ESTADÍSTICA
Segundo curso de Biología
Curso 2015-2016

ESTADÍSTICA DESCRIPTIVA

1. Estamos interesados en la variable X=“Tiempo de vida (en días)” de una especie de insectos.

   a) En una muestra pequeña de 11 insectos, los resultados muestrales fueron:

   20, 25, 13, 18, 32, 25, 20, 15, 28, 40, 27

   Hallar el tiempo medio de vida, el tiempo mediano de vida, y la desviación típica.

   b) En una muestra grande, los resultados obtenidos se resumen de la siguiente forma:

Percentil	30	50	70	100
Tiempo de vida	18	22	26	30

   Hallar el tiempo medio de vida (indicando previamente las clases, marcas de clase y frecuencias proporcionadas por la información muestral).

2. Con el fin de controlar la contaminación de un río, todas las semanas se hace una medición del nivel de ácido úrico.

   a) Las mediciones durante 9 semanas fueron:

   13 10 7 5 12 7 9 5 5

   Hallar el nivel medio de ácido úrico, el nivel mediano, y la desviación típica.

   b) En un estudio más completo, las mediciones semanales de ácido úrico se resumieron de la siguiente forma:

Percentil	20	40	70	100
Nivel de ácido úrico	6	8	12	18

   Hallar el nivel medio de ácido úrico y dibujar el histograma (indicando previamente las clases, marcas de clase y frecuencias proporcionadas por la información muestral).

3. Los babuinos constituyen un género de primates catarrinos de la familia *Cercopithecidae*. Se lleva a cabo un amplio estudio con 100 babuinos macho y los resultados sobre su altura hasta los hombros se resumen en la siguiente tabla:

Altura en cm	Entre 40 y 50	Entre 50 y 55	Entre 55 y 60	Entre 60 y 70
Número de individuos	19	28	30	23

   (a) Calcular la media muestral de estos datos.

   (b) Representar los datos en un histograma.

   (c) Calcular (aproximadamente) el primer cuartil y la mediana de los datos.

4. En una explotación ganadera, se midió la cantidad de triacilglicéridos (en gramos por litro) en la leche de 100 vacas, obteniéndose los siguientes resultados:

Gramos por litro	Entre 26 y 28	Entre 28 y 30	Entre 30 y 32	Entre 32 y 40
Número de reses	10	30	35	25

   (a) Obtener el número medio de gramos de triacilglicéridos por litro

   (b) Representar estos datos mediante un histograma.

   (c) Obtener (aproximadamente) el percentil 20 de estos datos.

5. Se piensa que la frecuencia del canto de los grillos (en una hora) tiene una estrecha relación con la temperatura. Para poder expresar la frecuencia del canto (Y) en función de la temperatura en grados Farenheit (X) se recogen cinco pares de datos que se resumen a continuación:

$$\sum x_i = 394 \qquad \sum y_i = 84 \qquad \sum x_i^2 = 31316 \qquad \sum y_i^2 = 1426 \qquad \sum x_i y_i = 6672$$

   (a) Obtener la recta de regresión para expresar la frecuencia del canto en función de la temperatura.

   (b) Evaluar el ajuste. La asociación entre X e Y, ¿es positiva o negativa?

6. El peso de los animales salvajes es siempre más difícil de determinar que su altura. Para poder expresar (aproximadamente) el peso de los ñúes azules (connochaetes taurinus) en función de su altura, se apresan 10 ejemplares adultos, y se anota su altura en cm (X) y su peso en Kg (Y). Los resultados se resumen a continuación:

$$\sum x_i = 1255 \qquad \sum y_i = 2275 \qquad \sum x_i^2 = 158325 \qquad \sum y_i^2 = 519875 \qquad \sum x_i y_i = 286850$$

   (a) Con los datos anteriores, halla la recta de regresión del peso (Y) sobre la altura (X).

   (b) ¿Cuál sería el peso aproximado de un ñu cuya altura es de 125 cm?

7. El equipo del profesor Manalis ha desarrollado recientemente, en el Instituto Tecnológico de Massachusetts (MIT), un dispositivo para medir la masa de partículas minúsculas como, por ejemplo, nanopartículas sintéticas y componentes biológicos de las celulas, cuyas masas se miden en attogramos (1 attogramo = 10^{-18} gramos).

   Con la idea de calibrar este dispositivo, se mide la masa de 10 nanopartículas sintéticas con un dispositivo más antiguo (X) y con el nuevo dispositivo (Y), obteniéndose una masa media de 46,4 attogramos con el antiguo dispositivo, y una masa media con el nuevo dispositivo de 50,2 attogramos. Además:

$$\sum x_i^2 = 22390 \qquad \sum y_i^2 = 26142 \qquad \sum x_i y_i = 24174$$

   (a) Halla la recta de regresión para expresar Y en función de X.

   (b) Estima cuanto valdría la masa de una nanopartícula con el nuevo procedimiento, si con el antiguo ha resultado ser de 60 attogramos.

   (c) Calcula y comenta el coeficiente de correlación entre X e Y.

8. (Ordenador) En 1778, H. Cavendish realizó una serie de 29 experimentos con objeto de medir la densidad de la tierra. Sus resultados, tomando como unidad la densidad del agua, fueron:

5’50	5’61	4’88	5’07	5’26	5’55	5’36	5’29	5’58	5’65
5’57	5’53	5’62	5’29	5’44	5’34	5’79	5’10	5’27	5’39
5’42	5’47	5’63	5’34	5’46	5’30	5’75	5’68	5’85	

   a) Representa los datos por medio de un diagrama de tallos y hojas.

   b) Representa los datos por medio de un diagrama de caja y bigotes.

   c) Halla la media y la desviación típica.

9. (Ordenador) Para evaluar la viabilidad de un proyecto de reforestación de una zona sometida a una fuerte actividad turística, se analiza la composición en mg por cm^3 de desechos orgánicos del territorio. Los datos que se obtienen son:

10.87	9.01	22.50	12.35	17.39	31.05	17.19	16.74	20.33
19.32	23.18	25.15	15.49	20.30	2.38	13.55	9.33	22.72
10.96	25.90	27.66	9.74	18.65	9.31	24.60	17.41	24.86
15.34	23.34	22.81	17.86	30.72	32.60	8.96	32.71	15.86
16.71	5.48	8.25	20.57	4.57	2.30	32.56	7.92	4.84
4.57	26.45	23.58	19.27	9.79	3.03	19.40	23.92	22.45
22.05	21.18	18.85	8.38	15.01	18.12	4.24	3.39	7.17
22.71	22.44	15.89	24.20	24.75	28.08	19.73	13.22	17.69
5.53	11.42	5.58	3.15	14.06	5.83	19.42	21.13	18.32
23.31	11.89	23.95	19.30	12.22	21.45	9.84	4.78	38.63
12.65	13.89	23.82	16.91	28.09	15.73	12.53	16.52	9.48
4.08								

Efectuar un estudio descriptivo como en el ejercicio anterior.

