BASES DE ESTADÍSTICA
Segundo curso de Ciencias Ambientales / Curso 2006-2007

ESTADÍSTICA DESCRIPTIVA

1. 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.

d) ¿Es la distribución aproximadamente normal?

2. 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. ¿Qué conclusiones se pueden obtener acerca de la dispersión y la forma de la distribución de los datos?

3. 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?

4. 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?

5. 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?

6. 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?

b) Dibuja un diagrama de dispersión con esos datos. ¿Qué nos dice el diagrama sobre la relación entre esas dos variables?

c) Las variables ¿están asociadas positiva o negativamente?

d) Describe la forma de la relación ¿Es lineal?

e) 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

7. 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) Quieres predecir el número de pasos por segundo a partir de la velocidad. Para ello, dibuja un diagrama de dispersión.

b) Describe la relación existente y halla la correlación.

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

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) $X \sim N(\mu, \sigma)$, (σ conocido).

   e) $X \sim N(\mu, \sigma)$, (μ conocido).

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

2. A pesar de que la mayoría de las encuestas se llevan a cabo de forma anónima, los encuestados pueden presentar reservas al contestar a ciertas preguntas comprometidas. Con el fin de evitar sesgos, las instrucciones para rellenar una encuesta sobre evasión fiscal son las siguientes:

   > Tire un dado.
   >
   > Si el resultado es 1 ó 2, conteste A en el caso de que usted haya cometido voluntariamente irregularidades en su declaración de la renta y B si no las ha cometido.
   >
   > Si el resultado es 3, 4, 5 ó 6, conteste A si *no* ha cometido irregularidades y B si *sí* las ha cometido.

   De 100 encuestados, 64 contestaron A. ¿Qué estimación puede darse del porcentaje de contribuyentes que ha cometido irregularidades voluntariamente?

3. 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.

4. Un test para detectar si el agua presenta cierto tipo de contaminación resulta positivo con probabilidad 0.99 si el agua está realmente contaminada (*sensibilidad del test*). Si el agua no está contaminada, resulta negativo con probabilidad 0.97 (*especificidad del test*). La sensibilidad y la especificidad se conocen debido a que se tiene mucha experiencia en el uso de la prueba.
   (a) ¿Qué relación existe entre la probabilidad de que el test dé positivo y la de que el agua esté contaminada?
   (b) Se aplica el test a muestras de agua de 15 lagos y resulta positivo en 2 de las muestras. Utiliza la relación del apartado (a) para estimar el porcentaje de lagos contaminados.

5. El coseno X del ángulo con el que se emiten los electrones en un proceso radiactivo es una variable aleatoria con función de densidad

   $$f_\theta(x) = \frac{1 + \theta x}{2}, \text{ si } -1 \leq x \leq 1,$$

   donde $\theta \in [-1, 1]$. Se dispone de una muestra aleatoria simple $X_1, X_2, \ldots, X_n$.
   (a) Calcula un estimador de θ por el método de los momentos.
   (b) Estudia si el estimador obtenido es insesgado.

6. En una gran piscifactoría hay una proporción desconocida de peces de una especie A. Para obtener información sobre esa proporción vamos a ir sacando peces al azar.

   a) Si la proporción de peces de la especie A es p, ¿cuál es la probabilidad de que el primer pez de la especie A sea el décimo que extraemos?

b) Tres personas realizan, independientemente unas de otras, el proceso de sacar peces al azar hasta encontrarse con el primero de tipo A:

La primera persona obtiene el primer pez tipo A en la décima extracción.

La segunda persona obtiene el primer pez tipo A en la decimoquinta extracción.

La tercera persona obtiene el primer pez tipo A en la decimoctava extracción.

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

7. 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, por máxima verosimilitud, la proporción de pollos enfermos en la granja, explicando todo el proceso seguido.

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) Obtener estimaciones insesgadas para la media y la varianza de la densidad.

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

2. 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.

