

ANÁLISIS DE DATOS  
2º de Biología — curso 2017/18  
MODELO DE DISEÑO DE EXPERIMENTOS (UN FACTOR)

1.- Se quiere comparar la capacidad pulmonar (en ml/kg de peso) en niños, adultos y ancianos, obteniéndose los siguientes resultados:

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| Niños    | 8,4 | 7,6 | 7,9 | 8,0 | 8,1 |
| Adultos  | 8,7 | 8,1 | 8,5 | 8,2 | 8,0 |
| Ancianos | 7,4 | 7,8 | 7,3 | 7,6 | 8,0 |

Hacer un estudio completo.

2.- Las precipitaciones caídas en un país han disminuido de manera preocupante durante el último año. Antes de tomar ninguna medida se decide hacer un estudio previo para saber si el descenso de las lluvias se produjo de forma homogénea. Para ello se seleccionan aleatoriamente cinco estaciones meteorológicas en cada una de las cuatro regiones del país, obteniéndose los siguientes porcentajes de disminución de las precipitaciones en cada una de ellas:

| Región Este | Región Norte | Región Oeste | Región Sur |
|-------------|--------------|--------------|------------|
| 10,4        | 12,8         | 11,2         | 13,9       |
| 12,8        | 14,2         | 9,8          | 14,2       |
| 15,6        | 16,3         | 10,7         | 12,8       |
| 9,2         | 10,1         | 6,3          | 15,0       |
| 8,7         | 12,0         | 12,4         | 13,7       |

- Plantear claramente todos los elementos y las hipótesis del modelo para comparar los porcentajes de disminución de las precipitaciones en las 4 regiones.
- ¿En qué zona parecen haber disminuido más las precipitaciones?
- Obtener la tabla ANOVA y contrastar la hipótesis nula de que las medias de disminución del porcentaje de lluvias en el país fueron las mismas en las cuatro regiones (tomar  $\alpha = 0,10$ ).
- Comparar las medias de las diferentes regiones de dos en dos, con un nivel de confianza global del 90 %.

3.- En un bosque próximo a una incineradora los árboles no crecen con normalidad. Se piensa que unos nuevos abonos americanos y australianos pueden ser la solución. Para ver si esta medida es efectiva, se utiliza el abono americano en un tercio de los árboles, el abono australiano en otro tercio, y para el tercio restante no se utiliza ningún abono. Después de 3 meses se han obtenido los siguientes resultados sobre el crecimiento en centímetros de 60 árboles en total:

$$\begin{aligned} \bar{y}_1 &= 6,57 & s_1^2 &= 0,70 \\ \bar{y}_2 &= 5,30 & s_2^2 &= 0,65 \\ \bar{y}_3 &= 3,20 & s_3^2 &= 0,50 \end{aligned}$$

¿Se puede afirmar que se obtienen diferencias en los resultados, con un nivel de significación 0,01? En caso necesario, efectuar una comparación de los crecimientos medios, con un nivel de significación conjunto de 0,15.

4.- En un estudio sobre la efectividad de los métodos para dejar de fumar se quiere saber si la reducción media en el número de cigarrillos diarios difiere de un método a otro entre hombres fumadores. Para ello se hace un experimento con 12 fumadores que consumían 60 cigarrillos diarios. Se aplica cada uno de los métodos a 4 de ellos, seleccionados aleatoriamente. El número de cigarrillos que deja de fumar cada individuo es:

|            |    |    |    |    |
|------------|----|----|----|----|
| Método I   | 50 | 51 | 51 | 52 |
| Método II  | 41 | 40 | 39 | 40 |
| Método III | 49 | 47 | 45 | 47 |

- Indicar claramente todos los elementos y las hipótesis del modelo para comparar la disminución de consumo de cigarrillos conseguidos con los tres métodos.
- Contrastar, mediante el análisis de la varianza, si la reducción media en el número de cigarrillos es similar para los tres métodos con un nivel de significación  $\alpha = 0,05$ .
- Obtener los intervalos de confianza para la diferencia entre las medias, con un nivel de confianza conjunto de 0,95. ¿Entre qué métodos se aprecian diferencias significativas?

5.- A continuación se muestran los datos recogidos en las inspecciones de cuatro gasolineras elegidas aleatoriamente. Los valores de la tabla reflejan los mililitros que faltan para completar un litro en distintas mediciones sobre el mismo surtidor de cada gasolinera.

|            |   |       |       |       |       |       |       |       |       |
|------------|---|-------|-------|-------|-------|-------|-------|-------|-------|
| Gasolinera | C | 17,80 | 18,00 | 17,98 | 18,20 | 18,00 | 17,99 | 18,10 | 17,90 |
|            | R | 18,01 | 17,75 | 18,00 | 17,77 | 18,01 | 18,01 | 18,12 | 18,20 |
|            | S | 18,10 | 17,92 | 18,01 | 17,88 | 18,30 | 18,22 | 18,56 | 18,10 |
|            | V | 18,05 | 18,01 | 17,94 | 18,23 | 18,20 | 18,00 | 17,84 | 18,11 |

Contrastar la hipótesis nula de que la cantidad media de gasolina que se sirve por litro no depende de la gasolinera (tomar  $\alpha = 0,05$ ).

6.- En un estudio sobre la incidencia del cáncer de garganta se analizan los resultados obtenidos en 10 ciudades de más de 100 000 habitantes, en otras 10 ciudades de menos de 100 000 habitantes, y en 20 pueblos. En cada población se registra la variable Y = «Número de casos detectados por cada 100 000 habitantes». Los resultados obtenidos se resumen a continuación:

|                   | $\bar{y}$ | $s^2$ |
|-------------------|-----------|-------|
| Ciudades grandes  | 120       | 105,2 |
| Ciudades pequeñas | 110       | 95,3  |
| Pueblos           | 70        | 101,8 |

- ¿Aportan estos datos evidencia estadística de que el hábitat influye en la incidencia del cáncer de garganta? Dar una respuesta razonada, con una confianza del 95 %, indicando el modelo y la metodología estadística empleada.
- Suponiendo que la varianza es la misma en cada uno de los tipos de población, dar una estimación insesgada de dicha varianza.
- Hallar intervalos de confianza para estimar simultáneamente las diferencias entre incidencias medias de la enfermedad en los tres tipos de población, con una confianza conjunta del 95 %. ¿Qué conclusiones estadísticas se pueden obtener?

7.- Se lleva a cabo un estudio para comparar las longitudes medias (en cm) de tres colonias de rata almizclera negra, situadas en Delaware, Maryland y Virginia. A continuación, se muestran algunos resultados del estudio:

| Estado   | N  | Media | Cuasi desviación típica |
|----------|----|-------|-------------------------|
| Delaware | 10 | 30,6  | 2,1                     |
| Maryland | 6  | 33,6  | 1,2                     |
| Virginia | 10 | 31,5  | 1,8                     |
| Total    | 26 | 31,6  | 2,1                     |

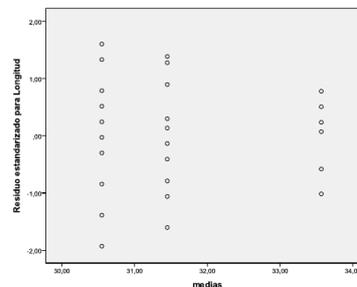
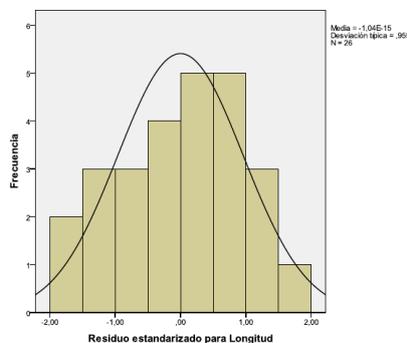
PRUEBA DE HOMOGENEIDAD DE VARIANZAS  
Longitud de la rata almizclera

| Estadístico de Levene | gl1 | gl2 | Sig. |
|-----------------------|-----|-----|------|
| ,854                  | 2   | 23  | ,439 |

### ANOVA

Longitud de la rata almizclera

|              | Suma de cuadrados | gl | Media cuadrática | F     | Sig. |
|--------------|-------------------|----|------------------|-------|------|
| Inter-grupos | 34,455            | 2  | 17,228           | 5,091 | ,015 |
| Intra-grupos | 77,823            | 23 | 3,384            |       |      |
| Total        | 112,278           | 25 |                  |       |      |



- Indicar el modelo utilizado para analizar estos datos y el significado de cada uno de sus elementos. Indicar los requisitos previos necesarios en el modelo utilizado y comentar su cumplimiento en base a la información aportada.
- ¿Qué conclusión se obtiene de la tabla ANOVA con nivel de significación 0,05? Indicar claramente las hipótesis nula y alternativa.
- Comparar todas las colonias de dos en dos, mediante los intervalos de confianza correspondientes y con un nivel de significación conjunto de 0,06.

### MODELO DE DISEÑO DE EXPERIMENTOS (VARIOS FACTORES)

1.- Se quiere estudiar la producción de fresa que se obtiene con diferentes variedades. La producción obtenida con 3 variedades y en 4 tipos de suelo diferentes, se ofrece a continuación:

| Variedades | Tipos de suelo |      |     |      |
|------------|----------------|------|-----|------|
|            | 1              | 2    | 3   | 4    |
| 1          | 6,3            | 6,9  | 5,3 | 6,2  |
| 2          | 10,1           | 10,8 | 9,8 | 10,5 |
| 3          | 8,4            | 9,4  | 9,0 | 9,2  |

Haz un estudio completo.

2.- En un estudio sobre el consumo de gasolina de distintos coches se realiza el siguiente experimento: se toman cuatro coches al azar de un fabricante español, cuatro de un francés, cuatro de un alemán, y cuatro de un japonés. Se prueba un coche de cada fabricante en una gran ciudad durante la hora punta, otro en ciudad fuera de la hora punta, otro se prueba en carretera de montaña y el otro en una carretera llana. El consumo en litros de gasolina por cada 100 kilómetros es:

|         | Ciudad (h. punta) | Ciudad (h. normal) | Carretera montaña | Carretera llana |
|---------|-------------------|--------------------|-------------------|-----------------|
| Español | 14,7              | 9,4                | 7,2               | 6,8             |
| Francés | 11,6              | 7,7                | 6,8               | 6,0             |
| Alemán  | 10,8              | 7,2                | 7,2               | 6,4             |
| Japonés | 16,0              | 10,0               | 9,3               | 7,7             |

- Plantear el modelo adecuado para estudiar el consumo de gasolina con dos factores.
- ¿Qué modelo de coche parece que consume más y qué modelo de coche parece que consume menos? ¿En qué condiciones parece que se consume más y en qué condiciones parece que se consume menos?
- Obtener la tabla de análisis de la varianza y decidir si el modelo de coche tiene una influencia significativa sobre el consumo (al nivel de significación 0,05).
- Comparar de dos en dos el consumo medio de los cuatro modelos de coche, con un nivel de confianza conjunto del 95 %. ¿Conclusiones?
- Finalmente, analizar los datos sin tener en cuenta las condiciones en que se conducen los coches, es decir, realizando un análisis de la varianza con un solo factor. Con este modelo, ¿influye el modelo de coche en el consumo de gasolina? ¿Coincide esta conclusión con la obtenida anteriormente? ¿Cuál sería el modelo adecuado y la conclusión correcta? ¿Por qué?

3.- Se quiere hacer un estudio de comparación pluviométrica entre 5 ciudades de una misma región. Para esto, se mide la lluvia recogida en esas 5 ciudades en 4 meses diferentes:

|          | Enero | Abril | Julio | Octubre | $y_i$            |
|----------|-------|-------|-------|---------|------------------|
| Ciudad A | 11    | 16    | 10    | 17      | 13,50            |
| Ciudad B | 9     | 9     | 8     | 12      | 9,50             |
| Ciudad C | 12    | 9     | 9     | 10      | 10,00            |
| Ciudad D | 11    | 10    | 10    | 12      | 10,75            |
| Ciudad E | 19    | 18    | 20    | 14      | 17,75            |
| $y_{.i}$ | 12,4  | 12,4  | 11,4  | 13,0    | $y_{..} = 12,30$ |

Además, la variabilidad total de los datos es  $SCT = 262,2$ .