10. (Ordenador) El maíz es un alimento importante para los animales. De todas formas, este alimento carece de algunos aminoácidos que son esenciales. Un grupo de científicos desarrolló una nueva variedad que sí contenía niveles apreciables de dichos aminoácidos. Para comprobar la utilidad de esta nueva variedad para la alimentación animal se llevó a cabo el siguiente experimento: a un grupo de 20 pollos de 1 día se les suministró un pienso que contenía harina de maíz de la nueva variedad. A otro grupo de 20 pollos (grupo de control) se le alimentó con un pienso que sólo se diferenciaba del anterior en que no contenía harina de la variedad mejorada de maíz. Los resultados que se obtuvieron sobre las ganancias de peso de los pollos (en gramos) al cabo de 21 días de alimentación fueron los siguientes:

    - *Variedad normal*
      380 321 366 356 283 349 402 462 356 410 329 399 350 384 316 272 345 455 360 431
    - *Variedad mejorada*
      361 447 401 375 434 403 393 426 406 318 467 407 427 420 477 392 430 339 410 326

    a) Para comparar las dos distribuciones, representa los dos diagramas de caja y bigotes en un mismo gráfico. ¿Qué se puede deducir de estos diagramas?

    b) ¿Cuáles son las medias y desviaciones típicas de los datos de ambos grupos? ¿Qué diferencias hay entre ambos?

11. (Ordenador) Las autoridades sanitarias de un municipio están interesadas en evaluar la calidad del agua para consumo en términos de colonias de bacterias tróficas en un acuífero próximo a la ciudad. Se consideran dos zonas diferentes del acuífero y se obtienen los siguientes resultados (número de colonias por 1000 mm de agua):

zona 1	194	199	191	202	215	214	197
	204	199	202	230	193	194	209
zona 2	158	161	143	174	220	156	156
	156	198	161	188	139	147	116

    Realizar un estudio comparativo de la calidad del agua en ambas zonas, utilizando resúmenes numéricos y diagramas de cajas. ¿Se puede considerar que ambas zonas son similares?

12. (Ordenador) Se tienen dos métodos, A y B, para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos (en calorías por gramo de masa para pasar de -0.72°C a 0°C) utilizando ambos métodos independientemente:

Método A	79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02
Método B	80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

    a) Hacer una comparación descriptiva de los dos métodos.

    b) Si por un error al transcribir los datos, se modifica el primer dato de cada método y se escribe 799.8 y 800.2, ¿cómo varían la media y la mediana?

13. (Ordenador) Los manatíes son unos animales grandes y dóciles que viven a lo largo de la costa de Florida. Cada año las lanchas motoras hieren o matan muchos de ellos. A continuación se presenta una tabla que contiene, para cada año, el número de licencias para motoras (expresado en miles de licencias) expedidas en Florida y el número de manatíes muertos en los años 1977 a 1990.

    a) Queremos analizar la relación entre el número de licencias expedidas anualmente en Florida y el número de manatíes muertos. ¿ Cual es la variable explicativa? Dibuja un diagrama de dispersión con esos datos. ¿Qué nos dice el diagrama sobre la relación entre esas dos variables? Las variables ¿están asociadas positiva o negativamente?

    b) Halla la recta de regresión para expresar el número de manatíes muertos en función del número de licencias.

    c) Describe la fuerza de la relación. ¿Se puede predecir con precisión el número de manatíes muertos cada año conociendo el número de licencias expedidas ese año? Si Florida decidiera congelar el número de licencias en 700.000, ¿cuántos manatíes matarían, aproximadamente, las lanchas motoras?

Año	Licencias	Manatíes	Año	Licencias	Manatíes
1977	447	13	1984	559	34
1978	460	21	1985	585	33
1979	481	24	1986	614	33
1980	498	16	1987	645	39
1981	513	24	1988	675	43
1982	512	20	1989	711	50
1983	526	15	1990	719	47

14. (Ordenador) Los corredores buenos dan más pasos por segundo a medida que aumentan la velocidad. He aquí el promedio de pasos por segundo de un grupo de corredoras de élite a distintas velocidades. La velocidad se expresa en metros por segundo.

Velocidad (m/s)	$4,83$	$5,14$	$5,33$	$5,67$	$6,08$	$6,42$	$6,74$
Pasos por segundo	$3,05$	$3,12$	$3,17$	$3,25$	$3,36$	$3,46$	$3,55$

    a) Se quiere predecir el número de pasos por segundo a partir de la velocidad. En primer lugar, dibuja un diagrama de dispersión.

    b) Halla la recta de regresión del número de pasos por segundo con relación a la velocidad.

    c) Halla el coeficiente de correlación lineal.

15. (Ordenador) La tabla siguiente presenta tres conjuntos de datos preparados por el estadístico Frank Anscombe para ilustrar los peligros de hacer cálculos sin antes representar los datos. *Los tres conjuntos de datos tienen la misma correlación y la misma recta de regresión.*

    Conjunto de datos A:

X	10	8	13	9	11	14	6	4	12	7	5
Y	$8,04$	$6,95$	$7,58$	$8,81$	$8,33$	$9,96$	$7,24$	$4,26$	$10,84$	$4,82$	$5,68$

    Conjunto de datos B:

X	10	8	13	9	11	14	6	4	12	7	5
Y	$9,14$	$8,14$	$8,74$	$8,77$	$9,26$	$8,10$	$6,13$	$3,10$	$9,13$	$7,26$	$4,74$

    Conjunto de datos C:

X	8	8	8	8	8	8	8	8	8	8	19
Y	$6,58$	$5,76$	$7,71$	$8,84$	$8,47$	$7,04$	$5,25$	$5,56$	$7,91$	$6,89$	$12,50$

a) Calcular la correlación y la recta de regresión para los tres conjuntos de datos y comprobar que son iguales.

b) Dibujar un diagrama de dispersión para cada uno de los conjuntos de datos con las rectas de regresión correspondientes.

c) ¿En cuál de los tres casos utilizaríamos la recta de regresión para predecir Y dado $X = 14$.

Conclusión: Representa siempre tus datos.

MODELOS DE PROBABILIDAD

1. El "tiempo de vida activa (en días)" de un plaguicida, X, viene representado por la función de densidad:
$$f(x) = \begin{cases} \frac{1}{500}e^{-\frac{x}{500}} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$
Calcular la mediana del tiempo de vida activa. ¿Cuál es su significado?

2. La variable aleatoria X="Tiempo transcurrido (en horas) hasta el fallo de una pieza" tiene función de densidad
$$f(x) = \begin{cases} \frac{1}{15000}\; e^{-\;\frac{x}{15000}} & \text{si } x > 0 \\ 0 & \text{en el resto.} \end{cases}$$
a) Calcular el tiempo medio transcurrido hasta el fallo.

b) Calcular el porcentaje de piezas que duran entre 10000 y 15000 horas.

