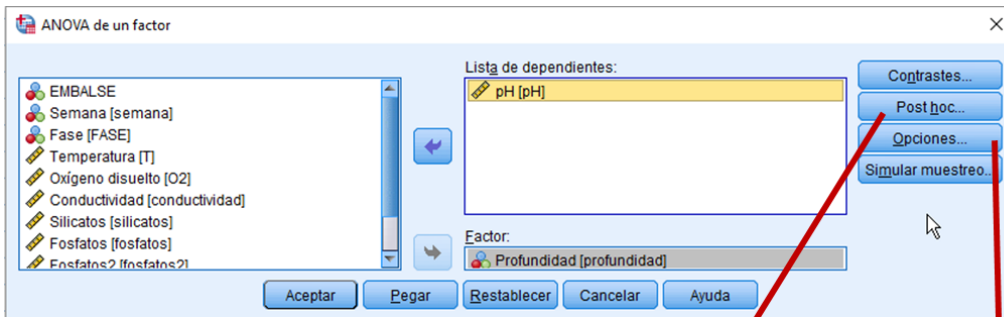
El contraste que se debe realizar es el test de Levene. Para obtener este contraste debemos hacer el Análisis de la Varianza. Dentro de las opciones nos permite elegir el test de Levene.

Para realizar el ANOVA debemos hacer lo siguiente:

- En el menú **Analizar**, elegir la opción **Comparar medias** y dentro de esta opción **ANOVA de un factor**.



Como en todos los análisis nos aparece la ventana donde tenemos que determinar cuáles son las variables que participan en el análisis. En Lista de dependientes se debe poner la variable cuantitativa, en este caso el *ph*, y en Factor, la variable que define los grupos (poblaciones), en este caso la *Profundidad*.

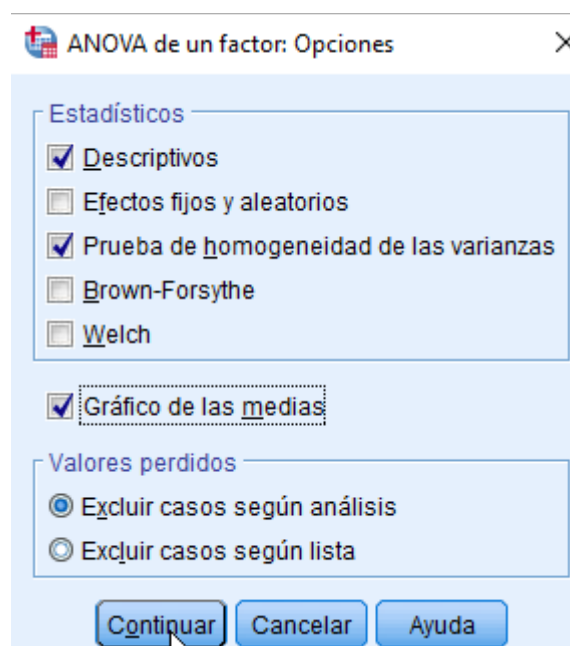


Presionando este botón nos aparece un menú donde están los diferentes test a posteriori para conocer donde están las diferencias.

Presionando este botón nos aparece un menú donde nos permite obtener Valores descriptivos y la prueba de homocedasticidad.

A la derecha hay varios botones que nos permiten obtener análisis adicionales al predeterminado por esta técnica.

Si presionamos el botón Opciones, se abre una nueva ventana donde se puede elegir mostrar los descriptivos de cada grupo (**Descriptivos**), la prueba de homocedasticidad (**Prueba de homogeneidad de varianzas**) y el gráfico de medias muestrales (Gráfico de las medias).



Para responder nuestra pregunta de si las varianzas poblacionales son iguales debemos ir a la página de resultados y buscar la tabla siguiente.

Prueba de homogeneidad de varianzas

pH

Estadístico de Levene	gl1	gl2	Sig.
2,110	6	63	,064

El p-valor (Sig.) es 0,064. Como es mayor que 0,05, los resultados son no significativos y por lo tanto no tenemos motivos para rechazar la hipótesis nula. En conclusión consideramos que se cumple el supuesto de igualdad de varianzas.

2. Teniendo en cuenta los resultados de la pregunta anterior que contraste debemos realizar para responder a la pregunta experimental planteada.

El contraste que se debe hacer es un Análisis de la varianza (ANOVA), debido a que se quiere contrastar si **más de dos grupos** presentan diferente media poblacional.

3. ¿Cuáles son la hipótesis nula y alternativa?

$$H_0 : \mu_1^2 = \mu_2^2 = \mu_3^2 = \mu_4^2 = \mu_5^2 = \mu_6^2 = \mu_7^2$$

$$H_a : \text{Algún } \mu_i^2 \text{ es diferente, } \forall i$$

4. Escriba el nivel o niveles de significación con los cuales quiere trabajar.

Ya que no me han indicado ningún nivel de significación concreto, trabajaré con los habituales del 5% y 1%

5.Cuál es el estadígrafo de contraste y su distribución de probabilidad bajo el supuesto de hipótesis nula cierta.

El estadígrafo de contraste es el cociente entre la varianza entre y la varianza dentro. Su distribución es una F de Snedecor con r-1, N-r grados de libertad (r es el número de grupos y N es el número total de observaciones). En nuestro caso particular los grados de libertad son 6 y 63.

6.Cuál es valor del estadígrafo experimental de nuestros datos. Escriba la tabla de resultados típica de este contraste.

En la hoja de resultados se tiene que buscar la tabla típica del ANOVA. Esta tabla es el resultado que da por defecto el programa. El valor del estadígrafo experimental es 10,898.