- Describir el modelo y sus hipótesis.
  - Construir la tabla ANOVA y contrastar, con nivel de significación 0,05, la hipótesis nula de que no hay diferencias pluviométricas entre las cinco ciudades.
  - Construir un intervalo de confianza al 95 % para la diferencia media de lluvia recogida entre las ciudades A y E. Con nivel de significación 0,05, ¿existe evidencia para rechazar que estas dos ciudades son iguales?
- 4.- En 12 grandes ciudades se hace un estudio sobre la tasa de contaminación (con cierto contaminante atmosférico). Se piensa que puede estar influida por dos factores:
- «Índice de pluviosidad», que actúa a 2 niveles (baja o alta pluviosidad).
  - «Densidad de industria contaminante», que actúa a 3 niveles (baja, media o alta densidad).

Se obtienen 2 réplicas para cada posible combinación de los niveles de los dos factores. Las tasas medias de contaminación de las dos réplicas se ofrecen a continuación:

|                  | Densidad baja | Densidad media | Densidad alta |
|------------------|---------------|----------------|---------------|
| Baja pluviosidad | 837,5         | 868            | 887           |
| Alta pluviosidad | 420,5         | 437            | 526,5         |

La variabilidad total viene dada por  $SCT = 513632,25$ .

Especificar modelo, hipótesis y obtener la tabla de análisis de la varianza adecuada para contestar, razonadamente, a las siguientes preguntas, al nivel de significación 0,05: ¿Existe interacción significativa entre los 2 factores? ¿Influye apreciablemente el índice de pluviosidad en la tasa de contaminación? ¿Influye la densidad de industria contaminante?

- 5.- Consideramos nuevamente los datos del ejercicio de las gasolineras del capítulo anterior, pero teniendo ahora en cuenta que los cuatro primeros datos para cada gasolinera, se tomaron inmediatamente después de adquirir los surtidores, y los cuatro últimos, 6 meses más tarde
- Con esta nueva información, plantear un modelo de diseño de experimentos de dos factores con posible interacción, y obtener la tabla ANOVA correspondiente. A partir de esta tabla, decidir si se podría simplificar el modelo (tomando  $\alpha = 0,05$ ).
  - Simplificar el modelo (si parece adecuado), obtener la tabla ANOVA para el modelo simplificado, y comentar los resultados.

6.- A fin de comparar dos productos distintos para la reducción del colesterol en pacientes con hipercolesterolemia benigna, se lleva a cabo un ensayo clínico durante 3 semanas con 30 de estos pacientes:

Un grupo de 10 pacientes siguió un tratamiento a base de estatinas. De este grupo, la mitad siguieron una dieta estricta, y la otra mitad una dieta moderada.

Otro grupo de 10 pacientes siguió un tratamiento a base de fibratos. De este grupo, la mitad siguieron una dieta estricta, y la otra mitad una dieta moderada.

Un último grupo de 10 pacientes recibió placebo. De este grupo, la mitad siguieron una dieta estricta, y la otra mitad una dieta moderada.

Al concluir las tres semanas, se anotó la reducción de colesterol para cada paciente (en mg/dl), y los datos obtenidos fueron:

|                | Estatinas | Fibratos | Placebo |
|----------------|-----------|----------|---------|
| Dieta estricta | 32        | 22       | 1       |
|                | 35        | 25       | 5       |
|                | 41        | 28       | 9       |
|                | 45        | 33       | 11      |
|                | 48        | 38       | 18      |
| Dieta moderada | 25        | 22       | 0       |
|                | 33        | 24       | 4       |
|                | 40        | 31       | 9       |
|                | 42        | 32       | 13      |
|                | 46        | 40       | 15      |

a. Plantear con detalle todos los elementos y las hipótesis del modelo de diseño de experimentos con dos factores y posible interacción para analizar estos datos.

b. Decidir, a partir de los gráficos de residuos, si las hipótesis del modelo son aceptables, explicando las respuestas.

c. Para un nivel de significación del 5 %, ¿el producto empleado o la dieta seguida son factores relevantes para explicar la posible reducción del colesterol? ¿Qué podemos decir sobre la interacción?

d. Como consecuencia de lo obtenido en el apartado anterior, simplificar el modelo todo lo posible, obtener la tabla ANOVA para el modelo simplificado, y sacar las conclusiones pertinentes (de nuevo, al 5 %).

7.- Para estudiar el efecto de la iluminación (A=natural, B=muy fuerte, C=escasa) en la velocidad de lectura se realiza un experimento. Se mide el número de palabras leídas en un minuto para distintos tipos de papel y tamaño de letra. Los resultados que se obtienen son los siguientes:

|              | Papel satinado | Papel blanco | Papel color |
|--------------|----------------|--------------|-------------|
| Letra grande | 258<br>A       | 230<br>C     | 240<br>B    |
|              | 235<br>B       | 270<br>A     | 240<br>C    |
| Letra normal | 220<br>C       | 225<br>B     | 260<br>A    |

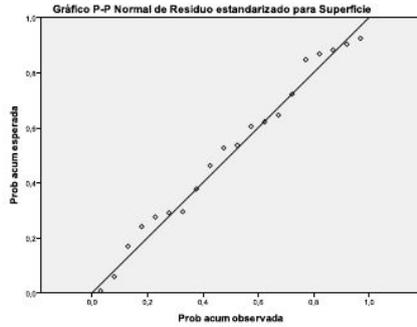
¿Cuántos factores se consideran en el experimento? Construir con SPSS la tabla de análisis de la varianza y contrastar, con un nivel de significación  $\alpha = 0,05$ , si los factores afectan a la velocidad de lectura.

8.- Un fisiólogo vegetal investiga el efecto del estrés mecánico (agitarlas durante 20 minutos dos veces al día) y del nivel de luz en el crecimiento de 20 plantas de soja. Se asignaron 5 plantas, al azar, a cada combinación del tipo de luz (baja y moderada) con el tipo de estrés (sin y con estrés) Después de 16 días de crecimiento, se midió la superficie de las hojas de cada planta (en  $\text{cm}^2$ ) Tras descartar la existencia de interacción entre el estrés y la luz, se obtuvieron los siguientes resultados:

#### ESTADÍSTICOS DESCRIPTIVOS

Variable dependiente: Superficie de las hojas de soja

| Estrés     | Nivel de luz | Media  | Desviación típica | N  |
|------------|--------------|--------|-------------------|----|
| Sin estrés | Luz baja     | 234,40 | 30,221            | 5  |
|            | Luz moderada | 316,60 | 15,159            | 5  |
|            | Total        | 275,50 | 48,836            | 10 |
| Con estrés | Luz baja     | 207,00 | 18,152            | 5  |
|            | Luz moderada | 272,20 | 36,697            | 5  |
|            | Total        | 239,60 | 43,884            | 10 |
| Total      | Luz baja     | 220,70 | 27,584            | 10 |
|            | Luz moderada | 294,40 | 35,331            | 10 |
|            | Total        | 257,55 | 48,797            | 20 |



**Contraste de Levene sobre la igualdad de las varianzas error**

Variable dependiente: Superficie de las hojas de soja

| F     | gl1 | gl2 | Sig. |
|-------|-----|-----|------|
| 1,833 | 3   | 16  | ,182 |

**Pruebas de los efectos inter-sujetos**

Variable dependiente: Superficie de las hojas de soja

| Origen           | Suma de cuadrados tipo III | gl | Media cuadrática | F        | Sig. |
|------------------|----------------------------|----|------------------|----------|------|
| Modelo corregido | 33602,500 <sup>a</sup>     | 2  | 16801,250        | 24,541   | ,000 |
| Intersección     | 1326640,050                | 1  | 1326640,050      | 1937,791 | ,000 |
| Estres           | 6444,050                   | 1  | 6444,050         | 9,413    | ,007 |
| Luz              | 27158,450                  | 1  | 27158,450        | 39,670   | ,000 |
| Error            | 11638,450                  | 17 | 684,615          |          |      |
| Total            | 1371881,000                | 20 |                  |          |      |
| Total corregida  | 45240,950                  | 19 |                  |          |      |

a. R cuadrado = ,743 (R cuadrado corregida = ,712)

- Indica el modelo que se ha utilizado, detallando sus elementos y el diagnóstico razonado de los requisitos previos.
- Estima (según este modelo) el efecto adicional que tiene ser sometida a estrés sobre la superficie media de las hojas.
- Justifica si el estrés o el nivel de luz tienen una influencia significativa (al 1 %) sobre la superficie, planteando correctamente los contrastes necesarios.
- Completa la tabla ANOVA que se habría obtenido a partir de los mismos datos, si hubiéramos utilizado un modelo de diseño de experimentos solamente con el factor estrés.

| Origen | Suma de cuadrados | gl | Media cuadrática | F | Sig.  |
|--------|-------------------|----|------------------|---|-------|
|        |                   |    |                  |   | 0,101 |
|        |                   |    |                  |   |       |
|        |                   |    |                  |   |       |

9.- Un laboratorio de medición atmosférica ha adquirido un nuevo equipo para medir ozono. Para evaluar si el nuevo equipo está calibrado, se realiza un pequeño experimento en 5 observatorios diferentes. En cada observatorio se toma una medida con el nuevo equipo (que llamaremos B) y otra con el equipo antiguo (que llamaremos A), obteniéndose los siguientes resultados (en  $\text{mg}/\text{m}^3$ ):

|          | Obs. 1 | Obs. 2 | Obs. 3 | Obs. 4 | Obs. 5 |
|----------|--------|--------|--------|--------|--------|
| Equipo A | 215    | 305    | 247    | 221    | 286    |
| Equipo B | 224    | 312    | 251    | 232    | 295    |

Teniendo en cuenta que la suma de cuadrados totales (SCT) es 12 491,6, proponer un modelo adecuado para explicar los niveles de ozono que se han observado en el experimento y contrastar si existen diferencias significativas entre los dos equipos, con nivel de confianza 0,95.

10.- En una investigación de laboratorio se emplean cámaras de crecimiento para estudiar el desarrollo de ciertos microorganismos cuando se varían las concentraciones de  $\text{CO}_2$  (baja y alta), y la temperatura (baja, media y alta). En distintas cámaras se cruzan todos los niveles de los dos factores y se obtienen tres réplicas completas del experimento. La siguiente tabla muestra los crecimientos medios que se obtienen para cada combinación de los dos factores:

|            | Temp. Baja | Temp. Media | Temp. Alta |
|------------|------------|-------------|------------|
| Conc. Baja | 51         | 46          | 42         |
| Conc. Alta | 59         | 54          | 48         |

Además, sabemos que  $\text{SCT} = 600$ . Nos planteamos las siguientes cuestiones: ¿Influye la concentración de  $\text{CO}_2$  sobre el crecimiento? ¿Influye la temperatura sobre el crecimiento? ¿Se produce alguna interacción apreciable entre la concentración de  $\text{CO}_2$  y la temperatura? Proponer un modelo adecuado y hacer el estudio para dar una respuesta razonada a estas tres preguntas a un nivel de significación 0,05.

11.– Se hace un estudio para ver de qué manera influyen el tipo de población (HABITAT) y el tipo de vivienda (TIPOVIV) sobre la cantidad de papel y cartón reciclados. Para esto, se toman datos del «número de kg reciclados por vivienda en un mes» en 9 viviendas pequeñas (3 en ciudades pequeñas, 3 en ciudades medianas y 3 en ciudades grandes), y en 9 viviendas grandes (3 en ciudades pequeñas, 3 en ciudades medianas y 3 en ciudades grandes). Se analizan los resultados con el SPSS, obteniéndose los siguientes resultados:

| Fuente          | Suma de cuadrados | g.l. | Media cuadrática | F      | Significación |
|-----------------|-------------------|------|------------------|--------|---------------|
| HABITAT         | 0,333             | 2    | 0,167            | 0,008  | 0,992         |
| TIPOVIV         | 410,889           | 1    | 410,889          | 19,210 | 0,001         |
| HABITAT*TIPOVIV | 4,111             | 2    | 2,056            | 0,096  | 0,909         |
| Error           | 256,667           | 12   | 21,389           |        |               |
| Total           | 672,000           | 17   |                  |        |               |

a. ¿Influye el tipo de población (HABITAT) sobre la cantidad reciclada? ¿Influye el tipo de vivienda (TIPOVIV) sobre la cantidad reciclada? ¿Existe interacción significativa entre los dos factores? Dar respuestas razonadas al nivel de significación 0,05 e indicar el modelo estadístico utilizado.

b. Con los mismos datos, consideramos ahora un modelo de diseño de experimentos con un solo factor (el tipo de vivienda). Construir la tabla ANOVA para este diseño y tomar una decisión razonada (al nivel 0,05) sobre si el tipo de vivienda influye o no sobre la cantidad reciclada.