3. Suponiendo que la probabilidad de que un niño que nace sea varón es 0,50, hallar la probabilidad de que una familia de 6 hijos tenga

a) por lo menos una niña,

b) por lo menos un niño,

c) por lo menos dos niños y una niña.

4. Una compañía de seguros con 10000 asegurados halla que el 0,005% de la población fallece cada año de un cierto tipo de accidente.

a) Hallar la probabilidad de que la compañía tenga que pagar a más de tres asegurados, por dicho accidente, en un año determinado.

b) ¿Cuál es el número medio de accidentes por año?

5. La probabilidad de que un individuo tenga una reacción alérgica al inyectarle un suero es 0,001. Hallar la probabilidad de que en 2000 individuos tengan reacción alérgica

a) exactamente tres,

b) más de 2.

6. Se considera que la variable aleatoria "Kg. de algodón recogidos por parcela" sigue una distribución $N(\mu = 100; \sigma = 10)$.

Hallar el porcentaje de parcelas en las que el número de Kg. recogidos será inferior a 115.

7. En 1969 se descubrió que los faisanes de Montana (Estados Unidos) padecían una apreciable contaminación por mercurio debida a que habían comido semillas tratadas para su crecimiento con metilo de mercurio. Se sabe que el nivel de mercurio (medido en ppm) de un faisán seleccionado aleatoriamente en la población es una variable aleatoria con distribución $N(\mu = 0,25; \sigma = 0,10)$.

(a) Calcula la probabilidad de que, al seleccionar aleatoriamente un faisán de la población, su nivel de mercurio supere 0,30 ppm.

(b) Si se seleccionan aleatoria e independientemente 100 faisanes, clacula la probabilidad de que al menos 45 de ellos tengan un nivel de mercurio superior a 0,25 ppm.

(c) Si se seleccionan aleatoria e independientemente cuatro faisanes, calcula la probabilidad de que su nivel medio de mercurio sea superior a 0,30 ppm.

8. Un zoólogo estudia una cierta especie de ratones de campo. Para ello captura ejemplares de una población grande en la que la proporción de dicha especie es p.

a) Si $p = 0,30$, hallar la probabilidad de que en 6 ejemplares capturados haya al menos 2 de los que le interesan.

b) Si $p = 0,03$, calcular la probabilidad de que en 200 haya exactamente 3 de los que le interesan.

c) Si $p = 0,40$, calcular la probabilidad de que en 200 haya entre 75 y 110 de los que le interesan.

9. Se sabe que el nivel de tensión sanguínea diastólica (en mmHg) en una población es una variable con distribución normal de media $\mu = 87$ y desviación típica $\sigma = 7.5$. Un individuo se clasifica como *hipertenso* si su presión es mayor de 90 mmHg.

   (a) Calcula la probabilidad de que un individuo seleccionado al azar en esta población sea *hipertenso*.

   (b) Si se seleccionan aleatoriamente 100 individuos de la población, calcula la probabilidad aproximada de que entre ellos haya más de 40 *hipertensos*.

   (c) Calcula el valor aproximado del primer cuartil de la población, es decir, el valor Q_1 tal que la tensión sanguínea del 25% de los individuos de la población es menor que Q_1.

   (d) Si se seleccionan aleatoriamente 9 individuos de la población, calcula la probabilidad de que su tensión media sea superior a 90 mmHg.

10. La capacidad de enrollar la lengua está controlada por una pareja de genes: el gen E que determina su enrollamiento y el gen e que lo impide. El gen E es dominante, de modo que una persona Ee será capaz de enrollar la lengua.

    En una ciudad grande se sabe que aproximadamente el 40% no puede enrollar la lengua y el 60% si puede.

    Si elegimos 200 personas al azar, ¿cuál es la probabilidad de que más de 70 no puedan enrollar su lengua?

11. En una granja dedicada a la helicicultura se crían dos tipos de caracoles: el común y el romano. La velocidad (en metros por hora) del caracol común de jardín (*Helix aspersa*) sigue una distribución $N(\mu = 50;\ \sigma = 5)$. La velocidad (en metros por hora) del caracol romano (*Helix pomatia*) sigue una distribución $N(\mu = 42;\ \sigma = 5)$.

    (a) Calcular el porcentaje de caracoles comunes de jardín que recorren menos de 60 metros en un hora.

    (b) Calcular la probabilidad de que un caracol de jardín elegido al azar recorra menos espacio en una hora que un caracol romano.

12. Se sabe que los niveles de triglicéridos (en mg/dL) en una población, tanto para los hombres como para las mujeres, tienen distribución normal. Para los hombres la distribución es $N(\mu = 100; \sigma = 30)$, y para las mujeres la distribución es $N(\mu = 90; \sigma = 25)$.

    (a) Seleccionando un hombre al azar, ¿cuál es la probabilidad de que su nivel de triglicéridos sea inferior a 130 mg/dL?

    (b) Si se seleccionan aleatoria e independientemente un hombre y una mujer, ¿cuál es la probabilidad de que el nivel de triglicéridos de la mujer sea superior al del hombre?

13. Una línea eléctrica se avería cuando la tensión sobrepasa la capacidad de la línea. Si la tensión es $N(100;\ 20)$ y la capacidad es $N(140;\ 10)$, calcular la probabilidad de avería.

14. La concentración de ácido úrico en sangre (mg/dl) en la población de pacientes con síndrome de apnea-hipopnea durante el sueño (SAHS) sigue una distribución normal, $N(\mu = 6,30; \sigma = 1,50)$. Se recomienda que la concentración de ácido úrico se mantenga por debajo de 7,20 mg/dl ya que, por encima de ese valor, comienzan a surgir problemas (cálculos renales, gota, ...).

    (a) Calcula el porcentaje de pacientes de la población anterior, cuya concentración de ácido úrico se mantiene por debajo de 7,20 mg/dl.

    (b) Si se seleccionan 50 pacientes al azar en esa población, calcula la probabilidad aproximada de que más de 30 de ellos tengan un nivel de ácido úrico aceptable (por debajo de 7,20 mg/dl).

    (c) Si se seleccionan al azar dos pacientes en esa población, ¿cuál es la probabilidad de que la diferencia de ácido úrico entre ellos sea inferior a 1 mg/dl?

ESTIMACIÓN PUNTUAL

1. Dada una muestra aleatoria de tamaño n de una variable X, calcular el estimador de máxima verosimilitud y el del método de los momentos, en los siguientes casos:

   a) $X \sim$ Bernoulli de parámetro p.

   b) $X \sim$ Poisson (λ).

   c) $X \sim$ Exponencial (λ); es decir, $f_\lambda(x) = \lambda e^{-\lambda x}$, para $x > 0$ $(\lambda > 0)$.

   d) $X \sim N(\mu, \sigma)$.