ANOVA

pH

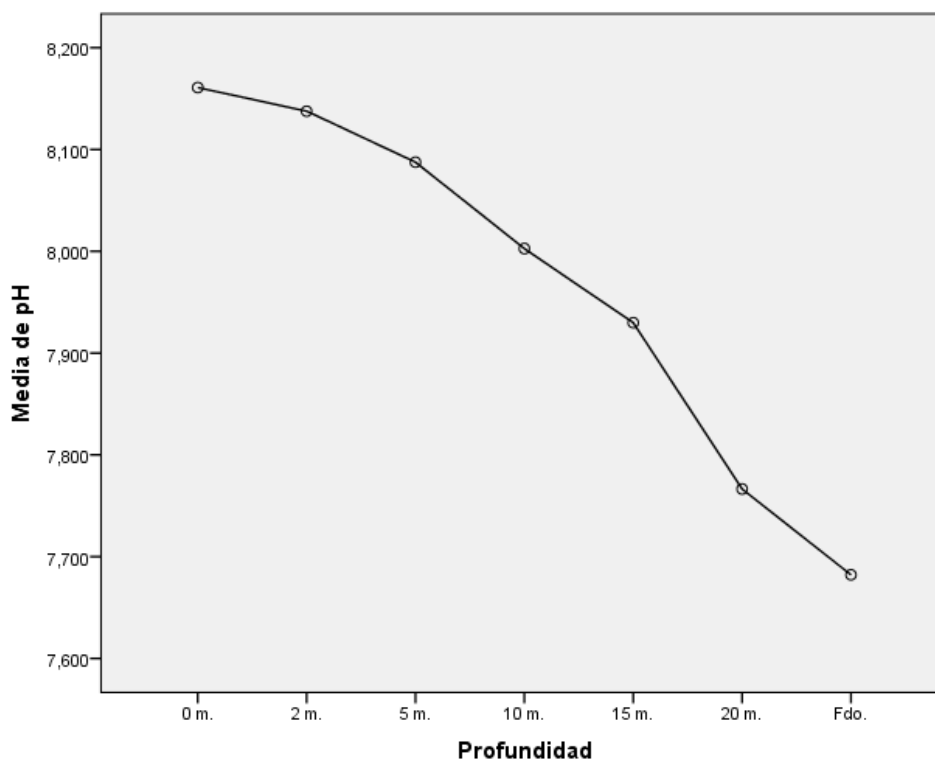
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	2,052	6	,342	10,898	,000
Dentro de grupos	1,977	63	,031		
Total	4,030	69			

Podemos también visualizar en la hoja de resultados la descriptiva básica y el gráfico de medias si se han marcado estos resultados en la ventana de opciones.

Descriptivos

pH

	N	Media	Desviación estándar	Error estándar	95% del intervalo de confianza para la media		Mínimo	Máximo
					Límite inferior	Límite superior		
0 m.	10	8,16083	,183248	,057948	8,02974	8,29192	7,900	8,425
2 m.	10	8,13763	,126109	,039879	8,04742	8,22784	8,000	8,329
5 m.	10	8,08753	,141091	,044617	7,98660	8,18846	7,800	8,306
10 m.	10	8,00265	,146829	,046431	7,89762	8,10769	7,800	8,293
15 m.	10	7,92979	,151354	,047862	7,82152	8,03806	7,700	8,222
20 m.	10	7,76648	,173113	,054743	7,64264	7,89032	7,600	8,116
Fdo.	10	7,68214	,275460	,087108	7,48509	7,87919	7,300	8,159
Total	70	7,96672	,241659	,028884	7,90910	8,02434	7,300	8,425

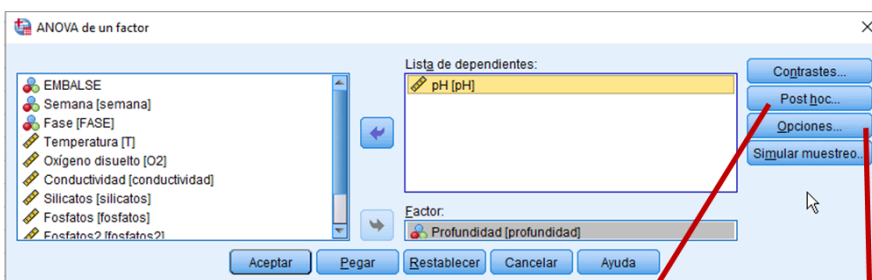


7. *Cuál es el p-valor del contraste. Interpretelo y escriba las conclusiones estadísticas y experimentales.*

El p-valor es menor de 0,00001, por lo tanto, los resultados son altamente significativos, lo que quiere decir que rechazamos la hipótesis nula y alguna de las medias poblacionales es diferente con la probabilidad de equivocarnos del 1%. Es decir, que hay diferencias en el Ph de al menos una de las profundidades analizadas.

8. *A partir de los resultados anteriores contraste entre que profundidades se ha detectado diferencias en las profundidades.*

Para responder a esta pregunta debemos de marcar en la ventana de ANOVA de un factor el botón de **post hoc**.



Presionando este botón nos aparece un menú donde están los diferentes test a posteriori para conocer donde están las diferencias.

Presionando este botón nos aparece un menú donde nos permite obtener Valores descriptivos y la prueba de homocedasticidad.

Nos aparece una nueva ventana con muchas propuestas para ver se encuentran las diferencias una vez el ANOVA ha salido significativo. Los test que se conocen son DMS (LSD), donde no se corrigen el nivel de significación de cada comparación; Tukey, donde la corrección del nivel de significación es el número de grupos (r); Bonferroni donde la corrección es el número de comparaciones ($(r*(r-1)/2$); Dunnett, si se quiere comparar las diferencias de los tratamientos contra un control, y su corrección es el número de grupos menos 1 ($r-1$). En el caso de Dunnett hay que indicar al programa en qué posición está el grupo control, el primero o último. Por defecto, viene la posición "últimos". Por lo tanto, hay que tener cuidado al definir las categorías del factor del análisis de la varianza cuando vamos a utilizar Dunnett porque al control se le debe asignar el menor o el mayor valor.

Para ver que entre qué profundidades existen diferencias marcaremos la casilla de Tukey porque todas las muestras tienen el mismo tamaño muestral, es decir, son balanceadas.

Las diferencias se encuentran entre:

Las profundidades más bajas 0, 2 y 5 con las más profundas, 20 y fondo.

Las profundidades de 10, 15 con el fondo.

Otra forma de indicarlo, el fondo se diferencia con todas las profundidades en Ph y las profundidades más bajas 0, 2, 5 también se diferencia con la segunda profundidad mayor, 20 metros.

