

**PROBLEMAS DE ESTADÍSTICA
CIENCIAS BIOLÓGICAS. CURSO 2000/01**

1. Análisis descriptivo con SPSS

1. Para evaluar la viabilidad de un proyecto de reforestación de una zona sometidas a stress turístico, para el que se ha solicitado una subvención pública, 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

Construir la tabla de frecuencias y el histograma correspondiente. ¿Qué conclusiones se pueden obtener acerca de la dispersión y la forma de la distribución de los datos?

2. 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 2 zonas diferentes del acuífero y se obtienen los siguientes resultados (número de colonias por 1000mm 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. Estudiar la dispersión, la asimetría y el apuntamiento de ambas muestras. ¿Se puede considerar que ambas zonas son semejantes?

3. En un estudio sobre el efecto de un contaminante procedente de una planta de transformación de aluminio se toma una muestra del número de peces muertos durante 15 días, obteniéndose los siguientes resultados:

14 17 23 15 16 18 12 16 26 30 13 22 17 15 24

Realizar un análisis descriptivo de la muestra obtenida.

4. Un investigador está estudiando si existe una relación entre el número de nacimientos de personas en una localidad y el número de cigüeñas observadas. Durante un año obtiene los siguientes datos cada mes:

nac.	36	27	19	45	41	33	34	41	31	25	45	44
cig.	83	64	46	106	97	78	79	96	71	61	105	105

¿Existe una relación lineal entre ambas variables?

5. En el Campeonato Mundial de Fútbol de 1998 participaron 32 selecciones nacionales: 5 de Africa, 8 de América, 4 de Asia y el resto de Europa. Representar la participación por continentes mediante un diagrama de rectángulos y un diagrama de sectores. Calcular el número medio de equipos por continente y la desviación típica.

6. En 1879 Michelson obtuvo los siguientes valores para la velocidad de la luz en el aire (se dan los resultados restando 299000 a los datos originales, en Km/seg., para facilitar su manejo): 850, 740, 900, 1070, 930, 850, 950, 980, 980, 880, 1000, 980, 930, 650, 760.

En 1882 Newcomb, utilizando otro procedimiento, obtuvo (restando de nuevo 299000): 883, 816, 778, 796, 682, 711, 611, 599, 1051, 781, 578, 796, 774, 820, 772. Se pide:

a) Diagramas de tallos y hojas para ambas distribuciones.

b) Medias y desviaciones típicas.

c) ¿Qué conclusiones pueden extraerse?

7. Los siguientes datos fueron utilizados por Lord Rayleigh (1894). En ellos aparecen los pesos del nitrógeno producido a partir de compuestos químicos o a partir del aire suprimiendo el oxígeno. Analizarlos mediante un diagrama de tallo y hojas:

<u>Fecha</u>	<u>Origen</u>	<u>Agente purificador</u>	<u>Peso del Nitrógeno</u>
29 – XI – 1893	NO	Acero caliente	2,30143
5 – XII – 1893	NO	Acero caliente	2,29816
6 – XII – 1893	NO	Acero caliente	2,30182
8 – XII – 1893	NO	Acero caliente	2,29890
12 – XII – 1893	Aire	Acero caliente	2,31017
14 – XII – 1893	Aire	Acero caliente	2,30986
19 – XII – 1893	Aire	Acero caliente	2,31010
22 – XII – 1893	Aire	Acero caliente	2,31001
26 – XII – 1893	N_2O	Acero caliente	2,29889
28 – XII – 1893	N_2O	Acero caliente	2,29940
9 – I – 1894	NH_4NO_2	Acero caliente	2,29849
13 – I – 1894	NH_4NO_2	Acero caliente	2,29889
27 – I – 1894	Aire	Hidrato de hierro	2,31024
30 – I – 1894	Aire	Hidrato de hierro	2,31030
1 – II – 1894	Aire	Hidrato de hierro	2,31028

8. Las puntuaciones de un test de inteligencia de 198 personas dieron los siguientes resultados:

Puntuación:	[30, 40],	(40, 50],	(50, 60],	(60, 70],	(70, 80],	(80, 90].
Nº de personas:	6	17	76	68	22	9

Analizar descriptivamente.

9. Se tienen los siguientes porcentajes de las tallas de los reclutas españoles del reemplazo de 1968:

Talla en cm.	menos de 150	(150, 155]	(155, 160]	(160, 165]
%	0,3	1,6	7,4	21,5
Talla en cm.	(165, 170]	(170, 175]	(175, 180]	más de 180
%	30,5	24,5	10,7	3,5

Analizar descriptivamente.