3. En una explotación minera, las rocas excavadas se someten a un análisis químico para determinar su contenido porcentual de cadmio. Se puede suponer que este contenido es una variable con distribución normal de media μ y varianza σ^2. Después de analizar 25 rocas se obtiene un contenido porcentual medio de 9.77 con una cuasidesviación típica de 3.164.
(a) Construye un intervalo de confianza de nivel 95% para el contenido porcentual medio de cadmio en la mina.
(b) Construye un intervalo de confianza de nivel 95% para σ.

4. 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.

5. 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?

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. En un experimento genético se ha medido la longitud del ala de 20 moscas, resultando los valores: 93, 90, 97, 90, 93, 91, 96, 94, 91, 91, 88, 93, 95, 91, 89, 92, 87, 88, 90, 86. Suponiendo que la longitud del ala de las moscas es una variable aleatoria con distribución $\mathcal{N}(\mu, \sigma^2)$, calcula un intervalo de confianza de nivel 0.90 para μ y otro para σ^2.

8. En una población, el consumo de agua anual por individuo es una variable aleatoria con distribución normal de media μ y desviación típica $\sigma = 7.5$. ¿Qué tamaño muestral hace falta para obtener un intervalo de confianza para μ con un margen de error ± 2 y un nivel de confianza de 0.955.

9. La concentración media habitual, X, de un determinado contaminante en la atmósfera es de 8 p.p.m. (partes por millón). Un día se hacen mediciones en 10 puntos de una ciudad (alejados unos de otros), obteniéndose los resultados que se resumen a continuación:

$$\sum_{i=1}^{10} x_i = 85.2 \qquad \sum_{i=1}^{10} x_i^2 = 773.82$$

Asumiendo Normalidad, calcula un intervalo de confianza de nivel 0.95 para la concentración media del contaminante. ¿Qué conclusiones se pueden sacar?

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. Se intenta estudiar la influencia de la hipertensión en los padres sobre la presión sanguínea de los hijos. Para ello se seleccionan dos grupos de niños, unos con padres de presión sanguínea normal (grupo 1) y otros con uno de sus padres hipertenso (grupo 2), obteniéndose las siguientes presiones sistólicas:

Grupo 1	104	88	100	98	102	92	96	100	96	96
Grupo 2	100	102	96	106	110	110	120	112	112	90

Halla un intervalo de confianza para la diferencia de medias, suponiendo que las varianzas en las dos poblaciones de niños son iguales.

12. a) En un estudio para el estado de la salud bucal de una ciudad, se tomó una muestra elegida al azar de 280 varones entre 35 y 44 años, y se contó el número de piezas dentarias en boca. Tras la revisión pertinente, los dentistas informaron que había 70 individuos con 28 o más dientes. Realizar una estimación por intervalos de confianza de la proporción de individuos de esta ciudad con 28 dientes o más, con un nivel de confianza 0.95.

    b) Utilizando los datos anteriores como muestra piloto, calcular cuál sería el tamaño muestral necesario para efectuar la anterior estimación con un error inferior a 0.01 (con el mismo nivel de confianza).

CONTRASTES DE HIPÓTESIS

1. La concentración media habitual, X, de un determinado contaminante en la atmósfera es de 8 p.p.m. (partes por millón). Un día se hacen mediciones en 10 puntos de una ciudad (alejados unos de otros), obteniéndose los resultados que se resumen a continuación:

$$\sum_{i=1}^{10} x_i = 85.2 \qquad \sum_{i=1}^{10} x_i^2 = 773.82$$

   Asumiendo Normalidad, ¿resulta aceptable, al nivel de significación $\alpha = 0.05$, que el grado de contaminación medio es el habitual?

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 en una cierta zona no es habitualmente mayor que 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

   Con un nivel de significación del 10%, ¿existen diferencias significativas entre los resultados medios proporcionados por los dos métodos?