ANOVA de un factor: Comparaciones múltiples post hoc

Asumiendo varianzas iguales

DMS S-N-K Waller-Duncan
 Bonferroni Tukey Tasa de errores tipo I/tipo II: 100
 Sidak Tukey-b Dunnett
 Scheffe Duncan Categoría de control: Último
 R-E-G-W F GT2 de Hochberg Prueba
 R-E-G-W Q Gabriel Bilateral < Control > Control

No asumiendo varianzas iguales

T2 de Tamhane T3 de Dunnett Games-Howell C de Dunnett

Nivel de significación: 0,05

Continuar Cancelar Ayuda

Se puede cambiar el nivel significación

Comparaciones múltiples

Variable dependiente: pH

HSD Tukey

(I) Profundidad	(J) Profundidad	Diferencia de medias (I-J)	Error estándar	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
0 m.	2 m.	,023202	,079229	1,000	-,21810	,26450
	5 m.	,073300	,079229	,967	-,16800	,31460
	10 m.	,158179	,079229	,427	-,08312	,39948
	15 m.	,231041	,079229	,069	-,01026	,47234
	20 m.	,394350 [*]	,079229	,000	,15305	,63565
	Fdo.	,478690 [*]	,079229	,000	,23739	,71999
2 m.	0 m.	-,023202	,079229	1,000	-,26450	,21810
	5 m.	,050097	,079229	,995	-,19120	,29140
	10 m.	,134977	,079229	,616	-,10632	,37628
	15 m.	,207838	,079229	,136	-,03346	,44914
	20 m.	,371148 [*]	,079229	,000	,12985	,61245
	Fdo.	,455488 [*]	,079229	,000	,21419	,69679
5 m.	0 m.	-,073300	,079229	,967	-,31460	,16800
	2 m.	-,050097	,079229	,995	-,29140	,19120
	10 m.	,084879	,079229	,934	-,15642	,32618
	15 m.	,157741	,079229	,431	-,08356	,39904
	20 m.	,321050 [*]	,079229	,003	,07975	,56235
	Fdo.	,405390 [*]	,079229	,000	,16409	,64669
10 m.	0 m.	-,158179	,079229	,427	-,39948	,08312

	2 m.	-,134977	,079229	,616	-,37628	,10632
	5 m.	-,084879	,079229	,934	-,32618	,15642
	15 m.	,072862	,079229	,968	-,16844	,31416
	20 m.	,236171	,079229	,059	-,00513	,47747
	Fdo.	,320511 ⁺	,079229	,003	,07921	,56181
15 m.	0 m.	-,231041	,079229	,069	-,47234	,01026
	2 m.	-,207838	,079229	,136	-,44914	,03346
	5 m.	-,157741	,079229	,431	-,39904	,08356
	10 m.	-,072862	,079229	,968	-,31416	,16844
	20 m.	,163309	,079229	,388	-,07799	,40461
	Fdo.	,247649 ⁺	,079229	,041	,00635	,48895
20 m.	0 m.	-,394350 ⁺	,079229	,000	-,63565	-,15305
	2 m.	-,371148 ⁺	,079229	,000	-,61245	-,12985
	5 m.	-,321050 ⁺	,079229	,003	-,56235	-,07975
	10 m.	-,236171	,079229	,059	-,47747	,00513
	15 m.	-,163309	,079229	,388	-,40461	,07799
	Fdo.	,084340	,079229	,936	-,15696	,32564
Fdo.	0 m.	-,478690 ⁺	,079229	,000	-,71999	-,23739
	2 m.	-,455488 ⁺	,079229	,000	-,69679	-,21419
	5 m.	-,405390 ⁺	,079229	,000	-,64669	-,16409
	10 m.	-,320511 ⁺	,079229	,003	-,56181	-,07921
	15 m.	-,247649 ⁺	,079229	,041	-,48895	-,00635
	20 m.	-,084340	,079229	,936	-,32564	,15696

*. La diferencia de medias es significativa en el nivel 0.05.

9. ¿Qué test ha utilizado para responder la pregunta anterior y por qué ha elegido ese test.

Como hemos indicado antes se ha utilizado el test de Tukey porque las muestras son balanceadas es decir presentan los mismos tamaños muestrales.

10. Escriba los pasos generales del contraste anterior en términos generales:

$$H_0: \mu_i = \mu_j, \forall i, j$$

$$H_a: \mu_i \neq \mu_j$$

Niveles de significación, 0,05 con la penalización de Tukey,

Estadígrafo de contraste

$$t = \frac{\bar{x}_i - \bar{x}_j}{S_D \sqrt{\frac{2}{n}}}$$

El estadígrafo de contraste sigue una distribución t de student con N-r grados de libertad y una penalización k=5

$$RA = \{t / |t| \leq t_{N-r, k, \alpha}\}$$

$$RR = \{t / |t| > t_{N-r, k, \alpha}\}$$

Práctica 6

TABLAS DE CONTINGENCIA

OBJETIVOS y PREGUNTAS

Un lago de aguas claras debe esta claridad a la falta de algas y de otras plantas acuáticas que no encuentran las condiciones de vida adecuadas a causa de la falta de fosfatos y nitratos que le son vitales.

Cuando se llevan mediante desagües y aguas filtradas estos elementos al agua de los lagos, aparece rápidamente un incremento excesivo (eutrofia) de algas que conduce al ensuciamiento de los lagos. Es especialmente temida en aquellos lagos adecuados para el baño, el alga llamada "oscillatoria rubescens", (es de la familia de las cianobacterias) que da al agua un colorido marrón rojizo poco atractivo. Existe, por lo tanto, una preocupación, por motivos turísticos, para mantener alejados los desagües domésticos, que contienen, debido al uso de detergentes habituales, grandes cantidades de fosfatos, de los lagos. El problema de la eutrofia no es tan importante en las aguas corrientes.

Los investigadores de este estudio piensan que puede haber una relación entre el Nivel de fosfatos (categorizado como Bajo, Medio y Alto) y la fase (Mezcla y Estratificación) en la que se encuentre el embalse, y por tanto se deban hacer más controles en relación al posible uso turístico de los mismos.

Analice si la sospecha de los investigadores es o no cierta.