10. En un estudio de consumo a un modelo de automóvil, se realizan 10 recorridos diferentes, estudiándose las características x = “longitud de recorrido” e y = “consumo en ese recorrido”.

Los resultados del estudio se resumen así:

$$\sum_{i=1}^{10} x_i = 1287'1, \quad \sum_{i=1}^{10} x_i^2 = 165696'84, \quad \sum_{i=1}^{10} y_i = 116'5, \quad \sum_{i=1}^{10} y_i^2 = 1381'11,$$

$$\sum_{i=1}^{10} x_i y_i = 15022'889.$$

- a) Calcular la longitud y el consumo medios de los recorridos. Calcular las varianzas v_x y v_y y comentarlas.
- b) Estudiar la dependencia lineal de x e y . Hallar la recta de regresión de y sobre x y juzgar cómo es de bueno el ajuste.
11. En un estudio para relacionar las variables X = “Número de semanas de gestación” con Y = “Peso en gramos del niño al nacer”, se obtuvieron los siguientes resultados:

$$n = 5, \quad \sum x_i = 197, \quad \sum y_i = 15555, \quad \sum x_i^2 = 7785$$

$$\sum y_i^2 = 49193521, \quad \sum x_i y_i = 617055.$$

Hallar la recta de regresión de Y sobre X y utilizarla para predecir el peso de un niño con 40 semanas de gestación. ¿Es bueno el ajuste realizado con esta recta de regresión?

12. La siguiente tabla proporciona tres indicadores económicos para los países de la OCDE:

	X_1 Tasa real de crecimiento de PNB media 72/82	X_2 Tasa de desempleo en 1986	X_3 Tasa de incremento de índices de precios 1985
R.F.A.	2,0	8,00	2,2
Australia	2,8	7,50	7,6
Austria	2,6	4,50	2,9
Bélgica	2,2	13,50	4,6
Canada	2,8	9,50	4,1
Dinamarca	1,8	8,75	3,9
España	2,6	22,50	7,4
U.S.A.	2,2	7,25	3,2
Finlandia	3,1	6,25	5,1
Francia	2,7	10,75	5,3
Grecia	3,1	9,00	20,1
Holanda	1,9	14,00	2,3
Irlanda	4,0	16,75	5,5
Islandia	3,4	1,00	32,7
Italia	2,6	10,50	9,1
Japón	4,3	2,75	1,7
Luxemburgo	1,7	1,75	3,2
Noruega	4,0	2,25	5,8
N. Zelanda	1,6	4,75	16,3
Portugal	3,8	11,50	15,9
G. Bretaña	1,5	11,50	5,9
Suecia	1,6	3,00	6,7
Suiza	0,6	0,50	3,4
Turquía	5,1	13,50	40,5

Representar gráficamente los datos considerando los indicadores de dos en dos, y hallar los correspondientes coeficientes de correlación.

2. Estimación puntual y por intervalos

1. 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 de tipo A en la 10^a extracción.

La segunda persona obtiene el primer pez de tipo A en la 15^a extracción.

La tercera persona obtiene el primer pez de tipo A en la 18^a extracción.

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

2. Estimar la proporción de estudiantes que están a favor de dedicar el 0.7 por 100 de ayuda al tercer mundo si una muestra aleatoria de 2000 estudiantes de las universidades españolas ha dado una proporción muestral a favor de 0.72. Construir un intervalo de confianza al 99%.

3. 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. Hállese un estimador puntual y un intervalo de confianza al 90% para λ .

4. (Ejercicio de ordenador) En 1798, 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	

Representar los datos con el SPSS y comprobar si existe alguna evidencia de que no proceden de una distribución normal. Estimar la media y la varianza del modelo con el SPSS. Calcular el error típico de estimación de la media y estimar con el SPSS la media mediante un intervalo de confianza al 95%.

5. (Ejercicio de ordenador) La siguiente tabla proporciona datos sobre la precipitación total registrada en 11 estaciones meteorológicas de dos provincias españolas:

Prov. A	100	89	84	120	130	105	60	70	90	108	130
Prov. B	120	115	96	115	140	120	75	90	108	130	135

Comprobar con el SPSS si las muestras son independientes y si se puede suponer normalidad de los datos. Dar una estimación puntual y un intervalo de confianza para la diferencia de las medias de pluviosidad entre las dos provincias.

6. Dada una muestra aleatoria de tamaño n de una variable X , calcular el estimador de máxima verosimilitud en los siguientes casos:

a) $X \sim \text{Bernoulli}(p)$.

b) $X \sim \text{Poisson}(\lambda)$.