12.– Se está estudiando la influencia del nivel de riego y del tipo de fertilizante sobre el crecimiento de cierto tipo de arbustos al finalizar su primer año de vida. Se anota la altura alcanzada por 4 arbustos sometidos a un nivel bajo de riego, de otros 4 arbustos sometidos a un nivel moderado de riego, y de otros 4 con un alto nivel de riego. La mitad de los arbustos de cada grupo han sido fertilizados con una mezcla de guano de pollo y cascarilla de arroz, mientras que la otra mitad han sido fertilizados con una mezcla de guano de vacuno y serrín de pino. Se muestra a continuación una parte de los resultados obtenidos al analizar los datos con la ayuda de SPSS:

#### Pruebas de los efectos inter-sujetos

Variable dependiente: Altura alcanzada

| Origen           | Suma de cuadrados tipo III | gl | Media cuadrática | F        | Sig. |
|------------------|----------------------------|----|------------------|----------|------|
| Modelo corregido | 300,500 <sup>a</sup>       | 3  | 100,167          | 111,814  | ,000 |
| Intersección     | 3008,333                   | 1  | 3008,333         | 3358,140 | ,000 |
| Riego            | 300,167                    | 2  | 150,083          | 167,535  | ,000 |
| Fertilizante     | ,333                       | 1  | ,333             | ,372     | ,559 |
| Error            | 7,167                      | 8  | ,896             |          |      |
| Total            | 3316,000                   | 12 |                  |          |      |
| Total corregida  | 307,667                    | 11 |                  |          |      |

a. R cuadrado = ,977 (R cuadrado corregida = ,968)

#### Comparaciones múltiples

Altura alcanzada  
Bonferroni

| (I) Nivel de riego      | (J) Nivel de riego      | Diferencia de medias (I-J) | Error tip. | Sig. | Intervalo de confianza 95% |                 |
|-------------------------|-------------------------|----------------------------|------------|------|----------------------------|-----------------|
|                         |                         |                            |            |      | Límite inferior            | Límite superior |
| Nivel bajo de riego     | Nivel moderado de riego | -6,00*                     | ,669       | ,000 | -8,02                      | -3,98           |
|                         | Nivel alto de riego     | -12,25*                    | ,669       | ,000 | -14,27                     | -10,23          |
| Nivel moderado de riego | Nivel bajo de riego     | 6,00*                      | ,669       | ,000 | 3,98                       | 8,02            |
|                         | Nivel alto de riego     | -6,25*                     | ,669       | ,000 | -8,27                      | -4,23           |
| Nivel alto de riego     | Nivel bajo de riego     | 12,25*                     | ,669       | ,000 | 10,23                      | 14,27           |
|                         | Nivel moderado de riego | 6,25*                      | ,669       | ,000 | 4,23                       | 8,27            |

Basadas en las medias observadas.  
El término de error es la media cuadrática(Error) = ,896.

\*. La diferencia de medias es significativa al nivel ,05.

a. Plantear con detalle todos los elementos e hipótesis del modelo de diseño de experimentos que se ha empleado.

b. La mezcla que se ha utilizado como fertilizante, ¿influye sobre el crecimiento de los arbustos al nivel de significación del 5%? En caso afirmativo, ¿entre qué mezclas de fertilizante encontramos diferencias significativas? Contestar razonadamente a partir de los resultados mostrados, al nivel de significación conjunto del 5%.

c. El nivel de riego, ¿influye sobre el crecimiento de los arbustos al nivel de significación del 5%? En caso afirmativo, ¿entre qué niveles de riego encontramos diferencias significativas? Contestar razonadamente a partir de los resultados mostrados, al nivel de significación conjunto del 5%.

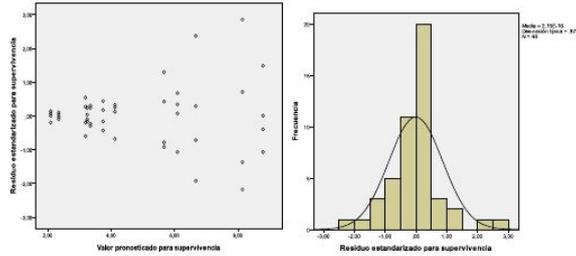
d. A partir de lo obtenido en los apartados anteriores, simplificar el modelo si se considera adecuado (explicando la razón) y obtener la nueva tabla ANOVA. Sacar las conclusiones pertinentes a partir de esta nueva tabla (al 5%).

13.- En un experimento sobre el efecto de 3 venenos y 4 posibles tratamientos sobre la supervivencia (en horas) de un tipo de animales, se asignaron al azar 4 animales a cada combinación de veneno y tratamiento. Se obtuvieron los siguientes resultados:

**Pruebas de los efectos inter-sujetos**  
Variable dependiente: supervivencia

| Origen               | Suma de cuadrados tipo III | gl | Media cuadrática | F       | Sig. |
|----------------------|----------------------------|----|------------------|---------|------|
| Modelo corregido     | 220,436 <sup>a</sup>       | 11 | 20,040           | 9,010   | ,000 |
| Intersección         | 1103,042                   | 1  | 1103,042         | 495,919 | ,000 |
| veneno               | 103,301                    | 2  | 51,651           | 23,222  | ,000 |
| tratamiento          | 92,121                     | 3  | 30,707           | 13,806  | ,000 |
| veneno * tratamiento | 25,014                     | 6  | 4,169            | 1,874   | ,112 |
| Error                | 80,072                     | 36 | 2,224            |         |      |
| Total                | 1403,550                   | 48 |                  |         |      |
| Total corregida      | 300,508                    | 47 |                  |         |      |

a. R cuadrado = ,734 (R cuadrado corregida = ,652)



**Contraste de Levene sobre la igualdad de las varianzas**

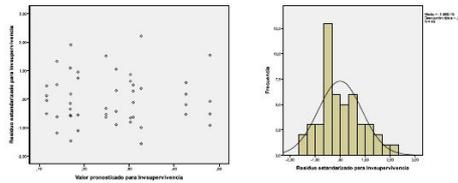
Variable dependiente: supervivencia

| F     | gl1 | gl2 | Sig. |
|-------|-----|-----|------|
| 4,854 | 11  | 36  | ,000 |

a. Diseño: Intersección + veneno + tratamiento + veneno \* tratamiento

a. Indicar el modelo utilizado, detallando todos sus elementos. Con la información aportada realizar un diagnóstico de los requisitos previos.

Se decide utilizar la transformación «inversa de la supervivencia» en vez de la variable «supervivencia». Con la nueva variable los resultados son:



**Contraste de Levene sobre la igualdad de las varianzas**

Variable dependiente: Inversa de la supervivencia

| F     | gl1 | gl2 | Sig. |
|-------|-----|-----|------|
| 1,576 | 11  | 36  | ,148 |

a. Diseño: Intersección + veneno + tratamiento + veneno \* tratamiento

**Pruebas de los efectos inter-sujetos**  
Variable dependiente: Inversa de la supervivencia

| Origen               | Suma de cuadrados tipo III | gl | Media cuadrática | F        | Sig. |
|----------------------|----------------------------|----|------------------|----------|------|
| Modelo corregido     | ,569 <sup>a</sup>          | 11 | ,052             | 21,531   | ,000 |
| Intersección         | 3,301                      | 1  | 3,301            | 1374,881 | ,000 |
| veneno               | ,349                       | 2  | ,174             | 72,635   | ,000 |
| tratamiento          | ,204                       | 3  | ,068             | 28,343   | ,000 |
| veneno * tratamiento | ,016                       | 6  | ,003             | 1,090    | ,387 |
| Error                | ,086                       | 36 | ,002             |          |      |
| Total                | 3,956                      | 48 |                  |          |      |
| Total corregida      | ,655                       | 47 |                  |          |      |

**Comparaciones múltiples (Bonferroni) Inversa de la supervivencia**

| (I)tratamiento | (J)tratamiento | Diferencia de medias (I-J) | Error típ. | Sig. | Intervalo de confianza 95% |                 |
|----------------|----------------|----------------------------|------------|------|----------------------------|-----------------|
|                |                |                            |            |      | Límite inferior            | Límite superior |
| 1              | 2              | ,1657 <sup>*</sup>         | ,02000     | ,000 | ,1099                      | ,2216           |
|                | 3              | ,0572 <sup>*</sup>         | ,02000     | ,042 | ,0014                      | ,1131           |
|                | 4              | ,1358 <sup>*</sup>         | ,02000     | ,000 | ,0800                      | ,1917           |
| 2              | 1              | -,1657 <sup>*</sup>        | ,02000     | ,000 | -,2216                     | -,1099          |
|                | 3              | -,1085 <sup>*</sup>        | ,02000     | ,000 | -,1644                     | -,0527          |
|                | 4              | -,0299                     | ,02000     | ,862 | -,0858                     | ,0259           |
| 3              | 1              | -,0572 <sup>*</sup>        | ,02000     | ,042 | -,1131                     | -,0014          |
|                | 2              | ,1085 <sup>*</sup>         | ,02000     | ,000 | ,0527                      | ,1644           |
|                | 4              | ,0786 <sup>*</sup>         | ,02000     | ,002 | ,0228                      | ,1345           |
| 4              | 1              | -,1358 <sup>*</sup>        | ,02000     | ,000 | -,1917                     | -,0800          |
|                | 2              | ,0299                      | ,02000     | ,862 | -,0259                     | ,0858           |
|                | 3              | -,0786 <sup>*</sup>        | ,02000     | ,002 | -,1345                     | -,0228          |

\*. La diferencia de medias es significativa al nivel ,05.

b. Con esta transformación ¿mejora o empeora el diagnóstico de los requisitos previos?

c. Indicar las conclusiones razonadas (a un nivel del 5 %) que se obtienen de la tabla ANOVA, especificando la hipótesis nula y alternativa en cada caso.

d. Indicar los contrastes realizados y las conclusiones que se obtienen en la tabla de resultados de Bonferroni.