2. La proporción de genes dañados en un tejido celular tras una sesión de radiación es una variable aleatoria continua X con función de densidad $f(x) = \theta x^{\theta-1}$, para valores de $x \in (0,1)$, siendo $\theta > 0$ un parámetro desconocido que depende del tipo de tejido.

   (a) Dada una muestra aleatoria $(X_1, \ldots, X_n)$ de X, calcúlese el estimador de θ por el método de los momentos y por el método de máxima verosimilitud.

   (b) Tras analizar una muestra de 3 tejidos celulares, se obtuvieron los valores $0,10$, $0,15$ y $0,25$. ¿Cuáles son la estimaciones concretas de θ con estos datos con los dos métodos?

3. En una gran piscifactoría hay una proporción desconocida, p, de cierto tipo de truchas. Para obtener información sobre esa proporción desconocida, vamos a ir sacando peces al azar hasta obtener una trucha de ese tipo, en tres ubicaciones diferentes:

   En la primera ubicación se obtiene la primera trucha de ese tipo en la décima extracción.

   En la segunda ubicación se obtiene la primera trucha de ese tipo en la decimoquinta extracción.

   En la tercera ubicación se obtiene la primera trucha de ese tipo en la decimoctava extracción.

   Escribir la función de verosimilitud y obtener la estimación de máxima verosimilitud de p.

4. Un modelo genético para las moscas de cierta variedad nos dice que pueden ser de tres tipos: homocigóticas AA (con probabilidad p^2), homocigóticas BB (con probabilidad q^2) y heterocigóticas AB (con probabilidad $2pq$), donde naturalmente $p + q = 1$.

   En una muestra aleatoria de 100 moscas obtenemos 10 de tipo AA, 50 de tipo BB, y 40 de tipo AB.

   Hallar la estimación de máxima verosimilitud de p con los datos obtenidos.

5. Una variable relacionada con el n'umero de mutaciones en una secuencia de ADN puede tomar los valores 0, 1 ó 2 con probabilidades $(1 + 2\theta)/3$, $(1 - \theta)/3$ y $(1 - \theta)/3$ respectivamente, donde θ es un parámetro desconocido. Se han obtenido 60 observaciones independientes de esta variable resultando la siguiente tabla de frecuencias absolutas:

Valores	0	1	2
Frecuencias	25	20	15

   (a) Estima el valor de θ a partir de las observaciones disponibles usando el método de los momentos.

   (b) Estima el valor de θ a partir de las observaciones disponibles usando el método de máxima verosimilitud.

6. En el artículo *"A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions"* publicado en *British Journal of Cancer* en 2013 se describe un análisis de aliento sencillo que se puede usar para el diagnóstico precoz de cáncer de estómago, a partir de la medición de cinco sustancias: 2-propenonitrilo, 2-butoxietanol, furfural, 6-metil-5-hepten-2-ona e isopreno.

   La sensibilidad de la prueba (probabilidad de dar positivo teniendo cáncer) es del 89%, y su especificidad (probabilidad de dar negativo estando sano) es del 94%.

   Este análisis de aliento se aplica a un grupo aleatorio de 50 personas de una población de riesgo, obteniéndose 10 resultados positivos y 40 negativos.

   (a) Estima q="Proporción de resultados positivos", razonando tu respuesta.

   (b) Halla la relación entre p="Proporción de personas con cáncer de estómago" en esa población de riesgo y q="Proporción de resultados positivos", y utilízala para estimar p.

7. Para estudiar la proporción p de caballos afectados por la peste equina se les va a someter a una prueba. Se sabe que la prueba resulta positiva si el animal está enfermo. Además, si el animal está sano, hay una probabilidad 0.04 de que la prueba resulte positiva.
   a) Estudia la relación entre la probabilidad p de que un caballo esté enfermo y la probabilidad q de que la prueba resulte positiva.
   b) Si se realizó la prueba a 500 caballos y resultó positiva en 95 casos, ¿cuál es el estimador de máxima verosimilitud de q? A partir del resultado del apartado (a), calcula una estimación de p.

8. Unos laboratorios desarrollan una prueba sencilla para detectar la *gripe del pollo*. La prueba tiene una fiabilidad muy aceptable: proporciona un 4% de falsos positivos (prueba positiva cuando el pollo está sano) y un 0% de falsos negativos (prueba negativa cuando el pollo está enfermo).

   En una granja avícola, se detecta un brote de *gripe del pollo*. Mediante la utilización de la prueba sencilla que se ha descrito anteriormente, se quiere estimar la incidencia de la enfermedad en esa granja. Para esto, se seleccionan al azar 100 pollos, se les efectúa la prueba y se obtienen 20 casos positivos. Estimar la proporción p de pollos enfermos en la granja, explicando todo el proceso seguido.

9. **La diferencia entre intención de voto directo y estimación de voto.**

   En cierto país con dos partidos políticos, A y B, se lleva a cabo un sondeo sobre opinión electoral (lo podemos llamar barómetro) en el que se pregunta:

   Suponiendo que mañana se celebrasen elecciones, ¿a qué partido votaría?

   Los resultados son: 23% votaría A, 27% votaría B, y 50% de indecisos. Por tanto, en la **intención de voto directo**, saldría triunfador el partido B.

   Pero hay un 50% de indecisos, de los cuales se puede intentar predecir el comportamiento, ya que en ese mismo sondeo, se hacen otras preguntas mediante las cuales se puede hacer un perfil (conservador o progresista) de las personas, y además, se les pregunta qué votaron en las últimas elecciones.

   Supongamos que los resultados de ese sondeo son los siguientes:

   (a) Del 50% de indecisos, un 30% tiene un perfil progresista. De estos, un 10% votaron A en las últimas elecciones, y un 60% votaron B (el resto no votaron o votaron en blanco).

   (b) Del 50% de indecisos, un 70% tiene un perfil conservador. De estos, un 60% votaron A en las últimas elecciones, y un 10% votaron B (el resto no votaron o votaron en blanco).

   A partir de toda esta información, y suponiendo que los indecisos terminarán manteniendo un patrón de comportamiento similar al de las últimas elecciones, se puede calcular la **estimación de voto**.

INTERVALOS DE CONFIANZA

1. En 1778, Henry Cavendish realizó mediciones de la densidad terrestre (mediante un experimento con una balanza de torsión) obteniendo los siguientes resultados:

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	4.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

   a) Asumiendo Normalidad, obtener un intervalo de confianza de nivel 0.95 para la media.

   b) ¿Cuántas observaciones habría necesitado para estimar la media con un error inferior a 0.04, al nivel de confianza 0.95?

2. La alimentación en cierta comarca es rica en alimentos con alto contenido en purinas (compuestos con nitrógeno), y esto puede producir un mayor nivel de ácido úrico en sangre. Se mide el nivel de ácido úrico en la sangre de 10 individuos elegidos al azar en esa comarca, y se obtienen los siguientes valores (con sus correspondientes resúmenes):

   Valores obtenidos: 5,1 6,1 5,2 5,9 5,4 5,5 5,9 6,1 4,8 6,5

   $\sum x_i = 56,5 \qquad \sum x_i^2 = 321,79$

   (a) Halla la mediana y los cuartiles.