- c) $X \sim \text{Exponencial}(\lambda)$; 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) $X \sim N(\mu, \sigma)$.
7. El procedimiento que se describe a continuación se utiliza para evitar respuestas falsas a preguntas delicadas en una encuesta (método de respuesta aleatorizada):
 Sea A una pregunta delicada (por ejemplo, ¿evade Vd. impuestos?) y sea B una pregunta inocua (por ejemplo, ¿su número de D.N.I. es par?). Se le pide al sujeto que lance una moneda en secreto; si sale cara contesta la pregunta A y si sale cruz la B . El encuestador recibe una sola respuesta (sí o no) y no sabe a qué pregunta corresponde.
 Si esta prueba se realiza a 1.000 sujetos y 600 de ellos contestan “sí”, ¿qué porcentaje de individuos se estima que evade impuestos?
8. Para estudiar la proporción p de caballos afectados por la peste equina se les va a someter a una prueba. Sabemos que la prueba será positiva si el animal está enfermo; si está sano, hay una probabilidad 0.04 de que la prueba resulte positiva.
- a) Hallar la relación entre la probabilidad p de estar enfermo y la probabilidad q de dar positivo en la prueba.
- b) Obtener la estimación de máxima verosimilitud de p si 500 ejemplares son sometidos a la prueba y resulta positiva en 95 casos.
- c) Si realmente hay un 20% de caballos afectados por la epidemia, ¿cuál es la probabilidad de que la prueba resulte positiva en al menos 95 ejemplares de los 500?

9. La distancia X entre un árbol cualquiera y el árbol más próximo a él en un bosque sigue una distribución de Rayleigh con función de densidad

$$f_\theta(x) = 2\theta x \exp(-\theta x^2) \quad \text{si } x \geq 0 \quad (\theta > 0).$$

Obtener los estimadores de máxima verosimilitud de θ y de

$$g(\theta) = E_\theta(X) = (1/2)(\pi/\theta)^{1/2},$$

basados en muestras de tamaño n .

10. La lectura de voltaje dada por un voltímetro conectado a un circuito eléctrico es una variable aleatoria con distribución uniforme en el intervalo $(\theta, \theta+1)$, siendo θ el verdadero valor (desconocido) del voltaje. Sea (X_1, \dots, X_n) una muestra aleatoria de lecturas de dicho voltímetro.
- a) Demostrar que la media muestral \bar{X} es un estimador sesgado de θ y calcular el sesgo.
- b) Calcular el error cuadrático medio de \bar{X} .
- c) Obtener, a partir de \bar{X} , un estimador insesgado de θ .
11. En una población se desea conocer la probabilidad de que un individuo sea alérgico al polen de las acacias. En 100 individuos tomados al azar se observaron 10 alérgicos. Hallar el intervalo de confianza de nivel 0.95 para la probabilidad pedida. ¿Cuántos individuos se deberían observar para que, con probabilidad 0.95, el error máximo en la estimación de la proporción de alérgicos sea del 0.01?.

12. Se supone que el número de erratas por página en un libro sigue una distribución de Poisson. Elegidas al azar 95 páginas, se obtuvieron los siguientes resultados:

Número de erratas	0	1	2	3	4	5
Número de páginas	40	30	15	7	2	1

Hallar el intervalo de confianza de nivel 0.9 para el número medio de erratas por página en todo el libro.

13. En una experiencia genética, se extraen 20 moscas de una caja experimental. Medida la longitud del ala en cada mosca, se obtuvieron 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 sigue una distribución normal, hallar intervalos de confianza de nivel 0.9 para ambos parámetros.

14. Se quiere comparar 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) usando reiteradamente ambos métodos:

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.

Obtener un intervalo de confianza de nivel 0.95 para comparar las mediciones medias obtenidas por ambos métodos.

15. La vida activa (en días) de cierto fármaco sigue una distribución normal $N(1200, 40)$. Se desea enviar un lote de medicamentos de modo que la vida media del lote no sea inferior a 1180 días, con probabilidad 0.95. Hallar el tamaño del lote.

16. 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

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

3. Contraste de hipótesis paramétricas

1. 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 confianza del 90%. Plantear las hipótesis, indicar cuáles son las regiones crítica y de aceptación.

2. 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.

3. 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 confianza del 95%, que es menor la proporción de viviendas con calefacción en Extremadura que en Galicia?

4. 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:

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

¿Qué se puede decir al respecto?

5. 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

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

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 confianza del 95%.