14.- Se piensa que puede haber dos factores que influyan sobre la mayor o menor contaminación por arsénico del suelo y del subsuelo. Un posible factor es la época del año (verano o invierno) y otro posible factor es el nivel de profundidad (en superficie, a media profundidad o a gran profundidad). Se toman 11 réplicas para cada combinación, obteniendo en total 66 datos sobre el contenido de arsénico (As). Se analizan estos datos con el SPSS mediante un modelo de diseño de experimentos y a continuación se ofrecen los resultados más interesantes:

#### Pruebas de los efectos inter-sujetos

Variable dependiente: As

| Fuente           | Suma de cuadrados tipo III | gl | Media cuadrática | F       | Significación |
|------------------|----------------------------|----|------------------|---------|---------------|
| Modelo corregido | 176,272 <sup>a</sup>       | 5  | 35,254           | 1,576   | ,181          |
| Intersección     | 3897,140                   | 1  | 3897,140         | 174,223 | ,000          |
| Epoca            | 12,498                     | 1  | 12,498           | ,559    | ,458          |
| Nivel            | 159,532                    | 2  | 79,766           | 3,566   | ,034          |
| Epoca * Nivel    | 4,242                      | 2  | 2,121            | ,095    | ,910          |
| Error            | 1342,125                   | 60 | 22,369           |         |               |
| Total            | 5415,537                   | 66 |                  |         |               |
| Total corregida  | 1518,396                   | 65 |                  |         |               |

a. R cuadrado = ,116 (R cuadrado corregida = ,042)

#### 1. Media global

Variable dependiente: As

| Media | Error típ. | Intervalo de confianza al 95%. |                 |
|-------|------------|--------------------------------|-----------------|
|       |            | Límite inferior                | Límite superior |
| 7,684 | ,582       | 6,520                          | 8,849           |

#### 2. Epoca del año

Variable dependiente: As

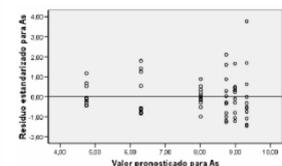
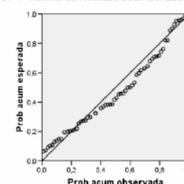
| Epoca del año | Media | Error típ. | Intervalo de confianza al 95%. |                 |
|---------------|-------|------------|--------------------------------|-----------------|
|               |       |            | Límite inferior                | Límite superior |
| Verano        | 7,249 | ,823       | 5,602                          | 8,896           |
| Invierno      | 8,119 | ,823       | 6,473                          | 9,766           |

#### 3. Nivel

Variable dependiente: As

| Nivel             | Media | Error típ. | Intervalo de confianza al 95%. |                 |
|-------------------|-------|------------|--------------------------------|-----------------|
|                   |       |            | Límite inferior                | Límite superior |
| Superficie        | 8,372 | 1,008      | 6,355                          | 10,389          |
| Profundidad media | 9,149 | 1,008      | 7,132                          | 11,166          |
| Gran profundidad  | 5,532 | 1,008      | 3,515                          | 7,549           |

Gráfico P-P Normal de Residuo estandarizado para As



#### 4. Epoca del año \* Nivel

Variable dependiente: As

| Epoca del año | Nivel             | Media | Error típ. | Intervalo de confianza al 95%. |                 |
|---------------|-------------------|-------|------------|--------------------------------|-----------------|
|               |                   |       |            | Límite inferior                | Límite superior |
| Verano        | Superficie        | 8,009 | 1,426      | 5,157                          | 10,862          |
|               | Profundidad media | 8,982 | 1,426      | 6,129                          | 11,834          |
|               | Gran profundidad  | 4,756 | 1,426      | 1,904                          | 7,609           |
| Invierno      | Superficie        | 8,735 | 1,426      | 5,882                          | 11,587          |
|               | Profundidad media | 9,316 | 1,426      | 6,464                          | 12,169          |
|               | Gran profundidad  | 6,307 | 1,426      | 3,455                          | 9,160           |

- Plantear con detalle el modelo utilizado.
- ¿Cuál es el contenido medio de arsénico en invierno? ¿Cuál es el contenido medio de arsénico en invierno a gran profundidad?
- ¿Influye la época del año sobre el contenido de arsénico? ¿Influye el nivel de profundidad? ¿Interaccionan los dos factores entre sí? Dar respuestas al nivel de significación 0,05. ¿Debería simplificarse el modelo?
- ¿Son aceptables las hipótesis del modelo?
- A partir de los mismos 66 datos, se quiere llevar a cabo un análisis de diseño de experimentos con sólo un factor: el nivel de profundidad. Construir la tabla ANOVA para este caso. ¿Influye el nivel de profundidad sobre el contenido de arsénico? Responder al nivel de significación 0,05.
- Con este modelo, hallar el intervalo de confianza (al nivel 0,95) para estimar la diferencia de contenido medio de arsénico que hay en superficie y a gran profundidad. ¿Qué conclusión puedes obtener de este intervalo?

1.- Se ha realizado un estudio por medio de técnicas de seguimiento de marcadores radiactivos para analizar la capacidad del organismo para absorber hierro y plomo. Participan en el estudio diez personas. A cada una de ellas se le da una dosis oral idéntica de hierro (en forma sulfato ferroso) y de plomo (cloruro de plomo-203). Después de doce días se mide la cantidad de cada componente retenida en el organismo y, a partir de esta, se determina el porcentaje absorbido por el cuerpo. Se obtuvieron los siguientes datos que representan los porcentajes de hierro (X) y plomo (Y) absorbidos por cada individuo de la muestra considerada:

|   |    |    |    |    |    |    |    |    |    |     |
|---|----|----|----|----|----|----|----|----|----|-----|
| X | 17 | 22 | 35 | 43 | 80 | 85 | 91 | 92 | 96 | 100 |
| Y | 8  | 17 | 18 | 25 | 58 | 59 | 41 | 30 | 43 | 58  |

- Plantear modelo e hipótesis. Mediante el análisis de los residuos, ¿qué se puede decir sobre dichas hipótesis?
- Obtener la recta de regresión y el coeficiente de correlación lineal r. ¿Qué indica el valor del coeficiente de correlación obtenido?
- ¿Influye el porcentaje de hierro absorbido sobre el porcentaje de plomo absorbido? Obtener una conclusión, al nivel de significación 0,01.
- Estimar, con una confianza del 95 %, el valor medio del porcentaje de plomo absorbido para aquellas personas que absorben un 30 % de hierro.

2.- El muestreo de áreas contiguas se utiliza en Ecología para contar el número de especies distintas de plantas por área. El recuento se realiza de manera que cada siguiente área contigua tiene el doble de superficie, empezando por un área de 1 metro cuadrado. El modelo que relaciona  $Y = \text{«Número de especies distintas»}$  con  $X = \text{«Superficie (en metros cuadrados)»}$  es  $Y = a \ln X + b$  ( $a = \text{«Índice de diversidad»}$ ,  $b = \text{«Número de especies por unidad de área»}$ ). Ajustar dicho modelo a los datos:

|                    |   |   |   |    |    |    |    |
|--------------------|---|---|---|----|----|----|----|
| Superficie         | 1 | 2 | 4 | 8  | 16 | 32 | 64 |
| Especies distintas | 2 | 4 | 7 | 11 | 16 | 19 | 21 |

3.- En un estudio sobre la resistencia a bajas temperaturas del bacilo de la fiebre tifoidea, se expusieron cultivos del bacilo durante diferentes periodos de tiempo a  $-5$  grados centígrados. Los siguientes datos representan:  
 $X = \text{«Tiempo de exposición»}$  (en semanas)       $Y = \text{«Porcentaje de bacilos supervivientes»}$

|   |     |     |    |     |     |      |      |       |
|---|-----|-----|----|-----|-----|------|------|-------|
| X | 0   | 0,5 | 1  | 2   | 3   | 5    | 9    | 15    |
| Y | 100 | 42  | 14 | 7,5 | 0,4 | 0,11 | 0,05 | 0,002 |

$$\begin{matrix} \sum x_i & \sum y_i & \sum \log y_i & \sum x_i^2 & \sum y_i^2 & \sum (\log y_i)^2 & \sum x_i y_i & \sum x_i \log y_i \\ 35,5 & 164,062 & 0,664 & 345,25 & 12\,016,42 & 99,52 & 52,23 & -125,394 \end{matrix}$$

Ajustar una recta y una exponencial a los datos. Interpretar los resultados.

4.- Se estudia la influencia sobre el nivel de contaminación por nitratos (Y) del porcentaje de población conectada a sistemas de tratamiento de residuos (X) en 20 áreas de la UE. Los datos obtenidos son los siguientes:

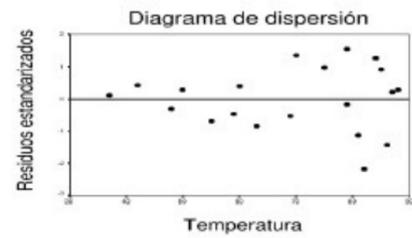
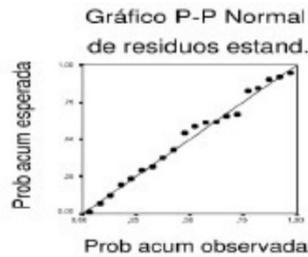
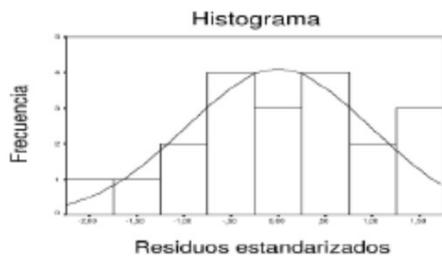
$$\begin{matrix} \sum x_i & \sum x_i^2 & \sum \log x_i & \sum (\log x_i)^2 & \sum y_i & \sum y_i^2 & \sum x_i y_i & \sum y_i \log x_i \\ 692 & 30\,430 & 67,34 & 235,01 & 82,7 & 432,49 & 2\,332,9 & 262,37 \end{matrix}$$

Ajustar un modelo de regresión logarítmico  $Y = \beta_0 + \beta_1 \ln X$ . ¿Es bueno este ajuste?

5.- Se lleva a cabo un estudio para tratar de explicar la supervivencia de cierta especie animal en función de las temperaturas máximas alcanzadas en los hábitats naturales en los que se desarrolla. Se seleccionan aleatoriamente 20 reservas naturales de esta especie y se mide el porcentaje de supervivientes al final del año, Y, y la temperatura máxima registrada en grados Fahrenheit, X. Los resultados que se obtienen son:

$$\begin{matrix} \sum x_i & \sum x_i^2 & \sum y_i & \sum y_i^2 & \sum x_i y_i \\ 1\,379 & 100\,055 & 823 & 42\,063 & 62\,103 \end{matrix}$$

- Calcular la recta de regresión y la varianza residual  $S_R^2$ .
- Realizar el contraste de la regresión. A nivel  $\alpha = 0,05$ , ¿podemos rechazar la hipótesis nula de que la temperatura no afecta a la supervivencia?
- A continuación se presentan algunos gráficos de residuos estandarizados. Analizar gráficamente si se cumplen las hipótesis de normalidad, homocedasticidad y linealidad.



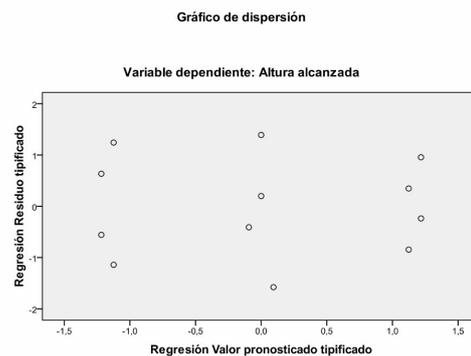
6.- Se está estudiando el crecimiento de cierto tipo de arbustos durante sus primeros años de vida. Se anota la altura (en centímetros) alcanzada por 12 arbustos, así como la edad (en meses) de cada uno de ellos. Se muestra a continuación una parte de los resultados obtenidos al analizar los datos con la ayuda de SPSS.

**Resumen del modelo<sup>b</sup>**

| Modelo | R                 | R cuadrado | R cuadrado corregida | Error típ. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,989 <sup>a</sup> | ,977       | ,975                 | ,839                        |

a. Variables predictoras: (Constante), Edad de la planta (en meses)

b. Variable dependiente: Altura alcanzada



**Coefficientes<sup>a</sup>**

| Modelo |                              | Coefficients no estandarizados |            | Coefficients tipificados | t      | Sig. |
|--------|------------------------------|--------------------------------|------------|--------------------------|--------|------|
|        |                              | B                              | Error típ. | Beta                     |        |      |
| 1      | (Constante)                  | 4,082                          | ,618       |                          | 6,608  | ,000 |
|        | Edad de la planta (en meses) | ,490                           | ,024       | ,989                     | 20,677 | ,000 |

a. Variable dependiente: Altura alcanzada

**ANOVA<sup>b</sup>**

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F       | Sig.              |
|--------|-----------|-------------------|----|------------------|---------|-------------------|
| 1      | Regresión | 300,635           | 1  | 300,635          | 427,530 | ,000 <sup>a</sup> |
|        | Residual  | 7,032             | 10 | ,703             |         |                   |
|        | Total     | 307,667           | 11 |                  |         |                   |

a. Variables predictoras: (Constante), Edad de la planta (en meses)

b. Variable dependiente: Altura alcanzada

**Estadísticos descriptivos**

|                              | N  | Mínimo | Máximo | Media | Desv. típ. |
|------------------------------|----|--------|--------|-------|------------|
| Altura alcanzada             | 12 | 9      | 23     | 15,83 | 5,289      |
| Edad de la planta (en meses) | 12 | 11     | 37     | 24,00 | 10,677     |
| N válido (según lista)       | 12 |        |        |       |            |

a. Plantear con detalle todos los elementos e hipótesis del modelo de regresión que se ha empleado. ¿Qué podemos decir sobre las hipótesis del modelo a partir de los resultados mostrados?

b. Estimar puntualmente todos los parámetros del modelo y escribir la recta de regresión que expresa la altura alcanzada en función de los meses de desarrollo. ¿Qué porcentaje de variabilidad explica esta recta?

c. La edad del arbusto, ¿ha resultado ser una variable significativa para explicar la altura? Justificar la respuesta con un nivel de significación del 1 %.

d. Estimar la altura media alcanzada por los arbustos al cabo de año y medio de vida. Dar un intervalo de confianza (al 99 %) para esta altura media.