   (b) Asumiendo normalidad, halla un intervalo (con nivel de confianza 0,90) para el nivel medio de ácido úrico en sangre en esa comarca.

   (c) Halla un intervalo (con nivel de confianza 0,90) para la varianza del nivel de ácido úrico en esa comarca.

3. Las toxinas producidas por cianobacterias (microorganismos procariotas capaces de realizar fotosíntesis oxigénica) en los embalses han sido reconocidas como un problema de salud (Chorus & Bartram, 2002); por ello, la nueva legislación europea sobre aguas de baño incluye el seguimiento y control de estas bacterias en los embalses.

   (a) De 33 embalses muestreados en España, 17 presentan una gran abundancia de cianobacterias, frente a 16 que no presentan una cantidad preocupante. A partir de estos datos, dar un intervalo de confianza para estimar la proporción de embalses españoles con una gran abundancia de cianobacterias (con un nivel de confianza del 90%).

   (b) ¿Cuántos embalses habría que muestrear para que el error en la estimación de esa proporción quede por debajo de 0,04?

4. Un equipo de investigadores quiere estimar la proporción p de vacas que sufren el mal de las vacas locas en una gran explotación ganadera, mediante un intervalo con un error máximo de 0.015 y nivel de confianza 0.95. ¿A cuántas vacas deben analizar para alcanzar aproximadamente este objetivo, sabiendo que en un pequeño sondeo orientativo (muestra piloto) resultó que el 15% de las vacas estaban afectadas por la enfermedad?

5. El maíz es un alimento importante para los animales. De todas formas, este alimento carece de algunos aminoácidos que son esenciales. Un grupo de científicos desarrolló una nueva variedad que sí contenía niveles apreciables de dichos aminoácidos. Para comprobar la utilidad de esta nueva variedad para la alimentación animal se llevó a cabo el siguiente experimento: a un grupo de 20 pollos de 1 día se les suministró un pienso que contenía harina de maíz de la nueva variedad. A otro grupo de 20 pollos (grupo de control) se le alimentó con un pienso que sólo se diferenciaba del anterior en que no contenía harina de la variedad mejorada de maíz. Los resultados que se obtuvieron sobre las ganancias de peso de los pollos (en gramos) al cabo de 21 días de alimentación fueron los siguientes:

   - *Variedad normal*

     380 321 366 356 283 349 402 462 356 410 329 399 350 384 316 272 345 455 360 431

   - *Variedad mejorada*

     361 447 401 375 434 403 393 426 406 318 467 407 427 420 477 392 430 339 410 326

Asumiendo normalidad e igualdad de varianzas, obtener un intervalo de confianza (al 95%) para estimar la diferencia entre las ganancias medias de peso con las dos variedades de pienso.

6. Se desea estimar la proporción p de ánades en la población de un parque natural que presenta altos niveles de contaminación por metales pesados. Para ello se realiza un sondeo preliminar con 50 ejemplares, de los cuales 9 resultaron tener altos niveles de contaminación.
a) Construir un intervalo de confianza, de nivel 0.95, para p a partir de los resultados.
b) ¿Qué tamaño muestral debería utilizarse en un nuevo sondeo para estimar p con un error máximo del 2.5% y un nivel de confianza del 0.92?

7. Una empresa de metalurgia está interesada en la temperatura media que alcanza cierta máquina utilizada en el proceso de fabricación. Para su estimación se obtienen 6 mediciones en grados centígrados:

   41,60 41,84 42,34 41,95 41,86 42,18

   Asumiendo Normalidad, se pide:

   a) Obtener el intervalo de confianza al 95% para la temperatura media, suponiendo que $\sigma = 0,30$.

   b) Deducir el tamaño muestral necesario para conseguir un intervalo de confianza al 95% con una longitud menor o igual que 0,1 grados.

   c) Obtener el intervalo de confianza al 95% para la temperatura media, suponiendo que desconocemos el valor de σ.

8. Estamos interesados en la concentración, X, de particulas contaminantes (expresada en ppm) en situaciones de altas presiones atmosféricas. Medimos dicha concentración en 6 estaciones de la Red de Control de la Calidad del Aire, tras cinco días de altas presiones, encontrando una concentración media de 150 ppm, con una cuasidesviación típica de 16 ppm.

   (a) Asumiendo Normalidad, calcula un intervalo de confianza para la concentración media de partículas contaminantes en esas condiciones, con un nivel de confianza del 95%.

   (b) ¿En cuántas estaciones deberíamos medir dicha concentración, para estimar la concentración media con un error inferior a 5 ppm (con el mismo nivel de confianza)?

9. La envergadura (en cm) del cóndor de California (X) en su edad adulta es modelizada con una distribución Normal. En una muestra de 5 cóndores adultos se obtiene que $\sum x_i = 1350$ y $\sum x_i^2 = 365150$.

   (a) Obtener un intervalo de confianza para estimar la envergadura media de toda la población, con una confianza del 90%.

   (b) ¿Cuántos cóndores sería necesario observar para poder estimar la envergadura media con la misma confianza y un error inferior a 5 cm?

10. Se admite que el número de microorganismos en una muestra de 1 mm cúbico de agua de un río sigue una distribución de Poisson de parámetro λ. En 40 muestras se han detectado, en total, 833 microorganismos. Calcula un estimador puntual y un intervalo de confianza al 90% para λ.

11. Nueve personas participan en el estudio de un producto que intenta reducir el apetito (clorofenilpiperacina). Cada uno de ello recibe este producto durante 2 semanas y placebo durante otras 2 semanas (naturalmente, el orden de los períodos de 2 semanas es aleatorio y ellos no lo conocen). Al final de cada período, se les pide que expresen su sensación de hambre (en una escala del 0 al 150). Los resultados son los siguientes:

Individuo	1	2	3	4	5	6	7	8	9
Después del producto	79	48	52	15	61	107	77	54	5
Después del placebo	78	54	142	25	101	99	94	107	64

   (a) Hallar un intervalo de confianza al 95% para la diferencia de las sensaciones medias de hambre con el producto y con placebo (asumir Normalidad).

   (b) Lo mismo, pero trabajando (equivocadamente) con las muestras como si fueran independientes (asumir Normalidad e igualdad de varianzas).

12. En una población se está estudiando la proporción p de individuos alérgicos al polen de las acacias. En 200 individuos tomados al azar se observaron 8 alérgicos. Calcula un intervalo de confianza para p con un nivel de confianza de 0.95. ¿Cuántos individuos necesitaríamos observar para estimar esa proporción con un error inferior al 1%, al nivel de confianza 0.95?