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 van a probar dos medicamentos A y B contra una enfermedad. Para esto tratamos 100 ratones enfermos con A y otros 100 con B. El número medio de horas que sobreviven con A es $\bar{x} = 1200$ y el número medio con B es $\bar{y} = 1400$. Suponiendo normalidad en ambos casos y sabiendo que:

$$\sum (x_i - \bar{x})^2 = 900000 \qquad \sum (y_i - \bar{y})^2 = 950000$$

   a) ¿Se puede aceptar igualdad de varianzas con $\alpha = 0.10$?
   b) ¿Es más efectivo el medicamento B? Plantear el contraste adecuado para estudiar esto con un nivel de significación del 5%.

10. Con objeto de estudiar si las pulsaciones en los hombres (X) pueden considerarse menores que en las mujeres (Y), se toman muestras de 16 hombres y 16 mujeres, obteniéndose los siguientes resultados:

$$\sum x_i = 1248 \qquad \sum x_i^2 = 97570 \qquad \sum y_i = 1288 \qquad \sum y_i^2 = 103846$$

    ¿Qué se puede decir al respecto?

11. Una compañía americana de distribución de gasolina quiere probar el rendimiento de un nuevo combustible. Para esto hace pruebas con 8 modelos de coches diferentes en una autopista. El número de millas recorridas por galón de gasolina con cada coche con el antiguo combustible y con el nuevo combustible se da a continuación:

Coche	1	2	3	4	5	6	7	8
Antiguo combustible	58	52	50	50	50	49	44	42
Nuevo combustible	60	55	52	51	48	50	42	46

    Asumiendo Normalidad, ¿se puede concluir que el nuevo combustible proporciona un mejor rendimiento? Dar una respuesta con un nivel de significación de 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 desea estudiar si la utilización de tratamientos para reducir el nivel de colesterol reduce también el riesgo de sufrir infartos. Para ello, 2.051 hombres de mediana edad recibieron un tratamiento para reducir el nivel de colesterol a base de *gemfibrozil* mientras que un grupo de control de 2.030 hombres recibió un placebo. Durante los cinco años siguientes, 56 hombres del grupo de *gemfibrozil* y 84 del grupo del placebo sufrieron infartos.
(a) ¿Existe evidencia empírica, al nivel $\alpha = 0,05$ de que el *gemfibrozil* es eficaz para reducir el riesgo de sufrir infartos?
(b)Determina razonadamente si el p-valor del contraste es mayor o menor que 0,05.

14. En una piscifactoría se desea contrastar la hipótesis (H_0) de que el porcentaje de peces adultos que miden menos de 20 cm. es como máximo el 10%. Se va a tomar una muestra de 6 peces y rechazaremos H_0 si encontramos más de un pez con longitud inferior a 20 cm. ¿Cuál es el nivel de significación de este contraste?

15. Algunos estudios parecen sugerir que tomar cada día una aspirina puede tener efectos beneficiosos para la salud. En un amplio estudio sobre 22071 personas sanas de más de 40 años, se administró una aspirina diaria durante cierto periodo de tiempo a 11037 personas mientras que el resto recibió un placebo. En la siguiente tabla aparece el número de personas que sufrieron infartos y embolias para cada grupo:

	Grupo aspirina	Grupo placebo
Infartos	139	239
Embolias	98	119

(a) ¿Aportan estos datos suficiente evidencia empírica al nivel $\alpha = 0,05$ de que tomar una aspirina es beneficioso para evitar sufrir embolias?

(b) Calcula un intervalo de confianza de nivel 0,95 para la la diferencia de proporciones de individuos que sufren un infarto entre los que han tomado y no han tomado aspirinas.