7.- Un estudio sobre el efecto de la temperatura en el rendimiento de un proceso químico proporciona los siguientes resultados:

|                 |    |    |    |    |    |   |   |    |    |    |    |
|-----------------|----|----|----|----|----|---|---|----|----|----|----|
| Temperatura (X) | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2  | 3  | 4  | 5  |
| Rendimiento (Y) | 1  | 5  | 4  | 7  | 10 | 8 | 9 | 13 | 14 | 13 | 18 |

a. Asumiendo el modelo  $Y_i = \beta_0 + \beta_1 x_i + u_i$ , obtener las estimaciones de  $\beta_0$  y  $\beta_1$ . ¿Cuál es la recta de regresión estimada? ¿Es bueno el ajuste?

b. A partir de la tabla ANOVA decidir, con un nivel de significación  $\alpha = 0,05$ , si la temperatura influye de manera significativa sobre el rendimiento.

c. Construir un intervalo de confianza al 95 % para  $\beta_1$ .

d. Construir un intervalo de confianza al 95 % para estimar el rendimiento medio de todos los procesos que se desarrollan a una temperatura de  $x = 3$ .

e. Construir un intervalo de confianza al 95 % para estimar el rendimiento de un nuevo proceso que se desarrolla a una temperatura de  $x = 3$ .

8.- Se lleva a cabo un análisis estadístico con 15 plantas para expresar la variable  $Y$ =«Superficie de las hojas de la planta» (en  $\text{cm}^2$ ) en función de la variable  $X$ =«Iluminación» (en lux). Se obtienen los siguientes resultados con SPSS.

#### Resumen del modelo<sup>b</sup>

| Modelo | R                 | R cuadrado | R cuadrado corregida | Error típ. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,872 <sup>a</sup> | ,761       | ,742                 | 20,212                      |

a. Variables predictoras: (Constante), Iluminación (lux)

b. Variable dependiente: Superficie de las hojas de soja

#### ANOVA<sup>b</sup>

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F      | Sig.              |
|--------|-----------|-------------------|----|------------------|--------|-------------------|
| 1      | Regresión | 16892,100         | 1  | 16892,100        | 41,349 | ,000 <sup>a</sup> |
|        | Residual  | 5310,833          | 13 | 408,526          |        |                   |
|        | Total     | 22202,933         | 14 |                  |        |                   |

a. Variables predictoras: (Constante), Iluminación (lux)

b. Variable dependiente: Superficie de las hojas de soja

#### Coefficientes<sup>a</sup>

| Modelo |                   | Coefficients no estandarizados |            | Coefficients tipificados | t     | Sig. |
|--------|-------------------|--------------------------------|------------|--------------------------|-------|------|
|        |                   | B                              | Error típ. | Beta                     |       |      |
| 1      | (Constante)       | 149,767                        | 19,872     |                          | 7,536 | ,000 |
|        | Iluminación (lux) | ,206                           | ,032       | ,872                     | 6,430 | ,000 |

a. Variable dependiente: Superficie de las hojas de soja

#### Estadísticos descriptivos

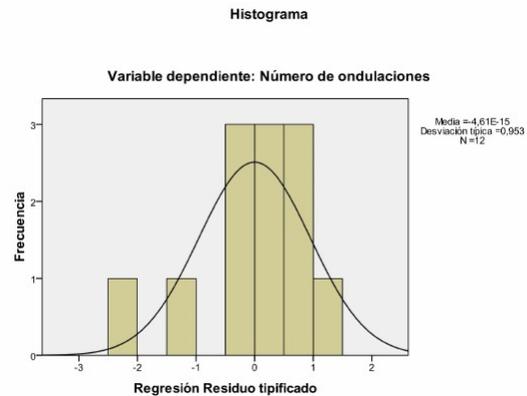
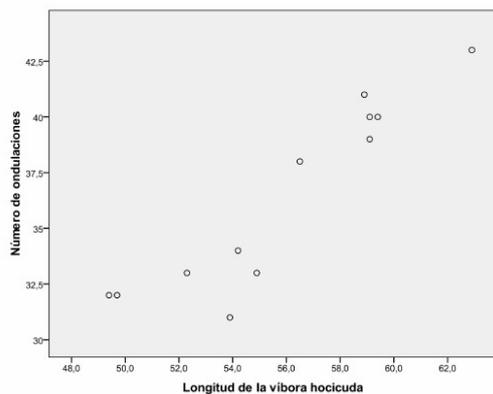
|                                 | Media    | Desviación típica | N  |
|---------------------------------|----------|-------------------|----|
| Superficie de las hojas de soja | 273,07   | 39,824            | 15 |
| Iluminación (lux)               | 600,0000 | 169,03085         | 15 |

a. Plantea el modelo utilizado describiendo todos sus elementos y sus requisitos. Escribe la recta de regresión estimada y evalúa su ajuste a los datos.

b. ¿Tiene la iluminación una influencia significativa (al 1 %) sobre la media de la superficie de las hojas? Justifica la respuesta.

c. Estima (al 99 % de confianza) la superficie media de las hojas de todas las plantas que crecen bajo una iluminación de 700 lux.

9.- La víbora hocicuda (*Vipera latastei*) es una especie de víbora presente en la Península Ibérica y en el norte del Magreb. A partir de los datos de 12 víboras, se desea hacer un estudio estadístico sobre el número de ondulaciones de la banda dorsal o zigzag en función de la longitud de estas víboras. A continuación, se ofrecen algunos gráficos y tablas obtenidas con SPSS.



ANOVA<sup>b</sup>

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F      | Sig.              |
|--------|-----------|-------------------|----|------------------|--------|-------------------|
| 1      | Regresión | 167,780           | 1  | 167,780          | 58,083 | ,000 <sup>a</sup> |
|        | Residual  | 28,886            | 10 | 2,889            |        |                   |
|        | Total     | 196,667           | 11 |                  |        |                   |

a. Variables predictoras: (Constante), Longitud de la víbora hocicuda

Coefficientes<sup>a</sup>

| Modelo |                                | Coeficientes no estandarizados |            | Coeficientes tipificados | t      | Sig. |
|--------|--------------------------------|--------------------------------|------------|--------------------------|--------|------|
|        |                                | B                              | Error tip. | Beta                     |        |      |
| 1      | (Constante)                    | -15,801                        | 6,858      |                          | -2,304 | ,044 |
|        | Longitud de la víbora hocicuda | ,933                           | ,122       | ,924                     | 7,621  | ,000 |

a. Variable dependiente: Número de ondulaciones

Estadísticos descriptivos

|                                | Media  | Desviación típica | N  |
|--------------------------------|--------|-------------------|----|
| Número de ondulaciones         | 36,33  | 4,228             | 12 |
| Longitud de la víbora hocicuda | 55,858 | 4,1845            | 12 |

- Planteamiento de todos los elementos, requisitos previos, etc., del modelo estadístico utilizado en el estudio. Con los gráficos aportados, ¿qué se puede decir sobre los requisitos previos del modelo?
- Estimar, mediante un intervalo de confianza al 95 %, el coeficiente de la variable explicativa del modelo de regresión utilizado.
- ¿Tiene la longitud de la víbora una influencia significativa sobre el número de ondulaciones? Escribir la hipótesis nula, la hipótesis alternativa, y obtener una conclusión razonada al 5 % de significación.
- Estimar, mediante un intervalo al 95 %, el número medio de ondulaciones de las víboras que tienen una longitud de 58 cm.

10.- Con el objetivo de reforestar la superficie calcinada tras un incendio, el gobierno de la Comunidad encarga un estudio a un equipo de investigación para analizar el efecto que tiene un fertilizante sobre una especie de pino. Para esto, anota el crecimiento medio anual obtenido con diferentes dosis de fertilizante (1, 2, 3 y 4 dosis) y desea ajustar un modelo de regresión logarítmico,  $Y = \beta_0 + \beta_1 \ln X$ , que explique el crecimiento medio anual en función de las dosis de fertilizante. Se obtienen los siguientes resultados con el SPSS.

Resumen del modelo

| Modelo | R                 | R cuadrado | R cuadrado corregida | Error tip. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,960 <sup>a</sup> | ,922       | ,883                 | 2,044                       |

a. Variables predictoras: (Constante), ln(dosis)

**Coefficientes<sup>a</sup>**

| Modelo |             | Coeficientes no estandarizados |            | Coeficientes estandarizados | t      | Sig. |
|--------|-------------|--------------------------------|------------|-----------------------------|--------|------|
|        |             | B                              | Error típ. | Beta                        |        |      |
| 1      | (Constante) | 21,181                         | 1,864      |                             | 11,360 | ,008 |
|        | ln(dosis)   | 9,526                          | 1,963      | ,960                        | 4,853  | ,040 |

a. Variable dependiente: Crecimiento medio anual

**ANOVA<sup>b</sup>**

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F      | Sig.              |
|--------|-----------|-------------------|----|------------------|--------|-------------------|
| 1      | Regresión | 98,396            | 1  | 98,396           | 23,556 | ,040 <sup>a</sup> |
|        | Residual  | 8,354             | 2  | 4,177            |        |                   |
|        | Total     | 106,750           | 3  |                  |        |                   |

a. Variables predictoras: (Constante), ln(dosis)

- ¿Cuál es el modelo de regresión logarítmico ajustado? ¿Es bueno el ajuste?
- ¿Cuál sería el crecimiento medio anual esperado utilizando 3 dosis?
- Obtener un intervalo de confianza, al 90 %, para estimar el parámetro  $\beta_1$ .
- ¿Se puede afirmar que el número de dosis de fertilizante influye sobre el crecimiento medio anual? Dar una respuesta con nivel de significación 0,10.

**MODELO DE REGRESIÓN MÚLTIPLE**

1.- En una asignatura se realizan dos exámenes parciales (calificado cada uno de ellos de 0 a 10) y una evaluación final (calificada también de 0 a 10). Se diseña un modelo para predecir la calificación de la evaluación final mediante las calificaciones de los dos parciales. Realizada una muestra de 10 alumnos elegidos al azar, se obtuvieron los siguientes datos:

|                           |     |     |     |     |     |     |     |     |     |     |
|---------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Calificación parcial 1    | 6,5 | 8,3 | 9,5 | 7,0 | 6,2 | 7,0 | 9,0 | 6,5 | 8,5 | 4,0 |
| Calificación parcial 2    | 2,0 | 5,5 | 5,5 | 7,0 | 2,7 | 8,5 | 6,0 | 5,0 | 4,5 | 2,7 |
| Calificación examen final | 4,9 | 5,7 | 7,5 | 8,5 | 3,0 | 8,0 | 8,9 | 4,7 | 7,9 | 5,4 |

Plantea el modelo e indica sus requisitos previos. Realiza el análisis adecuado y comenta las conclusiones.