13. La siguiente tabla proporciona los resultados de aplicar dos métodos diferentes a 5 lotes de pastillas para determinar su porcentaje de paracetamol (el método 1 consiste en un ensayo espectrométrico UV y el método 2 en espectroscopia de reflectancia en el infrarrojo cercano).

Lote	1	2	3	4	5
Método 1 (X)	84,63	84,38	84,08	84,41	83,82
Método 2 (Y)	83,15	83,72	83,84	84,20	83,92

Resúmenes numéricos para facilitar algunos cálculos:

$$\sum x_i = 421,32 \qquad \sum y_i = 418,83$$

$$\sum x_i^2 = 35502,51 \qquad \sum y_i^2 = 35084,31 \qquad \sum x_i y_i = 35292,04$$

(a) Determina la recta de regresión que expresa los porcentajes obtenidos por el Método 2 en función de los porcentajes obtenidos por el Método 1 (obteniendo previamente medias, varianzas y covarianza con 4 cifras decimales).

(b) Evalúa el ajuste proporcionado por la recta de regresión.

(c) Asumiendo Normalidad (e igualdad de varianzas si es preciso), calcula el intervalo de confianza para estimar la diferencia entre los resultados medios que se obtendrían con ambos métodos, con una confianza del 95%.

CONTRASTES DE HIPÓTESIS

1. El flamenco andino es una de las seis especies reconocidas de flamencos. Su tamaño medio es de 110 centímetros.

   Se detecta una nueva colonia de flamencos andinos en un humedal y se procede a su estudio. En una muestra de 9 flamencos se obtienen los siguientes tamaños:

   103 108 112 117 100 108 109 117 118

   (a) ¿Cuánto vale la mediana muestral? ¿Y los cuartiles?

   (b) Asumiendo Normalidad, ¿se puede aceptar que el tamaño medio de esta nueva colonia es también de 110 centímetros? Dar una respuesta al nivel de significación 0,05.

   Para facilitar los cálculos: $\sum x_i = 992 \qquad \sum x_i^2 = 109664$

   (c) En el apartado anterior, ¿el p-valor es mayor o menor que 0,05? Contestar razonadamente sin calcular el p-valor.

2. Se analiza un envío de botellas sobre las que se afirma que contienen 100 cl. de agua. Examinada una muestra de 5 botellas se obtiene que la media es de 95 cl. y la cuasivarianza muestral es $s^2 = 1.1$. Al nivel de significación 5%, ¿existe evidencia empírica para afirmar que la cantidad media de agua no es de 100 cl.?

3. Un sociólogo afirma que el 40% de los universitarios han viajado al extranjero al menos una vez. En una muestra de 100 universitarios, se observa que 36 han salido del país en alguna ocasión. Contrastar la hipótesis del sociólogo para un nivel de significación del 10%.

4. La concentración media de dióxido de carbono en el aire a cierta altura es habitualmente de unas 355 p.p.m. (partes por millón). Se sospecha que esta concentración es mayor en la capa de aire más próxima a la superficie. Para contrastar esta hipótesis se analizó el aire en 20 puntos elegidos aleatoriamente a una misma altura cerca del suelo. Resultó una media muestral de 580 p.p.m. y una cuasi-desviación típica muestral de 180. Suponiendo normalidad para las mediciones, ¿proporcionan estos datos suficiente evidencia estadística, al nivel 0.01, a favor de la hipótesis de que la concentración es mayor cerca del suelo? Indicar razonadamente si el p-valor es mayor o menor que 0.01.

5. Un fabricante de materiales para insonorización produce dos tipos A y B. De los 1000 primeros lotes vendidos, 560 fueron del tipo A. ¿Proporcionan estos datos suficiente evidencia estadística (al nivel de significación 0.01) para concluir que los consumidores prefieren mayoritariamente el tipo A?

6. Se tienen dos métodos, A y B, para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos (en calorías por gramo de masa para pasar de -0.72°C a 0°C) utilizando ambos métodos independientemente:

Método A	79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02
Método B	80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

   Asumiendo Normalidad e igualdad de varianzas, ¿existen diferencias significativas entre los resultados medios proporcionados por los dos métodos (a un nivel de significación del 10%)?

7. El maíz es un alimento importante para los animales. De todas formas, este alimento carece de algunos aminoácidos que son esenciales. Un grupo de científicos desarrolló una nueva variedad que sí contenía niveles apreciables de dichos aminoácidos. Para comprobar la utilidad de esta nueva variedad para la alimentación animal se llevó a cabo el siguiente experimento: a un grupo de 20 pollos de 1 día se les suministró un pienso que contenía harina de maíz de la nueva variedad. A otro grupo de 20 pollos (grupo de control) se le alimentó con un pienso que sólo se diferenciaba del anterior en que no contenía harina de la variedad mejorada de maíz. Los resultados que se obtuvieron sobre las ganancias de peso de los pollos (en gramos) al cabo de 21 días de alimentación fueron los siguientes:

- *Variedad normal*
  380 321 366 356 283 349 402 462 356 410 329 399 350 384 316 272 345 455 360 431
- *Variedad mejorada*
  361 447 401 375 434 403 393 426 406 318 467 407 427 420 477 392 430 339 410 326

a) Asumiendo Normalidad e igualdad de varianzas, ¿se puede considerar que hay suficiente evidencia estadística para afirmar que la ganancia media de peso es mayor con la variedad mejorada? Dar una respuesta con un nivel de significación 0,10.

b) ¿Era razonable aceptar la hipótesis de igualdad de varianzas? Dar una respuesta con un nivel de significación 0,10.

8. Se desea comparar la proporción de viviendas con calefacción en Extremadura y en Galicia. Se hace un muestreo en las dos comunidades con los siguientes resultados:

   Extremadura: De 500 viviendas elegidas al azar, 300 disponen de calefacción.

   Galicia: De 1000 viviendas elegidas al azar, 680 disponen de calefacción.

   ¿Hay suficiente evidencia estadística para concluir, con un nivel de significación del 5%, que es menor la proporción de viviendas con calefacción en Extremadura que en Galicia?

9. 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. Parece que la altura en Tanzania es mayor que la altura en 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:

   (a) Con un nivel de significación de 0.10, ¿podemos aceptar igualdad de varianzas de las alturas en las dos colonias?

   (b) ¿Disponemos de suficiente evidencia muestral para asegurar que la altura media en Tanzania es mayor que en Kenia (al nivel de significación 0.10)?

10. Un grupo de 23 voluntarios contrajo un resfriado tras haber sido expuesto a un virus. Diez de los voluntarios tomaron durante varios días un comprimido que contenía 1 g de vitamina C mientras que los trece restantes tomaron un placebo. Para todos ellos se observó el tiempo en días que duró el resfriado. Un resumen de los resultados aparece en la siguiente tabla:

	Vitamina C	Placebo
Media	$6,45$	7,13
Cuasidesviación típica	$0,76$	$0,88$

   (a) Suponiendo normalidad e igualdad de varianzas, ¿permiten estos datos afirmar a nivel $\alpha = 0,05$ que el tratamiento con vitamina C es eficaz para reducir la duración del resfriado? Determina razonadamente si el p-valor del contraste es mayor o menor que $0,05$.
   (b) Contrasta si la hipótesis de igualdad de varianzas es aceptable a nivel $\alpha = 0,05$.