Para ello, conteste a las siguientes cuestiones:

1. Defina claramente las hipótesis del contraste.
2. Fije el nivel o niveles de significación del contraste.
3. Escriba la expresión del estadígrafo de contraste.
4. ¿Cuál es la distribución teórica del estadígrafo de contraste?
5. ¿Cuál es el valor experimental del estadígrafo de contraste y su p-valor.
6. Explique cuál es la conclusión estadística y experimental.
7. Si las dos variables fueran independientes, ¿cuál sería el nivel de fosfatos bajo esperado en la fase de estratificación? y ¿Cómo se calcula este valor?:
8. ¿Qué celda de la tabla de contingencia es la más responsable de la relación entre el nivel de fosfatos y la fase? y ¿Cuál es su contribución al valor experimental?. Especificar cómo se calcula este valor.
9. ¿Qué porcentaje se da en la fase de estratificación y en el nivel bajo de fosfatos?.
10. ¿Qué porcentaje de la fase de estratificación tiene un nivel de fosfatos bajo? y ¿Qué porcentaje de nivel de fosfatos medio hay en la fase de mezcla?.

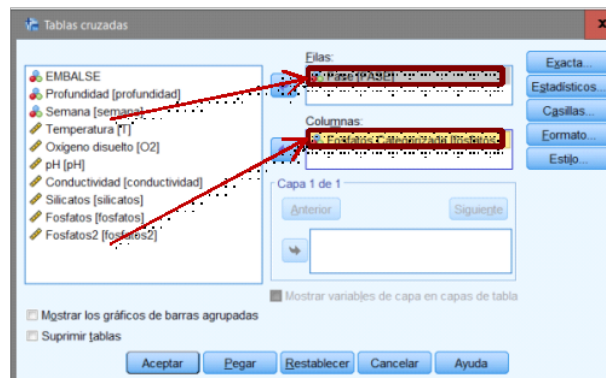
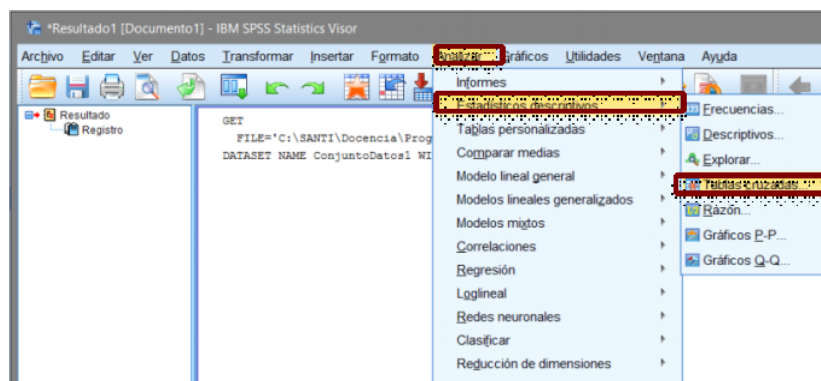
Práctica 6

TABLAS DE CONTINGENCIA

SOLUCIÓN

Para realizar un análisis de tablas de contingencia tenemos que ir al menú Analizar → Estadísticos descriptivos → Tablas de frecuencias.

Luego seleccionamos la variable que queremos poner en filas y la que queremos poner en columnas.



1. Defina claramente las hipótesis del contraste:

H_0 : El Nivel de fosfatos es independiente de la fase en la que se encuentre el embalse.

H_a : El Nivel de fosfatos está relacionado con la fase en la que se encuentre el embalse.

2. Fije el nivel o niveles de significación del contraste:

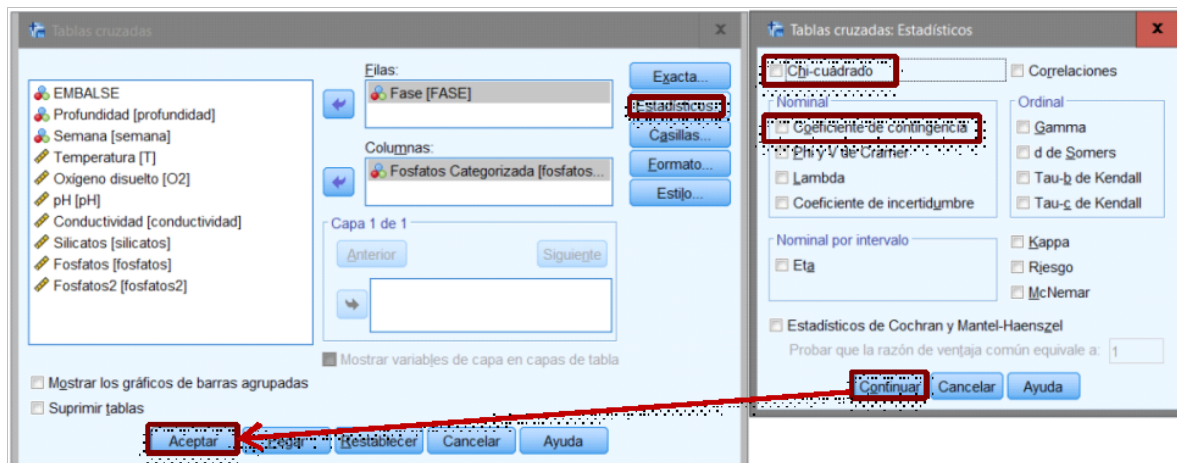
$\alpha = 5\%$ o 1%

3. Escriba la expresión del estadígrafo de contraste:

$$\chi^2_{(1-1)(j-1)} = \sum_i \sum_j \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

4. *¿Cuál es la distribución teórica del estadígrafo de contraste?, ¿Cuál es el valor experimental del estadígrafo de contraste y su p-valor?:*

Ji-cuadrado con 2 [(3-1) (2-1)] grados de libertad.



Los resultados se muestran en la siguiente tabla:

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson ^a	41,337 ^a	2	,000
Probabilidad de Fisher	41,498	1	,000
Asociación lineal por lineal	37,556	1	,000
N de casos válidos	140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 12,00.