2.- En un estudio <sup>1</sup> sobre el nivel de nitrógeno en los huesos de 36 esqueletos de hombres en dos cementerios, se consideraron las siguientes variables:

- Tiempo enterrado en años
- Profundidad en pies
- Nivel de nitrógeno (en gramos por 100 gramos de hueso)

Para detectar el posible efecto del tiempo de enterramiento y la profundidad sobre el nivel de nitrógeno, se llevó a cabo un análisis estadístico obteniendo los siguientes resultados:

**ANOVA<sup>a</sup>**

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F      | Sig.              |
|--------|-----------|-------------------|----|------------------|--------|-------------------|
| 1      | Regresión | ,622              | 2  | ,311             | 16,844 | ,000 <sup>b</sup> |
|        | Residuo   | ,609              | 33 | ,018             |        |                   |
|        | Total     | 1,231             | 35 |                  |        |                   |

a. Variable dependiente: Nitrógeno

b. Predictores: (Constante), Profundidad, Años de entierro

<sup>1</sup> Fuente: D.R. Jarvis (1997). «Nitrogen Levels in Long Bones from Coffin Burials Interred for Periods of 26-90 Years», Forensic Science International, Vol 85, pp 199-208.

**Coefficientes<sup>a</sup>**

| Modelo           | Coefficients no estandarizados |                | Coefficients estandarizados | t      | Sig. |
|------------------|--------------------------------|----------------|-----------------------------|--------|------|
|                  | B                              | Error estándar | Beta                        |        |      |
| 1 (Constante)    | 3,962                          | ,185           |                             | 21,374 | ,000 |
| Años de entierro | -,006                          | ,001           | -,600                       | -3,959 | ,000 |
| Profundidad      | ,020                           | ,018           | ,167                        | 1,103  | ,278 |

a. Variable dependiente: Nitrógeno

- Indica el modelo utilizado y todos sus requisitos previos.
- Indica el valor estimado de todos los parámetros del modelo ¿Cuál sería la variación del nivel de nitrógeno si aumentamos 10 años el tiempo de entierro?
- Indica la hipótesis nula y la alternativa del contraste que se puede realizar con la tabla ANOVA y su conclusión justificada (al 5 % de significación).
- ¿Tenemos evidencia significativa, al nivel 0,05, de que la profundidad influye? ¿y los años de entierro? Justifica la respuesta.
- Estima el nivel medio de nitrógeno de los huesos enterrados durante 40 años a una profundidad de 5 pies.

3.- El biólogo responsable de costes de una planta productiva de abonos estudia el consumo de agua en función de la media de producción diaria en el mes y de la temperatura media mensual. Se utilizan las variables *Productividad*: «media mensual de producción diaria» (en Tm), *Temperatura*: «temperatura media mensual» (en °C) y *Consumo*: «media de consumo diario de agua en el mes» (en m<sup>3</sup>). Decide realizar un primer análisis utilizando las dos variables como explicativas. Los resultados obtenidos se muestran a continuación:

**Resumen del modelo<sup>b</sup>**

| Modelo | R                 | R cuadrado | R cuadrado corregida | Error típ. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,657 <sup>a</sup> | ,431       | ,350                 | 76,817                      |

a. Variables predictoras: (Constante), Producción media diaria, Temperatura media mensual

**ANOVA<sup>a</sup>**

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F     | Sig.              |
|--------|-----------|-------------------|----|------------------|-------|-------------------|
| 1      | Regresión | 62672,800         | 2  | 31336,400        | 5,311 | ,019 <sup>b</sup> |
|        | Residual  | 82611,193         | 14 | 5900,800         |       |                   |
|        | Total     | 145283,993        | 16 |                  |       |                   |

a. Variable dependiente: Consumo diario de agua

b. Variables predictoras: (Constante), Producción media diaria, Temperatura media mensual

**Coefficientes<sup>a</sup>**

| Modelo                    | Coefficients no estandarizados |            | Coefficients tipificados | t     | Sig. |
|---------------------------|--------------------------------|------------|--------------------------|-------|------|
|                           | B                              | Error típ. | Beta                     |       |      |
| 1 (Constante)             | 352,7                          | 92,226     |                          | 3,824 | ,002 |
| Temperatura media mensual | 2,052                          | 2,586      | ,162                     | ,794  | ,441 |
| Producción media diaria   | ,379                           | ,117       | ,661                     | 3,243 | ,006 |

- Indica el modelo utilizado y todos sus requisitos previos.
- Explica las hipótesis y el resultado del contraste de la tabla ANOVA (utiliza una significación del 5 % en tu respuesta).
- Obtén intervalos de confianza del 95 % para los coeficientes de las variables Producción y Temperatura en el modelo. A la vista de estos intervalos (o de los resultados que se observan en las tablas dadas) indica si hay evidencia significativa de que alguna de las variables influya.
- Escribe el modelo ajustado y estima la media de consumo de agua en un mes en el que la producción media diaria ha sido de 350 Tm y la media de temperatura ha sido de 21 grados centígrados. ¿Cuál es el incremento medio del consumo de agua si se incrementa en una tonelada la producción?

4.- Se desea hacer una regresión lineal que explique el contenido en sales minerales del húmero dominante en función de  $X_1$ =«Contenido en sales minerales del radio dominante» y de  $X_2$  =«Contenido en sales minerales del radio no dominante». Se toman datos en 5 personas y se analizan con el SPSS (los resultados se ofrecen a continuación):

### Resumen del modelo

| Modelo | R    | R cuadrado | R cuadrado corregida | Error tıp. de la estimación |
|--------|------|------------|----------------------|-----------------------------|
| 1      | ,989 | ,977       | ,955                 | 3,4990E - 02                |

### ANOVA

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F      | Sig. |
|--------|-----------|-------------------|----|------------------|--------|------|
| 1      | Regresión | ,106              | 2  | 5,289E - 02      | 43,196 | ,023 |
|        | Residual  | 2,449E - 03       | 2  | 1,224E - 03      |        |      |
|        | Total     | ,108              | 4  |                  |        |      |

### Coefficientes

| Modelo |                            | Coefficients no estandarizados |            | Coefficients estandarizados | t     | Sig. |
|--------|----------------------------|--------------------------------|------------|-----------------------------|-------|------|
|        |                            | B                              | Error tıp. | Beta                        |       |      |
| 1      | (Constante)                | ,619                           | ,136       |                             | 4,544 | ,045 |
|        | Radio (parte dominante)    | ,471                           | ,326       | ,345                        | 1,444 | ,286 |
|        | Radio (parte no dominante) | ,956                           | ,342       | ,668                        | 2,798 | ,107 |

¿Son explicativas las variables  $X_1$  y  $X_2$ ? ¿Es explicativo el modelo en su conjunto? Dar respuestas razonadas, al nivel de significación 0,05. A partir de las respuestas anteriores, obtener conclusiones razonadas sobre cuál sería el procedimiento a seguir para quedarnos con un modelo de regresión adecuado.

5.- En una muestra de 20 mujeres sanas entre 20 y 34 años se ha medido su cantidad de grasa corporal, el espesor del pliegue cutáneo del tríceps, el perímetro del muslo y el perímetro del antebrazo. La cantidad de grasa corporal se mide mediante un complicado y caro procedimiento que requiere la inmersión de cada persona en agua, por lo que sería muy útil si un modelo de regresión a partir de las variables disponibles permitiera predecir de forma fiable dicha cantidad. Se ha ajustado el siguiente modelo de regresión múltiple:

$$\text{Grasa} = \beta_0 + \beta_1 \text{Triceps} + \beta_2 \text{MUSLO} + \beta_3 \text{ANTEBRAZO} + U$$

Se sabe que el estadístico  $F$  en la tabla Anova correspondiente a este modelo es  $F = 21,516$ . A continuación aparecen la matriz de correlaciones entre todas las variables involucradas y otros resultados del ajuste obtenidos con SPSS:

### Correlaciones

|           |                        | Grasa | Triceps | Muslo | Antebrazo |
|-----------|------------------------|-------|---------|-------|-----------|
| Grasa     | Correlación de Pearson | 1     | ,843    | ,878  | ,142      |
|           | Sig. (bilateral)       |       | ,000    | ,000  | ,549      |
|           | N                      | 20    | 20      | 20    | 20        |
| Triceps   | Correlación de Pearson | ,843  | 1       | ,924  | ,458      |
|           | Sig. (bilateral)       | ,000  |         | ,000  | ,042      |
|           | N                      | 20    | 20      | 20    | 20        |
| Muslo     | Correlación de Pearson | ,878  | ,924    | 1     | ,085      |
|           | Sig. (bilateral)       | ,000  | ,000    |       | ,723      |
|           | N                      | 20    | 20      | 20    | 20        |
| Antebrazo | Correlación de Pearson | ,142  | ,458    | ,085  | 1         |
|           | Sig. (bilateral)       | ,549  | ,042    | ,723  |           |
|           | N                      | 20    | 20      | 20    | 20        |

### Coefficientes<sup>a</sup>

| Modelo |             | Coefficients no estandarizados |            | Coefficients tipificados |        |      |
|--------|-------------|--------------------------------|------------|--------------------------|--------|------|
|        |             | B                              | Error tıp. | Beta                     | t      | Sig. |
| 1      | (Constante) | 117,085                        | 99,782     |                          | 1,173  | ,258 |
|        | Triceps     | 4,334                          | 3,016      | 4,264                    | 1,437  | ,170 |
|        | Muslo       | -2,857                         | 2,582      | -2,929                   | -1,106 | ,285 |
|        | Antebrazo   | -2,186                         | 1,595      | -1,561                   | -1,370 | ,190 |

a. Variable dependiente: Grasa

- Calcula el coeficiente de determinación correspondiente al modelo anterior. ¿Qué significa el valor obtenido?
- Con los resultados disponibles, ¿se puede afirmar a nivel 0,05 que el espesor del pliegue cutáneo del tríceps, el perímetro del muslo y el perímetro del antebrazo son útiles conjuntamente para explicar la cantidad de grasa corporal?
- Calcula un intervalo de confianza de nivel 95 % para el coeficiente  $\beta_2$ .
- Determina cuáles de las tres variables regresoras son individualmente significativas a nivel 0,05 para explicar la variable respuesta.
- A la vista de toda la información disponible, ¿existe algún problema con los resultados obtenidos en el apartado anterior? En caso afirmativo, ¿cómo se podría resolver?

6.- A fin de conocer la absorción de una droga por el hígado de la rata, se seleccionan aleatoriamente 19 ejemplares, se pesan y se les suministra un anestésico ligero a fin de suministrarles posteriormente una dosis oral de la droga. Se suministra a cada rata una determinada fracción de una dosis básica (variable «dosis») y tras un periodo de tiempo fijado se sacrifica cada ejemplar, se pesa su hígado y se estima la fracción de droga absorbida (variable «dosis en hígado»). Los estadísticos descriptivos de estas variables se recogen en la tabla siguiente:

|                        | N  | Media  | Desv. tip. | Varianza |
|------------------------|----|--------|------------|----------|
| Peso de la rata        | 19 | 171,53 | 16,490     | 271,930  |
| Peso del hígado        | 19 | 7,811  | 1,2229     | 1,495    |
| Dosis                  | 19 | ,8621  | ,08580     | ,007     |
| Dosis en hígado        | 19 | ,3353  | ,08847     | ,008     |
| N valido (segun lista) | 19 |        |            |          |

### PRIMER ANÁLISIS

Con los datos recogidos se trata de ver, en primer lugar, que efectivamente se puede predecir el peso del hígado por medio del peso de la rata. Los resultados obtenidos en este primer análisis se recogen en los siguientes cuadros y tablas.

Resumen del modelo

| Modelo | R                 | R cuadrado | R cuadrado corregida |
|--------|-------------------|------------|----------------------|
| 1      | ,500 <sup>a</sup> | ,250       | ,206                 |

a. Variables predictoras: (Constante), Peso de la rata

Coefficientes<sup>a</sup>

| Modelo |                 | Coefficients no estandarizados |            | Coefficients tipificados |       |      |
|--------|-----------------|--------------------------------|------------|--------------------------|-------|------|
|        |                 | B                              | Error tip. | Beta                     | t     | Sig. |
| 1      | (Constante)     | 1,450                          | 2,683      |                          | ,541  | ,596 |
|        | Peso de la rata | ,037                           | ,016       | ,500                     | 2,381 | ,029 |

a. Variable dependiente: Peso del hígado

ANOVA<sup>a</sup>

| Modelo | Suma de cuadrados | gl     | Media cuadrática | F     | Sig.  |
|--------|-------------------|--------|------------------|-------|-------|
| 1      | Regresión         | 6,730  | 1                | 6,730 | 5,667 |
|        | Residual          | 20,188 | 17               | 1,188 |       |
|        | Total             | 26,918 | 18               |       |       |

a. Variables predictoras: (Constante), Peso de la rata

b. Variable dependiente: Peso del hígado

Se pide:

- ¿De qué modelo se trata? Describe todos sus elementos y variables. ¿Cuál es el coeficiente de determinación? ¿Cómo se interpreta?
- ¿Cuál es el valor de la varianza residual? ¿Cuáles son los otros parámetros del modelo y qué estimaciones se obtienen? ¿Qué contraste de hipótesis se decide con el p-valor de la tabla ANOVA? Escribe claramente sus hipótesis nula y alternativa. ¿Cuál es la conclusión al nivel de significación  $\alpha = 0,05$ ?
- Da un intervalo de confianza del 90 % para el peso medio del hígado de las ratas cuyo peso en vivo es de 150 g.