11. Nueve personas participan en el estudio de un producto que intenta reducir el apetito (clorofenilpiperacina). Cada uno de ello recibe este producto durante 2 semanas y placebo durante otras 2 semanas (naturalmente, el orden de los períodos de 2 semanas es aleatorio y ellos no lo conocen). Al final de cada período, se les pide que expresen su sensación de hambre (en una escala del 0 al 150). Los resultados son los siguientes:

Individuo	1	2	3	4	5	6	7	8	9
Después del producto	79	48	52	15	61	107	77	54	5
Después del placebo	78	54	142	25	101	99	94	107	64

(a) Asumiendo Normalidad, ¿podemos afirmar, al nivel de significación del 5%, que la sensación media de hambre con el producto es menor que con el placebo?

(b) El p-valor, ¿es mayor o menor que 0,05?

12. Con el objeto de estudiar la efectividad de un agente diurético, se eligieron al azar 11 pacientes, aplicando a 6 de ellos dicho fármaco y un placebo a los restantes. La variable observada en esta experiencia fue la concentración de sodio en la orina a las 24 horas, la cual dio los resultados siguientes:

Diurético	20.4	62.5	61.3	44.2	11.1	23.7
Placebo	1.2	6.9	38.7	20.4	17.2	

Se supone que las concentraciones de sodio, en ambos casos, tienen una distribución $\mathcal{N}(\mu_1, \sigma)$ y $\mathcal{N}(\mu_2, \sigma)$ respectivamente. Contrasta, a un nivel de significación del 5%, si existe diferencia en el efecto medio al usar el agente diurético.

13. 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.

(a) ¿Permiten estos datos concluir (al nivel de significación del 5%) que la ingesta media de zinc en la población de esa edad es inferior a la recomendada?

(b) Calcula un intervalo de confianza de nivel 95% para la varianza poblacional de la ingesta diaria de zinc.

14. Se lleva a cabo un estudio morfométrico de cráneos de lobos de las Montañas Rocosas y de lobos árticos. Concretamente, medimos la anchura de la caja craneal de 16 lobos árticos y de 9 lobos de las Rocosas, y los resultados muestrales (expresados en centímetros) aparecen en la siguiente tabla:

	Número de datos	Media	Cuasidesviación típica
Lobos árticos	16	39,98	2,6883
Lobos de las Rocosas	9	42,61	2,1310

Asumiendo Normalidad e igualdad de varianzas cuando sea necesario:

(a) Calcula un intervalo de confianza para la media de la anchura de la caja craneal de los lobos de las Montañas Rocosas (con nivel de confianza 0,95).

(b) ¿Proporcionan estos datos suficiente evidencia estadística (al nivel de significación $0,05$) para concluir que los lobos de las Montañas Rocosas tienen una mayor anchura craneal que los lobos árticos, por término medio?

15. Se están comparando los jabalíes de Soria con los de Cáceres. Los de Soria parecen un poco más grandes. Se estudia una muestra de 16 jabalíes de Soria, y da un peso medio muestral de 98 Kg y una cuasi-varianza muestral de 130. También se estudia una muestra de 9 jabalíes de Cáceres, obteniéndose un peso medio muestral de 91 Kg y una cuasi-varianza muestral de 90. Asumimos normalidad para los pesos en las dos poblaciones de jabalíes.

(a) Con un nivel de significación de $0'10$, ¿podemos aceptar igualdad de varianzas de los pesos en las dos poblaciones?

(b) ¿Disponemos de suficiente evidencia muestral para asegurar (al nivel de significación $0'10$) que el peso medio en Soria es mayor que en Cáceres?

16. La existencia de trazas de metales en el agua afecta a su sabor y, si las concentraciones son altas, puede afectar a la salud. En un estudio se seleccionaron seis localizaciones en un río y, para cada localización, se determinó la concentración de zinc en el agua de la superficie y en el agua del fondo (en mg/l). Los resultados fueron los siguientes:

Localización	1	2	3	4	5	6
Concentración en el fondo (y)	0,43	0,27	0,57	0,53	0,71	0,72
Concentración en la superficie (x)	0,41	0,24	0,39	0,41	0,60	0,61

Con el fin de facilitar los cálculos, se recomienda trabajar con los siguientes resúmenes

$\sum_{i=1}^{n} x_i = 2,66$ $\sum_{i=1}^{n} x_i^2 = 1,278$ $\sum_{i=1}^{n} y_i = 3,23$ $\sum_{i=1}^{n} y_i^2 = 1,886$ $\sum_{i=1}^{n} x_i y_i = 1,546$

y se recomienda mantener tres cifras decimales en los cálculos, para evitar problemas con los redondeos.

(a) Calcula la recta de regresión que resulta útil para predecir la concentración de zinc en el fondo a partir de la concentración en la superficie.

(b) Evalúa el grado de ajuste de la recta a los datos. Si la concentración de zinc en la superficie es de 0,50 mg/l, ¿cuál sería la concentración de zinc predicha en el fondo en esa misma localización?

(c) Asumiendo Normalidad, ¿existe evidencia empírica para afirmar, con un nivel de significación $\alpha = 0,05$, que la concentración media de zinc en el fondo es diferente a la concentración media en la superficie?

17. El contenido medio habitual de arsénico en un Parque Nacional es de 9 p.p.m. Se cree que últimamente este contenido medio ha podido aumentar. Para estudiar esta posible contaminación, se toma una muestra del contenido de arsénico en 20 puntos diferentes del Parque, resultando una media muestral de 10 p.p.m. , y una cuasi-varianza muestral de 2,1.

¿Los datos son lo suficientemente concluyentes como para poder afirmar que, efectivamente, ha habido una contaminación, es decir, que el contenido medio en la actualidad es superior al habitual de 9 p.p.m.? Dar una respuesta, al nivel de significación 0,05, asumiendo Normalidad en los datos.

¿El p-valor de estos datos es superior o inferior a 0,05? Dar una respuesta razonada sin hacer cálculos adicionales.

18. El hematocrito es el porcentaje del volumen total de la sangre compuesto por glóbulos rojos y depende tanto del número de glóbulos rojos como de su tamaño. En una muestra de 21 hombres se obtuvo un valor medio de hematocrito de 48.8 y una cuasidesviación típica de 2.8. En otra muestra de 21 mujeres se obtuvo un valor medio de hematocrito de 40.6 y una cuasidesviación típica de 2.9. Asumiendo normalidad:

(a) Calcula un intervalo de confianza de nivel 95% para la varianza del nivel de hematocrito en la población de hombres.

