Problemas de repaso: temas 3 y 4

1. Se desea estimar la contaminación (en partes por millón) por metales pesados en el pescado que se vende en una lonja. Para ello se ha seleccionado una pequeña muestra de 12 unidades resultando que la contaminación media observada en ellos fue de 0.79 con una cuasi-desviación típica de 0.13. ¿Qué tamaño muestral habría que utilizar para estimar la contaminación media con un error máximo de 0.01 a un nivel de confianza de 0.95?

2. En una población de cierta especie el número de hijos de cada hembra es una variable aleatoria X que toma el valor 0 con probabilidad p, toma el valor 1 con probabilidad $2p$ y vale 2 con probabilidad $1-3p$. En una muestra de 50 hembras, hubo 20 sin hijos, 20 tuvieron uno y 10 tuvieron dos.
   (a) Calcula un estimador de p utilizando el método de momentos.
   (b) Calcula un estimador de p utilizando el método de máxima verosimilitud.

3. Con el fin de valorar la concentración de contaminantes orgánicos se ha medido la demanda biológica de oxígeno (DBO) en 10 muestras de agua tomadas en una estación experimental. La media resultante para las 10 muestras fue de 20 miligramos de oxígeno diatómico por litro (mgO_2/l) y la cuasivarianza para las 10 medidas fue de 8.
   (a) Calcula un intervalo de confianza de nivel 95 % para la DBO media de la estación experimental. (Se supone que las medidas tienen distribución normal).
   (b) Con un nivel de confianza del 90 % se desea garantizar que la diferencia entre la DBO media de la estación y la DBO media estimada con las muestras sea inferior a 0,5 mgO_2/l. ¿Cuántas muestras de agua se necesitan como mínimo?

4. Se ha producido una manifestación en una gran ciudad. Cada ciudadano que la apoyaba tenía una probabilidad del 50 % de acudir a la misma. Cada ciudadano que no la apoyaba tenía una probabilidad del 2 % de ir a la manifestación por diversas razones (curiosidad, presión del grupo, etc.)
   (a) Halla la relación entre los siguientes números:
   $p =$ proporción de ciudadanos que apoyan la manifestación
   $q =$ proporción de ciudadanos que acuden a la manifestación
   (b) En una muestra aleatoria de 1000 personas, resulta que 68 personas han asistido a dicha manifestación. Halla la estimación de máxima verosimilitud para q y deduce una estimación para p.

5. Se lleva a cabo un estudio sobre la longitud (medida desde la punta del pico hasta el final de la cola) de la cigüeña blanca en una colonia de la provincia de Lérida (en la que se están perdiendo los hábitos migratorios). Se obtienen los siguientes datos:

Tamaño muestral	Media muestral	Cuasi-varianza muestral
14	115	30

   Asumiendo normalidad,
   (a) Halla el intervalo de confianza (al nivel 0,90) para la longitud media de toda la colonia.
   (b) ¿Cuántas cigüeñas necesitaríamos en la muestra para estimar la longitud media con un error menor que 1 cm (al nivel de confianza 0,90)?

6. En una pequeña muestra de 20 árboles seleccionados al azar en un inmenso bosque, encontramos que 8 de ellos son cerezos.
   (a) Calcula el intervalo de confianza para estimar la proporción de cerezos en ese bosque (con un nivel de confianza del 90 %).
   (b) ¿Cuántos árboles tendríamos que muestrear en total para poder estimar la proporción de cerezos con un error inferior a 0,01 y con el mismo nivel de confianza del 90 %?

7. Se están estudiando dos colonias de ñúes azules, una que vive en un parque de Tanzania, y otra que vive en un parque de Kenia. Se estudia una muestra de 10 ñúes en Tanzania, obteniéndose una altura media muestral de 130 cm con una cuasi-varianza muestral de 80, y otra muestra de 15 ñúes en Kenia, obteniéndose una altura media muestral de 124 cm con una cuasi-varianza muestral de 75. Asumiendo normalidad para las alturas en las dos colonias, se pide calcular un intervalo de confianza de nivel 95 % para la diferencia de las alturas medias poblacionales en las dos colonias.

8. Se recomienda que una persona mayor de 50 años consuma 15 mg de zinc al día en su dieta. En un informe sobre los hábitos alimentarios de una muestra de 100 individuos mayores de 50 años se señala que éstos consumieron una media de 11,3 mg de zinc diarios. La cuasidesviación típica muestral correspondiente a estos datos fue de 6,43. Se supone que los datos siguen una distribución normal. Calcula un intervalo de confianza de nivel 95 % para la media poblacional y otro para la varianza poblacional de la ingesta diaria de zinc.

9. En un estudio se informa de los resultados de un experimento diseñado para comparar dos tratamientos de cierto tipo de cáncer, uno de ellos a base de quimioterapia y el otro basado en una combinación de quimioterapia y radioterapia (*New Engl. J. Med.*, 1997, pags. 956–962). De 154 individuos que recibieron únicamente quimioterapia, 76 sobrevivieron más de 15 años mientras que 98 de los 164 individuos que recibieron la combinación de quimioterapia y radioterapia sobrevivieron más de 15 años. Calcula un intervalo de confianza de nivel 95 % para la diferencia de las probabilidades de sobrevivir más de 15 años para ambos tratamientos.

10. Queremos comparar dos métodos rápidos para estimar la concentración de una hormona en una solución. Tenemos 10 dosis preparadas en el laboratorio y vamos a medir la concentración de cada una con los dos métodos. Se obtienen los siguientes resultados:

Dosis	1	2	3	4	5	6	7	8	9	10
Método A	10,7	11,2	15,3	14,9	13,9	15,0	15,6	15,7	14,3	10,8
Método B	11,1	11,4	15,0	15,1	14,3	15,4	15,4	16,0	14,3	11,2

Calcula un intervalo de confianza de nivel 95 % para la diferencia de las concentraciones medias medidas con ambos métodos.

11. Como parte de un estudio para la reducción de los gases de efecto invernadero que emiten los coches, se estudian los efectos de un determinado aditivo que reduce las emisiones. Sea X el número de kilómetros recorridos por un coche con un litro de gasolina sin el aditivo. Sea Y el número de kilómetros recorridos con un litro de gasolina con el aditivo. Se observan los kilómetros recorridos por litro de gasolina en ocho coches, cuatro de ellos sin aditivo. Los datos que se obtienen son:

$$\sum_{i=1}^{4} x_i = 25{,}4 \qquad \sum_{i=1}^{4} y_i = 31{,}2 \qquad \sum_{i=1}^{4} x_i^2 = 173{,}53 \qquad \sum_{i=1}^{4} y_i^2 = 261{,}22$$

Suponiendo que el aditivo puede cambiar la media pero no la varianza, y especificando las hipótesis necesarias, calcula un intervalo de confianza al 95 % para la diferencia de medias.