### SEGUNDO ANÁLISIS

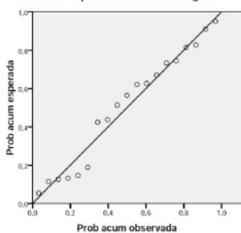
Posteriormente, se estudia si se puede determinar la fracción de droga absorbida («dosis en hígado») en función del peso de la rata y de la dosis suministrada. Los resultados obtenidos en este segundo análisis se recogen en los siguientes cuadros y figuras.

Resumen del modelo<sup>b</sup>

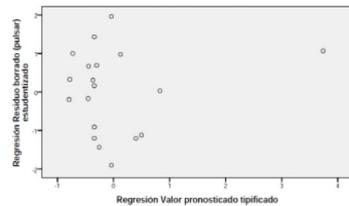
| Modelo | R                 | R cuadrado | R cuadrado corregida | Error tip. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,578 <sup>a</sup> | ,335       | ,251                 | ,07654                      |

a. Variables predictoras: (Constante), Dosis, Peso de la rata

Variable dependiente: Dosis en hígado



Variable dependiente: Dosis en hígado



ANOVA<sup>b</sup>

| Modelo | Suma de cuadrados | gl   | Media cuadrática | F    | Sig.  |
|--------|-------------------|------|------------------|------|-------|
| 1      | Regresión         | ,047 | 2                | ,024 | 4,024 |
|        | Residual          | ,094 | 16               | ,006 |       |
|        | Total             | ,141 | 18               |      |       |

a. Variables predictoras: (Constante), Dosis, Peso de la rata

Coefficientes<sup>a</sup>

| Modelo |                 | Coefficients no estandarizados |            | Coefficients tipificados |        |      |
|--------|-----------------|--------------------------------|------------|--------------------------|--------|------|
|        |                 | B                              | Error tip. | Beta                     | t      | Sig. |
| 1      | (Constante)     | ,286                           | ,191       |                          | 1,493  | ,155 |
|        | Peso de la rata | -,020                          | ,008       | -,3811                   | -2,608 | ,019 |
|        | Dosis           | 4,125                          | 1,506      | 4,001                    | 2,738  | ,015 |

a. Variable dependiente: Dosis en hígado

Se pide:

- Describir con detalle el modelo utilizado, indicando sus parámetros, sus elementos y los requisitos que deben cumplir. Comentar las figuras indicando cómo cada una de ellas justifica si se cumplen o no los requisitos previos del modelo.
- Determinar qué coeficientes del modelo son significativamente no nulos (al nivel de significación  $\alpha = 0,05$ ), indicando en cada caso el p-valor utilizado y el cuadro del cual procede. Estimar el valor medio de la dosis en hígado cuando el peso de la rata en vivo es de 160 g y la dosis suministrada es de 0,75.

7.- En 12 grandes ciudades se hace un estudio sobre la tasa de contaminación (con cierto contaminante atmosférico).

Se piensa que puede estar influida por tres variables:

$X_1$  = «Índice de pluviosidad»;

$X_2$  = «Densidad de industrias contaminantes en el término municipal»;

$X_3$  = «Millones de habitantes».

Se adjunta la salida del estudio de regresión lineal efectuado:

**Resumen del modelo<sup>b</sup>**

| Modelo | R                 | R cuadrado | R cuadrado corregida | Error típ. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1      | ,999 <sup>a</sup> | ,998       | ,997                 | 12,13473                    |

a. Variables predictoras: (Constante), Millones de habitantes, Índice de pluviosidad, Densidad de industrias contaminantes

b. Variable dependiente: Tasa registrada de contaminante

**ANOVA<sup>b</sup>**

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F        | Sig.              |
|--------|-----------|-------------------|----|------------------|----------|-------------------|
| 1      | Regresión | 512454,237        | 3  | 170818,079       | 1160,042 | ,000 <sup>a</sup> |
|        | Residual  | 1178,013          | 8  | 147,252          |          |                   |
|        | Total     | 513632,250        | 11 |                  |          |                   |

a. Variables predictoras: (Constante), Millones de habitantes, Índice de pluviosidad, Densidad de industrias contaminantes

b. Variable dependiente: Tasa registrada de contaminante

**Coefficientes<sup>a</sup>**

| Modelo |                                      | Coefficients no estandarizados |            | Coefficients estandarizados |  | t       | Sig. |
|--------|--------------------------------------|--------------------------------|------------|-----------------------------|--|---------|------|
|        |                                      | B                              | Error típ. | Beta                        |  |         |      |
| 1      | (Constante)                          | 1010,036                       | 14,648     |                             |  | 68,956  | ,000 |
|        | Índice de pluviosidad                | -2,023                         | ,035       | -,987                       |  | -58,205 | ,000 |
|        | Densidad de industrias contaminantes | 2,016                          | ,184       | ,196                        |  | 10,962  | ,000 |
|        | Millones de habitantes               | -3,271                         | 4,524      | -,013                       |  | -,723   | ,490 |
|        |                                      |                                |            |                             |  |         |      |

a. Variable dependiente: Tasa registrada de contaminante

a. Especificar claramente el modelo y las hipótesis que se han considerado.

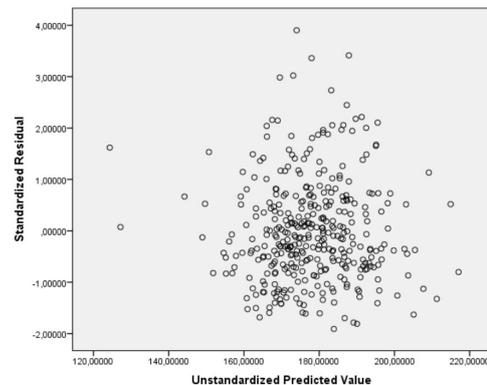
b. ¿Cuál sería la tasa de contaminante estimada con el modelo de regresión para una ciudad con 3 millones de habitantes, un índice de pluviosidad de 250 y una densidad de industria contaminante de 25?

c. ¿Qué variables son individualmente explicativas? ¿El modelo es globalmente explicativo? Dar respuestas razonadas al nivel de significación  $\alpha = 0,05$ . ¿Se detecta algún problema de multicolinealidad?

8.- En un estudio sobre el efecto del colesterol sobre el peso (en libras), se seleccionaron 367 personas de edades similares (200 mujeres y 167 hombres; a los hombres se les asignó el valor 0 y a las mujeres se les asignó el valor 1) a los que se les midió el colesterol, el peso y la estatura (en pulgadas). Los resultados obtenidos fueron:

**Descriptivos**

| Colesterol | hombres | Media               | 207,5749 |
|------------|---------|---------------------|----------|
|            |         | Desviación estándar | 45,09073 |
|            | mujeres | Media               | 208,3500 |
|            |         | Desviación estándar | 44,55393 |
| Altura     | hombres | Media               | 69,1257  |
|            |         | Desviación estándar | 2,93130  |
|            | mujeres | Media               | 63,7200  |
|            |         | Desviación estándar | 2,80194  |
| Peso       | hombres | Media               | 182,4072 |
|            |         | Desviación estándar | 38,34987 |
|            | mujeres | Media               | 174,0450 |
|            |         | Desviación estándar | 41,78640 |



**ANOVA<sup>a</sup>**

| Modelo |           | Suma de cuadrados | gl  | Media cuadrática | F      | Sig.              |
|--------|-----------|-------------------|-----|------------------|--------|-------------------|
| 1      | Regresión | 54450,564         | 3   | 18150,188        | 12,122 | ,000 <sup>b</sup> |
|        | Residuo   | 543526,193        | 363 | 1497,317         |        |                   |
|        | Total     | 597976,758        | 366 |                  |        |                   |

a. Variable dependiente: Peso

**Coefficientes<sup>a</sup>**

| Modelo |             | Coefficients no estandarizados |                | Coefficients estandarizados |  | t      | Sig. |
|--------|-------------|--------------------------------|----------------|-----------------------------|--|--------|------|
|        |             | B                              | Error estándar | Beta                        |  |        |      |
| 1      | (Constante) | -102,113                       | 50,346         |                             |  | -2,028 | ,043 |
|        | Colesterol  | ,070                           | ,045           | ,077                        |  | 1,540  | ,124 |
|        | Altura      | 3,907                          | ,708           | ,380                        |  | 5,514  | ,000 |
|        | Sexo        | 12,702                         | 5,578          | ,157                        |  | 2,277  | ,023 |
|        |             |                                |                |                             |  |        |      |

a. Variable dependiente: Peso

- a. Indica el modelo utilizado en este análisis identificando claramente las variables que intervienen.  
 b. Indica las hipótesis del contraste realizado en la tabla ANOVA y la conclusión que obtienes.  
 c. Indica los contrastes realizados en la tabla de coeficientes y las conclusiones que obtienes.  
 d. ¿Qué peso esperaríamos de una mujer de 61 pulgadas de altura y con 195 unidades de colesterol?  
 e. ¿Encuentras algo destacable en el diagrama de dispersión de los residuos?  
 Realizado el estudio por separado par a hombres y mujeres, se obtuvieron los siguientes resultados:

**Coefficientes para los hombres<sup>a</sup>**

| Modelo |             | Coeficientes no estandarizados |                | Coeficientes estandarizados |        | Sig. |
|--------|-------------|--------------------------------|----------------|-----------------------------|--------|------|
|        |             | B                              | Error estándar | Beta                        | t      |      |
| 1      | (Constante) | -122,072                       | 66,304         |                             | -1,841 | ,067 |
|        | Colesterol  | ,072                           | ,064           | ,076                        | 1,126  | ,261 |
|        | Altura      | 4,413                          | 1,013          | ,296                        | 4,357  | ,000 |

a. Variable dependiente: Peso

**Coefficientes para las mujeres<sup>a</sup>**

| Modelo |             | Coeficientes no estandarizados |                | Coeficientes estandarizados |       | Sig. |
|--------|-------------|--------------------------------|----------------|-----------------------------|-------|------|
|        |             | B                              | Error estándar | Beta                        | t     |      |
| 1      | (Constante) | -62,926                        | 70,372         |                             | -,894 | ,373 |
|        | Colesterol  | ,066                           | ,064           | ,078                        | 1,031 | ,304 |
|        | Altura      | 3,350                          | ,987           | ,256                        | 3,393 | ,001 |

a. Variable dependiente: Peso

- f. Escribe y compara los 2 planos de regresión ajustados para los hombres y los 2 ajustados para mujeres.  
 g. Utilizando los 3 modelos (el conjunto y los separados por sexo), indica qué peso esperaríamos de una mujer de 61 pulgadas de altura y 195 unidades de colesterol y de un hombre con 70 pulgadas de altura y 200 de colesterol.