(b) ¿Podemos aceptar igualdad de varianzas en la población de hombres y de mujeres (a nivel 0.10)?

(c) ¿Permiten los datos afirmar (a nivel 0.05) que el nivel medio de hematocrito en la población de hombres es superior al nivel medio en la población de mujeres? Determina razonadamente si el p-valor del contraste es mayor o menor que 0.05.

19. Se desea estudiar la efectividad de un insecticida ecológico contra los áfidos en la cosecha de patatas.

Por un lado, se tratan 100 plantas con el insecticida y se comprueba que 10 de ellas tienen áfidos.

Por otro lado, se tiene un grupo de control de otras 100 plantas diferentes que no reciben tratamiento y se comprueba que 15 de ellas tienen áfidos.

¿Podemos concluir, al nivel de significación 0,05, que la proporción de plantas con áfidos es menor cuando son tratadas con el insecticida?

20. La producción de trigo (en Tm/Ha) que se obtiene un año en 5 parcelas es la siguiente:

11,04 15,13 9,04 20,60 31,26

Se quiere estudiar la efectividad de un nuevo fertilizante. Para esto, se observa la producción de trigo que se obtiene al año siguiente, usando el nuevo fertilizante. Los resultados en las mismas 5 parcelas son los siguientes:

12,08 17,28 10,82 18,90 32,03

¿Podemos afirmar que el nuevo fertilizante aumenta la producción media? Responder razonadamente, al nivel de significación 0,01.

21. El nivel ceráunico de un lugar es el número de días al cabo del año en los que hay tormenta (se considera día con tormenta a aquel en el que al menos se oye un trueno). En cierta comarca, se conoce por datos históricos que el nivel ceráunico sigue una distribución de Poisson con una media (λ) de 20 días.

    Sin embargo, se piensa que últimamente su nivel ceráunico ha aumentado, como consecuencia del cambio climático. En un seguimiento de los últimos 50 años, se ha obtenido que el número medio de días al año con tormenta ha sido de 22. ¿Se puede afirmar que efectivamente se ha producido un aumento del nivel ceráunico? Dar una respuesta razonada, al nivel de significación del 5%.

BONDAD DE AJUSTE

1. Después de lanzar un dado 300 veces, se han obtenido las siguientes frecuencias:

	1	2	3	4	5	6
Frecuencias	43	49	56	45	66	41

Al nivel de significación 0,05, ¿se puede aceptar que el dado es regular?

2. En 1778, H. Cavendish realizó una serie de 29 experimentos con objeto de medir la densidad de la tierra. Sus resultados, tomando como unidad la densidad del agua, fueron:

5'50	5'61	4'88	5'07	5'26	5'55	5'36	5'29	5'58	5'65
5'57	5'53	5'62	5'29	5'44	5'34	5'79	5'10	5'27	5'39
5'42	5'47	5'63	5'34	5'46	5'30	5'75	5'68	5'85	

Al nivel de significación 0.05, ¿se puede aceptar que la densidad de la tierra se ajusta a una distribución Normal?

Observación:

Utilizar los intervalos A_1="Menor o igual que 5,30", A_2=(5,30; 5,45] , A_3=(5,45; 5,60] , A_4="Mayor que 5,60", y el hecho de que, a partir de los datos, se obtiene que $\hat{\mu} = 5,45$ y que $\hat{\sigma} = 0,22$.

3. Nos dicen que un programa de ordenador genera observaciones de una distribución $N(0;1)$. Como no estamos seguros de ello, obtenemos una muestra aleatoria de 450 observaciones, mediante dicho programa, obteniéndose los siguientes resultados:

30 observaciones menores que -2;

80 observaciones entre -2 y -1;

140 observaciones entre -1 y 0;

110 observaciones entre 0 y 1;

60 observaciones entre 1 y 2;

30 observaciones mayores que 2.

¿Se puede aceptar, al nivel $\alpha = 0,01$, que el programa funciona correctamente?

4. La tabla que aparece a continuación, muestra la frecuencia de la cifra final del *gordo* de 200 sorteos de la Lotería de Navidad:

Cifra final	0	1	2	3	4	5	6	7	8	9
Frecuencia	20	8	13	20	27	30	26	20	20	16

¿Se puede aceptar, al nivel de significación del 1%, que todas las terminaciones son igualmente probables?

5. Se desea estudiar el número de erratas por página que se producen en la edición de una enciclopedia, antes de su lanzamiento. Para ello se analizan 200 páginas seleccionadas al azar, encontrando los siguientes resultados:

Número de erratas por página	0	1	2	3
Número de páginas	150	42	5	3

(a) Asumiendo que el "número de erratas por página" sigue un modelo de Poisson, calcula un intervalo (al 90% de confianza) para el número medio de erratas por página en toda la enciclopedia.

(b) Con los datos obtenidos, ¿es realmente aceptable, al nivel de significación del 10%, que el "número de erratas por página" sigue una distribución de Poisson?

6. Un modelo genético para las moscas de cierta variedad nos dice que pueden ser de tres tipos: homocigóticas AA (con probabilidad p^2), homocigóticas BB (con probabilidad q^2) y heterocigóticas AB (con probabilidad $2pq$), donde naturalmente $p + q = 1$.

   En una muestra aleatoria de 100 moscas obtenemos 10 de tipo AA, 50 de tipo BB, y 40 de tipo AB.

   ¿Se ajustan los datos a dicho modelo genético, al nivel de significación 0,05?

7. Se clasificaron 1000 individuos de una población según el sexo y según fueran normales o daltónicos.

	Masculino	Femenino
Normal	442	514
Daltónicos	38	6

   Según un modelo genético, las probabilidades deberían ser:

$$\begin{array}{cc} \frac{1}{2}\,p & \frac{1}{2}\,p^2 + pq \\ \\ \frac{1}{2}\,q & \frac{1}{2}\,q^2 \end{array}$$

   donde $q = 1 - p =$ proporción de genes defectuosos en la población.

   A partir de la muestra se ha estimado que $q = 0,087$. ¿Concuerdan los datos con el modelo, al nivel de significación 0,05?

8. Un Ayuntamiento decide poner 4 contenedores para reciclar papel en una zona de la ciudad, con la idea de que sean utilizados por la misma cantidad de personas (aproximadamente). Para ver si esto es cierto, hace una encuesta en la zona a 300 personas, preguntándoles que contenedor utilizan. Los resultados obtenidos son los siguientes:

   El contenedor 1 es utilizado por 80 personas.

   El contenedor 2 es utilizado por 70 personas.

   El contenedor 3 es utilizado por 85 personas.

   El contenedor 4 es utilizado por 65 personas.

   a) Como consecuencia de estos resultados, ¿resulta aceptable que los 4 contenedores tienen el mismo nivel de utilización? Dar una respuesta razonada, con un nivel de significación de 0.10.

   b) El p-valor del contraste anterior, ¿es inferior o superior a 0.10? Dar una respuesta razonada.