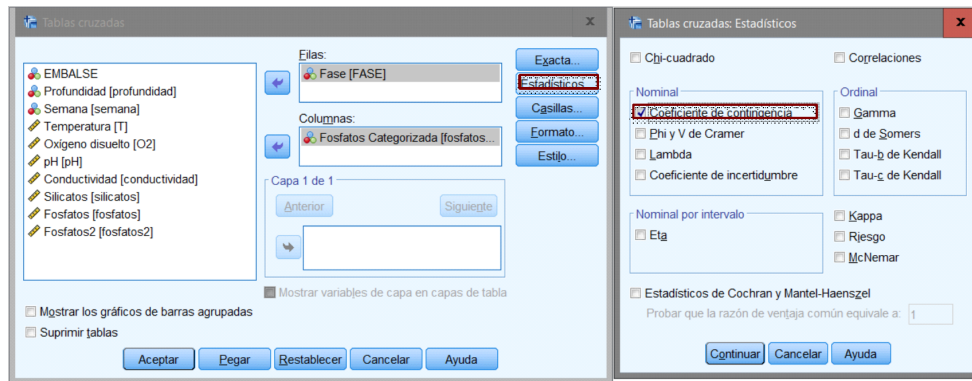
Valor experimental: **41,337** ; p-valor: **0,000**

5. *Explique cuál es la conclusión estadística y experimental.*

Conclusión estadística: *Se rechaza la hipótesis nula con resultados altamente significativos p-valor = 0,000 < α = 0,01.*

Conclusión experimental: *Asumiendo un riesgo de equivocarnos del 1% en relación a la hipótesis de independencia entre las dos variables, podemos decir que "El Nivel de fosfatos está relacionado con la fase en la que se encuentre el embalse".*

6. *¿Existe un alto grado de relación entre las variables?. ¿Cuál es su valor? y ¿Cómo se calcula?:*



¿Existe un alto grado de relación entre las variables?:

Para responder a esta pregunta se utiliza el Coeficiente de contingencia cuya fórmula es

$$CC = \frac{\chi_{exp}^2}{\chi_{exp}^2 + n}; \text{ Valor: } 0,477; \%CC = \frac{CC}{\sqrt{\frac{r-1}{r}}} r = \text{Min}(I, J); \%CC = \frac{0,477}{\sqrt{\frac{1}{2}}} = 0,674 \text{ relación}$$

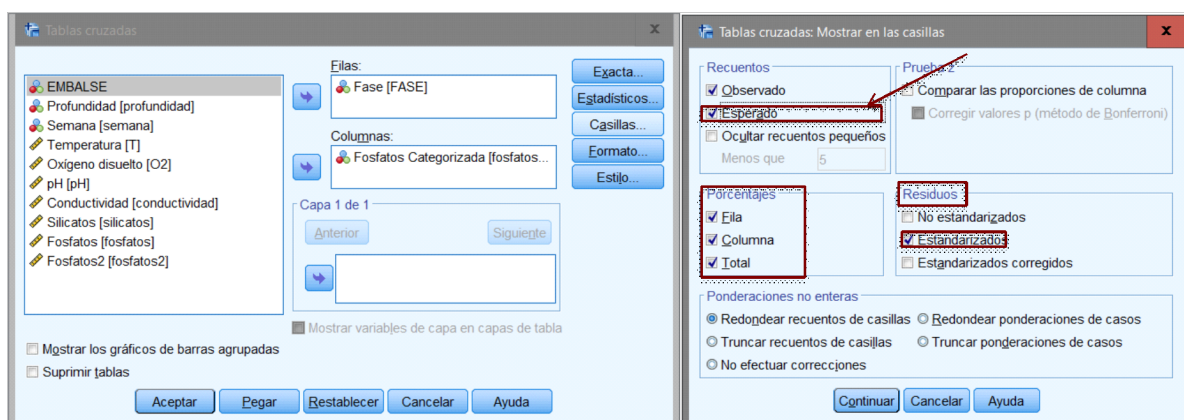
relativamente alta.

Medidas simétricas			
		Valor	Significación aproximada
Nominal por Nominal	Coeficiente de contingencia	,477	,000
N de casos válidos		140	

7. Si las dos variables fueran independientes, ¿cuál sería el nivel de fosfatos bajo esperado en la fase de estratificación? y ¿Cómo se calcula este valor?:

Nivel de fosfatos esperado en la fase de estratificación: **42,5**; ¿Cómo se calcula este

$$\text{valor?: } \hat{f}_{ij} = \frac{f_{ig} \cdot f_{gj}}{f_{gg}} = \frac{70 \cdot 85}{140} = 42,5$$



			Fosfatos Categorizada			Total
			baja	media	alta	
Fase	MEZCLA	Recuento	24	19	27	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	34,3%	27,1%	38,6%	100,0%
		% dentro de Fosfatos Categorizada	28,2%	79,2%	87,1%	50,0%
		% del total	17,1%	13,6%	19,3%	50,0%
		Residuo estandarizado	-2,8	2,0	2,9	
ESTRATIFICACION		Recuento	61	5	4	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	87,1%	7,1%	5,7%	100,0%
		% dentro de Fosfatos Categorizada	71,8%	20,8%	12,9%	50,0%
		% del total	43,6%	3,6%	2,9%	50,0%
		Residuo estandarizado	2,8	-2,0	-2,9	
Total		Recuento	85	24	31	140
		Recuento esperado	85,0	24,0	31,0	140,0
		% dentro de Fase	60,7%	17,1%	22,1%	100,0%
		% dentro de Fosfatos Categorizada	100,0%	100,0%	100,0%	100,0%
		% del total	60,7%	17,1%	22,1%	100,0%

8. ¿Qué celda de la tabla de contingencia es la más responsable de la relación entre el nivel de fosfatos y la fase? y ¿Cuál es su contribución al valor experimental? Especificar cómo se calcula este valor: ¿Qué celda de la tabla de contingencia es la más responsable de la relación entre el nivel de fosfatos y la fase?:

En este caso hay dos celdas cuya relación es máxima [fase de mezcla y nivel alto de fosfatos] y [fase de estratificación y nivel alto de fosfatos]. Se emplea el residuo estandarizado

The image shows two SPSS dialog boxes for a crosstabs analysis. The first dialog, 'Tablas cruzadas', has 'Fase [FASE]' in the 'Filas' field and 'Fosfatos Categorizada [fosfatos...]' in the 'Columnas' field. The second dialog, 'Tablas cruzadas: Mostrar en las casillas', has 'Residuos' checked with 'Estandarizados' selected. The 'Recuentos' section has 'Observado' and 'Esperado' checked. The 'Prueba z' section has 'Comparar las proporciones de columna' checked. The 'Ponderaciones no enteras' section has 'Redondear recuentos de casillas' selected.