6. Los datos de temperatura recogidos en 1000 estaciones meteorológicas en todo el mundo en el año 1950 dieron una temperatura promedio de 57°F. En 1988 la temperatura media en esas mismas estaciones fue 57.6°F. Se calculó la diferencia entre la lectura de 1988 y la de 1950 en cada estación y se obtuvo que la cuasi-desviación típica de la muestra resultante de las 1000 diferencias fue $s = 4.1^\circ\text{F}$. ¿Proporcionan estos datos suficiente evidencia estadística, al nivel 0.02, a favor de la hipótesis de que la temperatura media ha aumentado en 1988 respecto a 1950? ¿El p -valor del contraste es mayor o menor que 0.01?

Indicar claramente las suposiciones previas en las que se basa la validez del contraste empleado.

7. Se considera buena la edición de un libro si el número medio de erratas por página no supera el 0.1 (H_0). Dadas las pruebas de imprenta, se eligen 10 páginas al azar y se rechazan las pruebas si se observan 2 o más erratas. Se supone que el número de erratas por página sigue una Poisson.
- ¿Qué nivel de significación tiene el contraste?. ¿Con qué probabilidad aceptaremos un libro si realmente tiene una media de 0.2 erratas por página?.
8. 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 del 10%. Para ello 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.
- a) ¿Cuál es el nivel de significación de este contraste?
- b) Calcular la potencia del contraste si en realidad hay un 20% de peces que miden menos de 20 cm.
9. Para estudiar si una prueba de laboratorio puede resultar demasiado nociva para la salud, contrastaremos la hipótesis (H_0) de que la probabilidad de que una persona sometida a esa prueba resulte afectada sea como máximo 0.001. Para esto sometemos a esa prueba a 1000 personas elegidas al azar, y aceptamos H_0 si como máximo ha habido un afectado.
- a) Nivel de significación del contraste.
- b) Si en realidad la prueba afecta a la salud de una persona con probabilidad 0.003, ¿cuál es la probabilidad de aceptar H_0 ?
10. Se han realizado diversos estudios analizando la incidencia de la malaria en niños africanos heterocigóticos respecto al gen asociado a la anemia falciforme, y en niños homocigóticos normales. En un estudio realizado por Allison y Clyde se obtuvieron los siguientes datos: de 136 niños heterocigóticos 36 sufrieron la malaria, mientras que de 407 homocigóticos la sufrieron 152.
- ¿Hay evidencia estadística (al nivel 0.05) de que los heterocigóticos están mejor protegidos frente a la malaria que los homocigóticos normales?
11. Se recibe un envío de latas de conserva de las que se afirma que el peso medio son 1000 gramos. Examinada una muestra de 5 latas se obtiene un peso medio de 995 gr. con una cuasivarianza $s^2 = 19.6$. Al nivel de significación 0.05, ¿se puede aceptar que el peso medio son 1000 gr.?
12. La concentración media de dióxido de carbono en el aire en una cierta zona no es habitualmente mayor que 355 p.p.m.v. (partes por millón en volumen). Se sospecha que esta concentración es mayor en la capa de aire más próxima a la superficie. Para contrastar esta hipótesis se analiza el aire en 20 puntos elegidos aleatoriamente a una misma altura cerca del suelo. Resultó una media muestral de 580 p.p.m.v. 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.
13. La duración media de una muestra de 10 bombillas es $\bar{x} = 1250$ horas, con una cuasidesviación típica muestral de $s_X = 115$. Se cambia el material del filamento por otro nuevo y, entonces, de una muestra de 12 bombillas se obtuvo una duración media de $\bar{y} = 1340$ horas, con una cuasidesviación típica muestral de $s_Y = 106$.

- a) ¿Puede aceptarse que las varianzas, antes y después del cambio de filamento, son iguales? ¿Bajo qué hipótesis?
- b) ¿Ha aumentado la duración media de las bombillas?
14. 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, se pide:
- a) ¿Se puede aceptar igualdad de varianzas si sabemos que $\sum(x_i - \bar{x})^2 = 900000$ y $\sum(y_i - \bar{y})^2 = 950000$? (Tomar $\alpha = 0.10$).
- b) ¿Es más efectivo el medicamento B? Plantear el contraste adecuado para estudiar esto con un nivel de significación de 0.05.
15. Con objeto de estudiar si las pulsaciones en los hombres pueden considerarse menores que en las mujeres, se tomaron muestras de 16 hombres (X_i) y 16 mujeres (Y_i), obteniéndose los siguientes datos:

X_i :	74	77	71	76	79	74	83	79	83	72	79	77	81	79	84	80
Y_i :	81	84	80	73	78	80	82	84	80	84	75	82	79	82	79	85

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

¿Qué se puede decir al respecto?