16. Una compañía petrolífera está considerando la posibilidad de introducir un nuevo aditivo en su gasolina, esperando incrementar el Kilometraje medio por litro. Los ingenieros del grupo de investigación prueban 10 coches con la gasolina habitual y otros 10 coches con la gasolina con el nuevo aditivo. El resumen de los resultados es:
"Kilometraje medio sin aditivo"= 14'2 Km/l $\quad s_1^2 = 3'24$
"Kilometraje medio con aditivo"= 15'4 Km/l $\quad s_2^2 = 5'76$
a) ¿Se puede considerar probado que el nuevo aditivo aumenta el kilometraje medio por litro? Plantear el modelo correspondiente (asumiendo Normalidad e igualdad de varianzas) y obtener una conclusión con una confianza del 95%.
b) Con los datos disponibles, ¿era razonable trabajar con la hipótesis de igualdad de varianzas en el apartado anterior? Dar una respuesta razonada con un nivel de significación de 0'10.

17. 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

Contrastar si los dos métodos proporcionan, en media, las mismas estimaciones (tomar un nivel de confianza del 90%).

18. Un fisioterapeuta afirma que con un nuevo procedimiento de rehabilitación que él aplica, determinada lesión tiene un tiempo medio de recuperación de 15 días. Se seleccionan al azar 38 personas que sufren dicho tipo de lesión para verificar su afirmación, y se obtiene un tiempo medio de recuperación de 17 días y una cuasivarianza de 9.

    Asumiendo normalidad para el tiempo de recuperación, ¿resulta aceptable la afirmación del fisioterapeuta? Dar una respuesta razonada, con un nivel de significación del 5%.

19. Se desea estudiar la efectividad de un insecticida ecológico contra los áfidos en la cosecha de patatas. Para esto, por un lado se tratan 40 plantas con el insecticida, obteniéndose una media de 15 áfidos por planta con una cuasi-varianza de 9,1. Por otro lado, tenemos un grupo de control de 30 plantas que no reciben tratamiento, obteniéndose una media de 18,3 áfidos por planta con una cuasi-varianza de 10,5.

    a) Asumiendo Normalidad e igualdad de varianzas, ¿podemos concluir que el insecticida es eficaz? Es decir, ¿podemos concluir que el número medio de áfidos por planta es menor cuando se utiliza el insecticida ecológico? Dar una respuesta razonada con un nivel de significación de 0,10.

    b) ¿Era aceptable asumir la igualdad de varianzas? Dar una respuesta razonada con un nivel de significación de 0,10.

20. En otro lugar, se lleva a cabo otro estudio diferente con el insecticida ecológico del ejercicio anterior. En este nuevo estudio, queremos comparar proporciones de plantas afectadas.

    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?

CONTRASTES χ^2

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 afirmar que el dado es regular?

2. 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?

3. 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?

4. Para estudiar si el grupo sanguíneo de los individuos tiene relación con la predisposición a padecer diabetes, se han seleccionado al azar 400 sujetos de los que se ha determinado el grupo sanguíneo y el nivel de glucosa en sangre en idénticas condiciones experimentales. Clasificada la segunda medida en niveles bajo, medio y alto, los resultados han sido:

	Bajo	Medio	Alto	Total
0	137	86	35	258
A	42	23	11	76
B	19	17	7	43
AB	14	7	2	23
Total	212	133	55	400

   Compruebe si existe independencia entre el grupo sanguíneo y el nivel de glucosa con un nivel de significación del 5%.

5. Para estudiar el número de ejemplares de cierta especie en peligro de extinción que viven en un bosque, se divide el mapa del bosque en nueve zonas y se cuenta el número de ejemplares de cada zona. Se observa que 60 ejemplares viven en el bosque repartidos en las 9 zonas de la siguiente forma:

8	7	3
5	9	11
6	4	7

   Mediante un contraste de hipótesis, analiza si estos datos aportan evidencia empírica de que los animales tienen tendencia a ocupar unas zonas del bosque más que otras.

6. Se desea estudiar el número de accidentes por día que se producen en cierto regimiento. Para ello se toman al azar los partes de 200 días dentro de los últimos 5 años, encontrando los siguientes resultados:

Número de accidentes/día	0	1	2	3	4	5	6
Número de días	58	75	44	18	3	1	1

a) ¿Se puede aceptar, con nivel de confianza del 90%, que el número de accidentes por día sigue una distribución de Poisson?

b) Independientemente del resultado de a), suponemos que la distribución del número de accidentes por día es Poisson (λ). ¿Hay suficiente evidencia estadística (tomar nivel de significación $\alpha = 0'05$) de que el verdadero valor medio λ del número de accidentes por día es menor que 1,35? Dada la aceptación o el rechazo en el test usado, ¿el $p-$valor es mayor o es menor que 0,05?