			Fosfatos Categorizada			Total
			baja	media	alta	
Fase	MEZCLA	Recuento	24	19	27	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	34,3%	27,1%	38,6%	100,0%
		% dentro de Fosfatos Categorizada	28,2%	79,2%	87,1%	50,0%
		% del total	17,1%	13,6%	19,3%	50,0%
		Residuo estandarizado	-2,8	2,0	2,9	
ESTRATIFICACION	ESTRATIFICACION	Recuento	61	5	4	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	87,1%	7,1%	5,7%	100,0%
		% dentro de Fosfatos Categorizada	71,8%	20,8%	12,9%	50,0%
		% del total	43,6%	3,6%	2,9%	50,0%
		Residuo estandarizado	2,8	-2,0	-2,9	
Total	Total	Recuento	85	24	31	140
		Recuento esperado	85,0	24,0	31,0	140,0
		% dentro de Fase	60,7%	17,1%	22,1%	100,0%
		% dentro de Fosfatos Categorizada	100,0%	100,0%	100,0%	100,0%
		% del total	60,7%	17,1%	22,1%	100,0%

$$\text{Residuo estandarizado: } r_{ij} = \frac{f_{ij} - \hat{f}_{ij}}{\sqrt{\hat{f}_{ij}}}; r_{13} = \frac{27 - 15,5}{\sqrt{15,5}} = 2,9 \text{ y } r_{23} = \frac{4 - 15,5}{\sqrt{15,5}} = -2,9$$

¿Cuál es su contribución al valor experimental? Especificar cómo se calcula este valor:

$$c_{ij} = \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}} = \sqrt{r_{ij}}; c_{ij} = (\pm 2,9)^2 = 8,41$$

9. ¿Qué porcentaje se da en la fase de estratificación y en el nivel bajo de fosfatos?:

43,6%

The image shows two windows from the SPSS software interface. The left window is the 'Tablas cruzadas' dialog box, where 'Fase [FASE]' is selected as the row variable and 'Fosfatos Categorizada [fosfatos...]' is selected as the column variable. The right window is the 'Tablas cruzadas: Mostrar en las casillas' options dialog, where the following options are checked: 'Observado' and 'Esperado' under 'Recuentos'; 'Estandarizados' under 'Residuos'; and 'Fila', 'Columna', and 'Total' under 'Porcentajes'. A red arrow points to the 'Columna' option in the 'Porcentajes' section.

Tabla cruzada Fase 'Fosfatos Categorizada'						
			Fosfatos Categorizada			Total
			baja	media	alta	
Fase	MEZCLA	Recuento	24	19	27	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	34,3%	27,1%	38,6%	100,0%
		% dentro de Fosfatos Categorizada	28,2%	79,2%	87,1%	50,0%
		% del total	17,1%	13,6%	19,3%	50,0%
		Residuo estandarizado	-2,8	2,0	2,9	
ESTRATIFICACION	Recuento	61	5	4	70	
	Recuento esperado	42,5	12,0	15,5	70,0	
	% dentro de Fase	87,1%	7,1%	5,7%	100,0%	
	% dentro de Fosfatos Categorizada	71,8%	20,8%	12,9%	50,0%	
	% del total	43,6%	3,6%	2,9%	50,0%	
	Residuo estandarizado	2,8	-2,0	-2,9		
Total	Recuento	85	24	31	140	
	Recuento esperado	85,0	24,0	31,0	140,0	
	% dentro de Fase	60,7%	17,1%	22,1%	100,0%	
	% dentro de Fosfatos Categorizada	100,0%	100,0%	100,0%	100,0%	
	% del total	60,7%	17,1%	22,1%	100,0%	

10. ¿Qué porcentaje de la fase de estratificación tiene un nivel de fosfatos bajo? y ¿Qué porcentaje de nivel de fosfatos medio hay en la fase de mezcla?:

87,1% y 79,2%

Tabla cruzada Fase 'Fosfatos Categorizada'						
			Fosfatos Categorizada			Total
			baja	media	alta	
Fase	MEZCLA	Recuento	24	19	27	70
		Recuento esperado	42,5	12,0	15,5	70,0
		% dentro de Fase	34,3%	27,1%	38,6%	100,0%
		% dentro de Fosfatos Categorizada	28,2%	79,2%	87,1%	50,0%
		% del total	17,1%	13,6%	19,3%	50,0%
		Residuo estandarizado	-2,8	2,0	2,9	
ESTRATIFICACION	Recuento	61	5	4	70	
	Recuento esperado	42,5	12,0	15,5	70,0	
	% dentro de Fase	87,1%	7,1%	5,7%	100,0%	
	% dentro de Fosfatos Categorizada	71,8%	20,8%	12,9%	50,0%	
	% del total	43,6%	3,6%	2,9%	50,0%	
	Residuo estandarizado	2,8	-2,0	-2,9		
Total	Recuento	85	24	31	140	
	Recuento esperado	85,0	24,0	31,0	140,0	
	% dentro de Fase	60,7%	17,1%	22,1%	100,0%	
	% dentro de Fosfatos Categorizada	100,0%	100,0%	100,0%	100,0%	
	% del total	60,7%	17,1%	22,1%	100,0%	

Anexo II

GUÍA DE APRENDIZAJE DE LAS PRÁCTICAS CON SPSS DE LAS ASIGNATURAS DE ESTADÍSTICA DE LA FACULTAD DE QUÍMICAS (II)

Puesta en práctica:

Trabajo autónomo del alumno

UNIVERSIDAD DE SALAMANCA

PROYECTO DE INNOVACIÓN DOCENTE: **ID2015/0063**

Coordinadora:

M^a José Fernández Gómez

Miembros del Equipo:

Inmaculada Barrera Mellado

Javier Martín Vallejo

Santiago Vicente Tavera

Rosa Amanda Sepúlveda Correa

Jesús Martín Rodríguez

INTRODUCCIÓN

Imagine que usted trabaja en una compañía alimentaria donde se pretende producir una marca especial de cereales. El gerente de la empresa le manda realizar un estudio los cereales que se venden en el mercado para poder conocer mejor este mercado y fabricar un tipo de cereales que pueda competir con los que ya llevan mucho tiempo en venta al público.