16. Se tienen algunos indicios de que el consumo de tabaco tiende a provocar problemas de trombosis debidos a un aumento en la capacidad de coagulación. Para estudiar esta hipótesis, Levine (1973) extrajo muestras de sangre de 11 individuos antes y después de que fumasen un cigarrillo y midió la capacidad de agregación de las plaquetas, obteniendo los datos siguientes (correspondientes al máximo porcentaje de plaquetas que se agregaron después de haber sido sometidas a un estímulo adecuado):

Antes:	25	25	27	44	30	67	53	53	52	60	28
Después:	27	29	37	56	46	82	57	80	61	59	43

¿Hay suficiente evidencia estadística (al nivel de significación 0.01) a favor de la hipótesis de que los fumadores presentan una mayor tendencia a la formación de coágulos? Indicar las condiciones que se requieren para que el test utilizado sea válido.

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 significación del 0.1).

18. Con el fin de estudiar el efecto de los rayos X sobre la viabilidad huevo-larva en “*Tribolium castaneum*” se irradiaron 1000 huevos de los que resultaron 572 larvas. Sabiendo que la viabilidad normal de esta especie es del 63%, contrastar la hipótesis nula de que la radiación no ha tenido efecto sobre la viabilidad (al nivel de significación 0.05).
19. Un fabricante de lavadoras produce un determinado modelo en dos colores A y B. De las 1000 primeras lavadoras vendidas, 560 fueron del color A. ¿Proporcionan estos datos suficiente evidencia estadística (al nivel de significación 0.01) para concluir que los consumidores prefieren mayoritariamente el color A?
20. Un método de tratamiento contra la leucemia mieloblástica aguda consiste en someter al paciente a quimioterapia intensiva. Se sabe que este tratamiento proporciona un porcentaje de remisión de un 70%. Se aplica un nuevo método de tratamiento a 50 voluntarios. ¿Cuál es el mínimo número de casos de remisión de la enfermedad que debe observarse para poder afirmar (a un nivel de significación del 0.025) que el nuevo método produce una tasa de remisión más alta que el antiguo?
21. Un examen consta de 100 preguntas con 4 alternativas cada una. Se desea que los alumnos que superen la prueba sepan, al menos, el 50% de la asignatura. ¿Cuál debe ser el número mínimo de respuestas correctas para aprobar el examen, con un nivel de significación del 0.01? Estudiar las dos formas posibles de plantear el contraste. (Sólo una de las alternativas es correcta y las preguntas abarcan uniformemente toda la asignatura).
22. Una prueba de detección de la hepatitis vírica produce un 2% de falsos positivos (prueba positiva en una persona sana) y un 5% de falsos negativos (prueba negativa en una persona enferma). Se aplica esta prueba a 800 personas independientes tomadas al azar en la población.
 - a) Hallar la relación entre p =“Probabilidad de dar positivo” y r =“Probabilidad de padecer hepatitis vírica”.
 - b) ¿Cuál es el máximo número de pruebas positivas que podríamos obtener, entre las 800, para considerar estadísticamente probado que la enfermedad afecta a menos del 8% de la población? Tomar $\alpha = 0.01$.

4. Contrastes de hipótesis no paramétricas

1. Un estudio sobre tabaquismo en tres comunidades, mediante tres muestras aleatorias de tamaño 100, proporciona los siguientes resultados:

Comunidad	fumadores	no fumadores
A	13	87
B	17	83
C	18	82

¿Pueden considerarse homogéneas las tres poblaciones en cuanto a sus hábitos fumadores, al nivel 0.05?

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

Resultado	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?

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

	Masculino	Femenino
Normales	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?

4. 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 0.01, que el programa funciona correctamente?

5. En el transcurso de dos horas, el número de llamadas por minuto, solicitadas a una centralita telefónica fue:

Núm. llamadas/minuto	0	1	2	3	4	5	6
Frecuencia	6	18	32	35	17	10	2

¿Se puede aceptar que el número de llamadas por minuto sigue una distribución de Poisson?

6. Una fábrica de automóviles quiere averiguar si la preferencia de modelo tiene relación con el sexo de los clientes. Se toman dos muestras aleatorias de 1000 hombres y 1000 mujeres observándose las siguientes preferencias:

SEXO	MODELO		
	A	B	C
Mujer	340	400	260
Hombre	350	270	380

¿Son homogéneas las preferencias entre hombres y mujeres, al nivel de significación 0.01?

7. 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

O	A	B	AB
r^2	$p^2 + 2pr$	$q^2 + 2qr$	$2pq$

donde $p + q + r = 1$.

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.

8. 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?

9. 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.