7. Se desea evaluar la efectividad de una nueva vacuna antigripal. Para ello se decide suministrar dicha vacuna, de manera voluntaria y gratuita, a una pequeña comunidad. La vacuna se administra en dos dosis, separadas por un período de dos semanas, de forma que algunas personas han recibido una sola dosis, otras han recibido las dos y otras personas no han recibido ninguna. La siguiente tabla indica los resultados que se registraron durante la siguiente primavera en 1000 habitantes de la comunidad elegidos al azar.

	No vacunados	Una dosis	Dos dosis
Gripe	24	9	13
No gripe	289	100	565

¿Proporcionan estos datos suficiente evidencia estadística (al nivel de significación 0,05) para indicar una dependencia entre la clasificación respecto a la vacuna y la protección frente a la gripe?

8. Se quiere comparar la biodiversidad de dos montes cercanos. Para esto hacemos lo siguiente:

En uno de los montes se eligen al azar 50 zonas, de 4 m^2 cada una, y se hace el recuento del número de especies vegetales diferentes que hay en cada una, con los siguientes resultados:

En 20 zonas había menos de 6 especies diferentes.

En 17 zonas había entre 6 y 8 especies diferentes.

En 13 zonas había más de 8 especies diferentes.

En el otro monte se se hace el mismo recuento en otras 40 zonas, obteniéndose los siguientes resultados:

En 12 zonas había menos de 6 especies diferentes.

En 20 zonas había entre 6 y 8 especies diferentes.

En 8 zonas había más de 8 especies diferentes.

¿Son similares los dos montes en lo que se refiere a su biodiversidad? Hacer el contraste correspondiente con un nivel de significación del 0,10.

9. 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?

10. Se está estudiando la distribución de los grupos sanguíneos O, A, B, AB en dos comunidades. Los resultados obtenidos fueron

	O	A	B	AB
Comunidad 1	121	120	79	33
Comunidad 2	118	95	121	30

a) ¿Se puede considerar que son homogéneas ambas comunidades?

b) Consideremos ahora sólo los datos de la comunidad 1. El modelo teórico asigna las siguientes probabilidades a cada uno de los grupos:

$$\begin{array}{cccc} O & A & B & AB \\ r^2 & p^2+2pr & q^2+2qr & 2pq \end{array} \qquad (p+q+r=1)$$

A partir de los datos de la muestra se han obtenido las siguientes estimaciones de los parámetros: $\hat{p} = 0,2465$ y $\hat{q} = 0,1732$. Obtener las frecuencias esperadas según el modelo teórico y contrastar la hipótesis de que los datos se ajustan a él.

11. Hemos desarrollado un modelo teórico para las diferentes clases de una variedad de moscas. Este modelo nos dice que la mosca puede ser de tipo L con probabilidad p^2, de tipo M con probabilidad q^2 y de tipo N con probabilidad $2pq$ $(p+q=1)$. Para confirmar el modelo experimentalmente tomamos una muestra de 100 moscas, obteniendo 10, 50 y 40, respectivamente.

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

b) ¿Se ajustan los datos al modelo teórico, al nivel de significación 0'05?

12. 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.

13. Se han clasificado los alumnos de un curso según su grupo sanguíneo y su Rh, obteniendo la siguiente tabla:

Rh/ Grupo	A	B	AB	0
+	48	32	12	40
-	16	6	5	12

a) ¿Es independiente el Rh del grupo sanguíneo? Dar una respuesta razonada al nivel de significación 0,05.

b) El p-valor de los datos ¿es inferior o superior a 0,05? Dar una respuesta razonada.