Para ello, le proporcionan un archivo donde se han elegido al azar 76 marcas de cereales donde se estudian características relacionadas con la composición nutritiva, en términos de oligoelementos, proteínas y grasas. También se tiene en cuenta las calorías, si están enriquecidos con vitaminas (Bajo=0, Medio=1, Alto=2). Por último se ha considerado el precio de los cereales de forma cualitativa (0=Baratos, 1=Caros) y si se toman fríos (C) con leche en los desayunos o calientes (H) como ingredientes de comidas.

En el archivo "Cereal.sav", aparece la información de los cereales con las siguientes variables

Nombre de la marca del cereal, **Empresa** que fabrica la marca de cereal, **Mfr**: la primera inicial de cada empresa, **Consumo**: si se toman en frío (C) o en caliente (H), **Calorias**, **Proteinas**: número de proteínas que contienen, **Grasa (gr)**, **Sodio (mgr)**, **Fibra(gr)**: fibra(gr) **Carbohidratos complejos(gr)**, **Azúcar (gr)** **Caloriasgrasa**, **Potasio (mgr)** **Enriquecidos**: si están enriquecidos con vitaminas, **Fibrarec**: fibra recodificada en baja, media, alta, **Precio**. Además se han considerado dos variables nuevas donde se pretende comprobar si los valores que dicen que son en términos medios correctos o no. Para ello la propia empresa realiza sus mediciones del sodio (**ValoraciónSodio**) y del potasio (**ValoraciónPotasio**).



INFORMACIÓN NUTRICIONAL Valores medios aproximados				
	Por 100g	Por galleta (12,5g)	%CDO por galleta	%CDO por 4 galletas (50g)
Valor energético	439 kcal 1844 kJ	55 kcal 231 kJ	3 %	12 %
Proteinas	8,5 g	1,1 g	2 %	8 %
Hidratos de carbono	65 g	8,1 g	3 %	12 %
de los cuales: azúcares	29 g	3,6 g	4 %	16 %
Grasas	16 g	2 g	3 %	12 %
de las cuales: saturadas	4,3 g	0,5 g	3 %	12 %
Fibra alimentaria	6,8 g	0,9 g	3 %	12 %
Sodio	0,26 g	0,03 g	1 %	4 %
Vitaminas y Minerales				
	Por 100g	**%CDR por 100g	Por galleta (12,5g)	%CDO por galleta
Vitamina E	4,70 mg	39 %	0,59 mg	5 %
Vitamina B1	0,507 mg	46 %	0,06 mg	5 %
Vitamina B3	4,02 mg	25 %	0,50 mg	3 %
Magnesio	126 mg	34 %	15,75 mg	4 %
Hierro	7,70 mg	55 %	0,96 mg	7 %

*CDO - Cantidad Diaria Orientativa, que representa la cantidad aproximada de calorías y nutrientes para un adulto medio, calculada sobre la base de una dieta de 2.000 kcal.

**CDR - Cantidad diaria recomendada.

Para realizar el informe conteste a los siguientes bloques de prácticas:

PRÁCTICA 1 – Estadística Descriptiva

1. Qué variables son cuantitativas y cuales cualitativas.
2. Haga una distribución de frecuencias de las variables cualitativas y elija un gráfico para presentar la información. Comente los resultados.
3. Realice una descriptiva básica de las variables sodio y total de carbohidratos (Totcarbo). En la descriptiva básica elija las medidas (y/o gráficos que no sean box-plot) que crea que son importantes para saber cómo se comportan los datos originales. Comente los resultados obtenidos.
4. Haga los box-plot para ambas variables. Comente los resultados.
5. Qué variable de ambas presenta mayor variabilidad. Discuta los resultados.
6. Realice los análisis descriptivos (con estadísticos y gráficos) para las variables calorías, proteínas, grasa para cada industria por separado y discuta las diferencias del tipo cereales fabricados por dichas industrias en función de esas variables.

PRÁCTICA 2 – Intervalos de confianza

1. Calcular el intervalo de confianza para las calorías a un nivel de confianza del 95% e interpretar el resultado.
2. Realizar de nuevo la pregunta anterior pero con un nivel de significación del 99%.
3. Calcule los intervalos de confianza para los carbohidratos complejos para los cereales con un precio bajo y alto a un nivel de confianza del 95%. Interprete los resultados.
4. Represente los intervalos de confianza en un gráfico. Podría decir a la vista de los intervalos obtenidos según el precio que hay diferencias a nivel poblacional entre caros y baratos.
5. Calcule los intervalos de confianza al 90% para el sodio para los distintos grupos definidos por la variable Fiber Gr Interprete todos los intervalos y explique porque los intervalos tienen diferentes amplitudes. ¿Considera que todos los grupos tienen la misma media poblacional?.
6. Realice el box-plot para la variable sodio. ¿Considera que los datos pueden proceder de una población normal?

PRÁCTICA 3 – Contrastes de Hipótesis.

1. Se quiere contrastar si la cantidad de sodio de que llevan incorporados los cereales está por encima de lo que la OMS considera saludable y cuyo valor lo ha fijado en 200 mg. Se supone que el sodio sigue una distribución Normal. Conteste a las siguientes preguntas:

- H_0
- H_a
- Niveles de significación
- *Escriba la expresión del estadígrafo de contraste . Justifique la elección del estadígrafo de contraste y escriba el valor experimental del estadígrafo de contraste en la muestra p-valor*
- *Conclusión estadística*
- *Conclusión experimental.*

2. Contraste si efectivamente el sodio presenta una distribución Normal. Utilice el test de Shapiro-Wilks

- H_0
- H_a
- Niveles de significación
- *Valor del estadígrafo de contraste:*
- *p-valor*
- *Conclusión estadística*

3. Ahora queremos contrastar la cantidad de sodio es diferente en los cereales cuyo precio es considerado bajo con relación a los considerados caros. Plantee todos los pasos del contraste

3.1.- Analice si la variable estudiada sigue una distribución Normal. Presente todos los pasos del contraste.