#### MODELO DE REGRESIÓN LOGÍSTICA

1.- El 4 de julio de 1999 una tormenta con vientos que excedían las 90 millas por hora azotó el nordeste de Minnesota, en EE. UU., causando graves daños en los bosques de un parque natural de la zona. Los científicos analizaron los efectos de la tormenta determinando para más de 3600 árboles del parque su diámetro en cm (variable  $D$ ), y una medida de la severidad local de la tormenta (variable  $S$ ). Los árboles que habían muerto fueron codificados con  $Y = 1$ , y los árboles supervivientes con  $Y = 0$ . Utilizaremos un modelo de regresión logística para expresar la probabilidad de no supervivencia de los árboles mediante las variables  $D$  y  $S$ . Usando SPSS, se obtiene la siguiente tabla:

**Variables en la ecuación**

|                       | B      | E.T. | Wald    | gl | Sig. | Exp(B) |
|-----------------------|--------|------|---------|----|------|--------|
| Paso 1 <sup>a</sup> D | ,097   | ,005 | 346,022 | 1  | ,000 | 1,102  |
| S                     | 4,424  | ,189 | 545,122 | 1  | ,000 | 83,412 |
| Constante             | -3,543 | ,127 | 774,463 | 1  | ,000 | ,029   |

a. Variable(s) introducida(s) en el paso 1: D, S.

Se pide:

- a. Hallar las estimaciones puntuales de los parámetros del modelo.  
 b. Obtener un intervalo de confianza de nivel 95 % para el coeficiente del diámetro,  $\beta_1$ .  
 c. La variable  $D$ , ¿tiene una influencia significativa en el modelo? ¿Y la variable  $S$ ? Dar respuestas al nivel de significación del 5%.  
 d. Estimar la probabilidad de que no sobreviva un árbol cuyo diámetro es de 30 cm situado en una zona en que la fuerza de la tormenta viene dada por  $S = 0,8$ .

Con los mismos datos, ajustamos un modelo de regresión logística simple que incluye sólo el diámetro como variable regresora para explicar la probabilidad de no supervivencia de los árboles, obteniendo la siguiente tabla con SPSS:

**Variables en la ecuación**

|                       | B      | E.T. | Wald    | gl | Sig. | Exp(B) |
|-----------------------|--------|------|---------|----|------|--------|
| Paso 1 <sup>a</sup> D | ,098   | ,005 | 405,223 | 1  | ,000 | 1,102  |
| Constante             | -1,702 | ,083 | 422,608 | 1  | ,000 | ,182   |

a. Variable(s) introducida(s) en el paso 1: D.

- e. A nivel  $\alpha = 0,01$ , ¿es posible afirmar que el diámetro influye en la probabilidad de que un árbol sobreviva?  
 f. Escribe la regla para clasificar un árbol como superviviente o no superviviente en función de su diámetro.

2.- Se dispone de medidas de la longitud y la anchura del pétalo y del sépalo de 100 lirios correspondientes a dos especies diferentes: *Iris versicolor* ( $Y = 0$ ) e *Iris virginica* ( $Y = 1$ ). Se ha ajustado un modelo de regresión logística a los datos con el fin de estudiar la probabilidad de que un lirio pertenezca a cada una de las dos especies en función de las cuatro medidas. Los resultados más relevantes obtenidos con SPSS se muestran a continuación:

**Variables en la ecuación**

|                             | B       | E.T.   | Wald  | gl | Sig. | Exp(B)    |
|-----------------------------|---------|--------|-------|----|------|-----------|
| Paso 1 <sup>a</sup> LSepalo | -2,465  | 2,394  | 1,060 | 1  | ,303 | ,085      |
| ASepalo                     | -6,681  | 4,480  | 2,224 | 1  | ,136 | ,001      |
| LPetalo                     | 9,429   | 4,737  | 3,962 | 1  | ,047 | 12448,870 |
| APetalo                     | 18,286  | 9,743  | 3,523 | 1  | ,061 | 8,741E7   |
| Constante                   | -42,638 | 25,708 | 2,751 | 1  | ,097 | ,000      |

a. Variable(s) introducida(s) en el paso 1: LSepalo, ASepalo, LPetalo, APetalo.

- a. A nivel  $\alpha = 0,05$ , ¿qué variables son significativas?  
 b. Calcula un intervalo de confianza de nivel 95 % para el coeficiente correspondiente a la anchura del sépalo.  
 c. Escribe la regla de clasificación lineal que proporciona el modelo con las cuatro variables. Usando esta regla, ¿en cuál de las dos especies se clasifica un lirio tal que la longitud de su pétalo es 5 cm, la anchura de su pétalo es 2 cm, la longitud de su sépalo es 6 cm y la anchura de su sépalo es 3 cm?

3.- La altura (en cm) es una variable que se puede utilizar para intentar clasificar con rapidez a un chimpancé en una de las dos especies existentes: chimpancé común (que codificaremos con 1) y bonobo (que codificaremos con 0). A partir de las 10 siguientes alturas:

|                    |     |     |     |     |     |
|--------------------|-----|-----|-----|-----|-----|
| BONOBOS            | 94  | 97  | 100 | 105 | 108 |
| CHIMPANCÉS COMUNES | 102 | 108 | 112 | 114 | 120 |

obtenemos con SPSS la siguiente tabla:

**Variables en la ecuación**

|                            | B       | E.T.   | Wald  | gl | Sig. | Exp(B) |
|----------------------------|---------|--------|-------|----|------|--------|
| Paso 1 <sup>a</sup> Altura | ,308    | ,187   | 2,701 | 1  | ,100 | 1,361  |
| Constante                  | -32,608 | 19,860 | 2,696 | 1  | ,101 | ,000   |

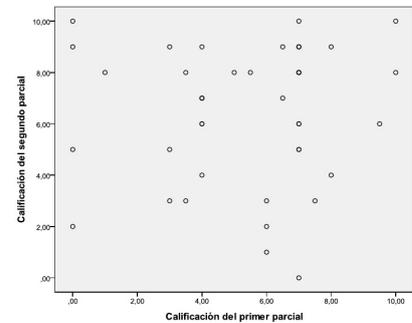
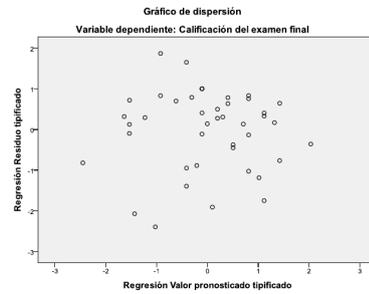
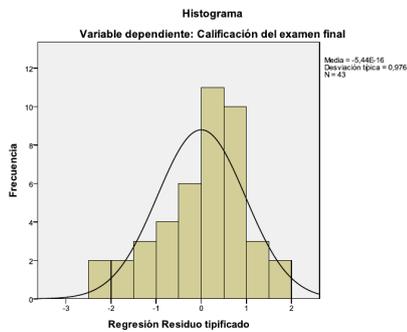
a. Variable(s) introducida(s) en el paso 1: Altura.

- a. Escribe el modelo de regresión logística ajustado que expresa la probabilidad de clasificar un chimpancé como chimpancé común en función de su altura. ¿Cuánto vale esa probabilidad si el chimpancé tiene una altura de 110 cm?  
 b. Escribe la regla de clasificación de los chimpancés y aplícala a un chimpancé con una altura de 100 cm.

4.- En una asignatura con 43 alumnos se realizan dos exámenes parciales (calificado cada uno de ellos de 0 a 10) y una evaluación final (calificada también de 0 a 10). Estos datos van a ser analizados con diferentes modelos estadísticos:

**MODELO DE REGRESIÓN LINEAL MÚLTIPLE**

Este modelo lo vamos a aplicar para intentar predecir la calificación de la evaluación final mediante las calificaciones de los dos parciales. Las siguientes tablas y gráficos son obtenidos con SPSS:



### ANOVA<sup>b</sup>

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F     | Sig.              |
|--------|-----------|-------------------|----|------------------|-------|-------------------|
| 1      | Regresión | 13,889            | 2  | 6,945            | 3,835 | ,030 <sup>a</sup> |
|        | Residual  | 72,435            | 40 | 1,811            |       |                   |
|        | Total     | 86,324            | 42 |                  |       |                   |

- a. Variables predictoras: (Constante), Calificación del segundo parcial, Calificación del primer parcial  
 b. Variable dependiente: Calificación del examen final

### Coefficientes<sup>a</sup>

| Modelo |                                  | Coefficients no estandarizados |            | Coefficients tipificados | t     | Sig. |
|--------|----------------------------------|--------------------------------|------------|--------------------------|-------|------|
|        |                                  | B                              | Error típ. | Beta                     |       |      |
| 1      | (Constante)                      | 5,050                          | ,671       |                          | 7,526 | ,000 |
|        | Calificación del primer parcial  | ,117                           | ,079       | ,214                     | 1,473 | ,148 |
|        | Calificación del segundo parcial | ,176                           | ,079       | ,324                     | 2,231 | ,031 |

- a. Variable dependiente: Calificación del examen final

a. Plantear todos los elementos e hipótesis previas del modelo, y diagnosticar dichas hipótesis utilizando los gráficos disponibles.

b. ¿Es el modelo conjuntamente explicativo? ¿Son las calificaciones de cada parcial individualmente explicativas? Plantear los contrastes de hipótesis adecuados, responder razonadamente (con un nivel de significación del 5%), y obtener las conclusiones que parezcan oportunas.

### MODELO DE REGRESIÓN LINEAL SIMPLE

Este modelo lo vamos a aplicar para intentar predecir la calificación de la evaluación final utilizando solamente la calificación del segundo parcial. Las siguientes tablas son obtenidas con SPSS:

### Coefficientes<sup>a</sup>

| Modelo |                                  | Coefficients no estandarizados |            | Coefficients tipificados | t      | Sig. |
|--------|----------------------------------|--------------------------------|------------|--------------------------|--------|------|
|        |                                  | B                              | Error típ. | Beta                     |        |      |
| 1      | (Constante)                      | 5,623                          | ,555       |                          | 10,141 | ,000 |
|        | Calificación del segundo parcial | ,184                           | ,080       | ,340                     | 2,312  | ,026 |

- a. Variable dependiente: Calificación del examen final

### ANOVA<sup>b</sup>

| Modelo |           | Suma de cuadrados | gl | Media cuadrática | F     | Sig.              |
|--------|-----------|-------------------|----|------------------|-------|-------------------|
| 1      | Regresión | 9,958             | 1  | 9,958            | 5,346 | ,026 <sup>a</sup> |
|        | Residual  | 76,366            | 41 | 1,863            |       |                   |
|        | Total     | 86,324            | 42 |                  |       |                   |

- a. Variables predictoras: (Constante), Calificación del segundo parcial  
 b. Variable dependiente: Calificación del examen final

**Estadísticos descriptivos**

|                                  | Media  | Desviación típica | N  |
|----------------------------------|--------|-------------------|----|
| Calificación del examen final    | 6,8116 | 1,43364           | 43 |
| Calificación del segundo parcial | 6,4419 | 2,63947           | 43 |

**c.** ¿Tiene la calificación del segundo parcial una influencia significativa sobre la calificación de la evaluación final? Plantear el contraste de hipótesis adecuado, y dar una respuesta razonada (con un nivel de significación del 5%).

**d.** Predecir (mediante un intervalo al 95 %) la calificación de la evaluación final de un único alumno que haya obtenido un 6 en el segundo parcial.

**MODELO DE REGRESIÓN LOGÍSTICA**

Aplicamos este modelo a la predicción (aproximada) de la probabilidad de que un alumno obtenga más de un 7 en la evaluación final a partir de las calificaciones de los dos parciales. Para hacer esto, codificaremos con  $Y = 1$  a los alumnos que obtienen más de 7 en la evaluación final, y con  $Y = 0$  a los que obtienen menos. La siguiente tabla se obtiene con SPSS:

**VARIABLES EN LA ECUACIÓN**

|                              | B      | E.T.  | Wald  | gl | Sig. | Exp(B) |
|------------------------------|--------|-------|-------|----|------|--------|
| Paso 1 <sup>a</sup> Parcial1 | ,212   | ,135  | 2,474 | 1  | ,116 | 1,236  |
| Parcial2                     | ,247   | ,133  | 3,434 | 1  | ,064 | 1,280  |
| Constante                    | -2,573 | 1,203 | 4,570 | 1  | ,033 | ,076   |

a. Variable(s) introducida(s) en el paso 1: Parcial1, Parcial2.

**e.** Utilizar el modelo ajustado para calcular la probabilidad de que un alumno, que ha obtenido un 7 en el primer parcial y un 6 en el segundo, obtenga más de un 7 en la evaluación final.