3.2.- En función del resultado anterior conteste a la pregunta de si hay diferencias en la cantidad de sodio en función del precio de los cereales.

Conteste a cada uno de estas preguntas.

- H_0
- H_a
- Niveles de significación
- *Escriba la expresión del estadígrafo de contraste. Justifique la elección del estadígrafo de contraste y escriba el valor experimental del estadígrafo de contraste en la muestra.*
- *p-valor*
- *Conclusión estadística*
- *Conclusión experimental.*

4. Se pretende conocer si las valoraciones de Sodio en términos medios son correctas comparándolas con las valoraciones obtenidas por la compañía.

4.1.- Analice si la variable estudiada sigue una distribución Normal (Shapiro-Wilks). Presente todos los pasos del contraste.

4.2.- En función del resultado anterior conteste a la pregunta de si hay diferencias en la cantidad de sodio entre las dos valoraciones.

- H_0
- H_a
- Niveles de significación
- *Escriba la expresión del estadígrafo de contraste. Justifique la elección del estadígrafo de contraste y escriba el valor experimental del estadígrafo de contraste en la muestra.*
- *p-valor*
- *Conclusión estadística*
- *Conclusión experimental.*

5. Realice y replantee el mismo contraste anterior pero en el caso del potasio.

5.1.- Analice si la variable estudiada sigue una distribución Normal (Shapiro-Wilks). Presente todos los pasos del contraste.

5.2.- En función del resultado anterior conteste a la pregunta de si hay diferencias en la cantidad de potasio entre las dos valoraciones.

- H_0
- H_a
- Niveles de significación
- *Escriba la expresión del estadígrafo de contraste. Justifique la elección del estadígrafo de contraste y escriba el valor experimental del estadígrafo de contraste en la muestra.*
- p -valor
- Conclusión estadística
- Conclusión experimental.

6. Contraste si el azúcar sigue una distribución Normal en cereales que se toman en frío y en caliente.

FRIOS

- H_0
- H_a
- Niveles de significación
- *Valor experimental del estadígrafo de contraste:*
- p -valor
- Conclusión estadística
- Conclusión experimental.

CALIENTES

- H_0
- H_a
- Niveles de significación
- *Valor experimental del estadígrafo de contraste:*
- p -valor
- Conclusión estadística
- Conclusión experimental.

7. A partir de los resultados obtenidos anteriormente, ¿qué contraste elegiría para ver si hay diferencias en la cantidad de azúcar entre los cereales que se toman en frío y en caliente? Realice el contraste con los pasos correspondientes.

- H_0
 - H_a
 - *Niveles de significación*
 - *Escriba la expresión del estadígrafo de contraste. Justifique la elección del estadígrafo de contraste y escriba el valor experimental del estadígrafo de contraste en la muestra.*
 - *p-valor*
 - *Conclusión estadística*
 - *Conclusión experimental.*
-

PRÁCTICA 4 – ANOVA.

1. Ahora se pretende conocer si hay diferencias entre tres de enriquecimiento de los cereales (Bajo, Medio, Alto), con relación a la variable calorías grasa. Para ello rellene los siguientes apartados.

- H_0
- H_a
- Niveles de significación
- Defina el Estadígrafo de contraste y su distribución.
- Valor experimental del estadígrafo de contraste para la muestra analizada. Escriba la tabla típica de este análisis.
- p -valor
- Conclusión estadística
- Conclusión experimental.

2. Si ha encontrado diferencias en el apartado anterior. ¿A que son debidas esas diferencias?. Plantee el contraste de forma general

- H_0
- H_a
- Niveles de significación
- Defina el Estadígrafo de contraste y su distribución.
- Valores experimentales para los distintos contrastes y su p -valores.
- Conclusión estadística
- Conclusión experimental.

3. ¿Qué contraste ha utilizado y por qué?

4. ¿Qué penalización ha utilizado y por qué.

5. Se han cumplido Los supuestos de partida para hacer este análisis. En que se basa para responder a esta pregunta. ¿Qué test ha utilizado?. ¿Cuáles han sido los resultados de las pruebas?

PRÁCTICA 5 – REGRESIÓN Y CORRELACIÓN.

Se pretende estudiar las asociaciones entre diferentes variables recogidas para las marcas de cereales.

1.- Estudiar si hay relación lineal y su grado entre las variables Grasa, sodio, azúcar, calorías/grasa, potasio. Qué coeficiente ha utilizado para contestar a esta pregunta.

2.- Calcular la recta de regresión para explicar la cantidad de fibra a partir de la cantidad de Potasio.

3.- Interprete el valor de b (pendiente) y el valor de a (ordenada en el origen) de la recta.

4.-¿Qué porcentaje de variaciones en el tamaño del tanque de combustible son explicadas por el peso del coche. Qué coeficiente ha utilizado

6.-Considera que el poder predictivo del modelo es bueno. ¿Qué información necesita para responder a esta pregunta?.

7.- Calcule otros modelos (parabólico, exponencial, potencial, logarítmico) y compárelos con el modelo lineal.

PRÁCTICA 6 – TABLAS DE CONTINGENCIA.

Se pretende conocer si hay asociación entre el grado de enriquecimiento vitamínico y el precio de los cereales. Para ello conteste a las siguientes preguntas.

1.- ¿Qué contraste debería realizar para contestar esta pregunta?

2.-Rellene las siguiente cuestiones.

- Ho:
- Ha:
- nivel de significación
- Estadígrafo de contraste y distribución bajo la hipótesis nula cierta.
- Valor experimental
- Conclusión estadística.
- Conclusión experimental.

3.- De los cereales que presentan un enriquecimiento más bajo qué porcentaje de ellos presentan valores altos.

4.- Qué porcentaje de cereales presentan un enriquecimiento medio y precio alto.

5.- De los cereales con precio caro cuál es la distribución porcentual para la variable enriquecimiento.

6.-Si el contraste ha sido significativo busque cuáles son las combinaciones de categorías de ambas variables con mayor influencia en la asociación. ¿Qué información ha utilizado para ello y por qué?.

7.- Por último, indique el grado de asociación entre las variables- ¿Qué coeficiente ha utilizado para ello?
