

Capítulo 1: Estadística descriptiva

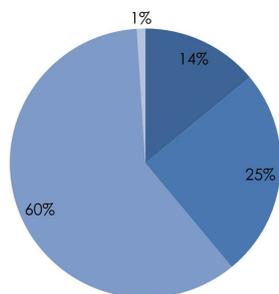
1. El Ministerio de Sanidad, Servicios Sociales e Igualdad ha publicado recientemente el Barómetro Sanitario nacional, correspondiente a 2011.
 - a) Según los resultados del estudio, de los 7757 encuestados, 2487 afirman ser fumadores en la actualidad, 5262 afirman ser no fumadores y el resto no contesta. Construye la tabla de frecuencias correspondiente y realiza una gráfica representativa de la distribución de la muestra según el hábito de fumar.
 - b) A aquellos que habían declarado ser fumadores se les preguntó a continuación como habían influido en su consumo las medidas de la nueva Ley del tabaco. A continuación se muestra una tabla incompleta que resume las respuestas. Completa la tabla y realiza una gráfica representativa.

Influencia de la Ley del tabaco	n_i	f_i
Fuma menos que antes		
Fuma más que antes	125	
No ha influido en el consumo		0.68275
No contesta	28	

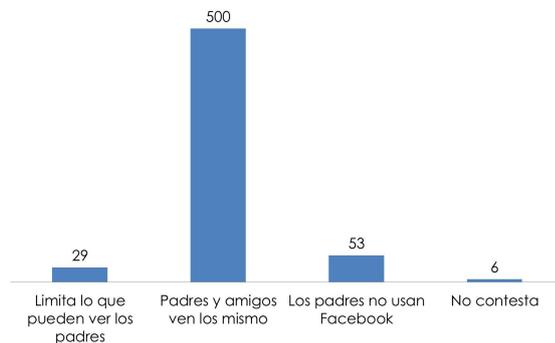
2. Un estudio llevado a cabo por el Pew Research Center's Internet & American Life Project (<http://www.pewinternet.org>) tiene como objetivo analizar la actitud de los jóvenes en EEUU ante las redes sociales y su configuración de la privacidad. Para ello se ha llevado a cabo una encuesta entre usuarios de Facebook. A continuación se muestran dos gráficas con datos de dicho estudio.

Configuración de la privacidad del perfil

■ Público
■ Parcialmente privado
■ Privado (solo amigos)
■ No sabe



¿Compartes la misma información con tus padres y amigos?

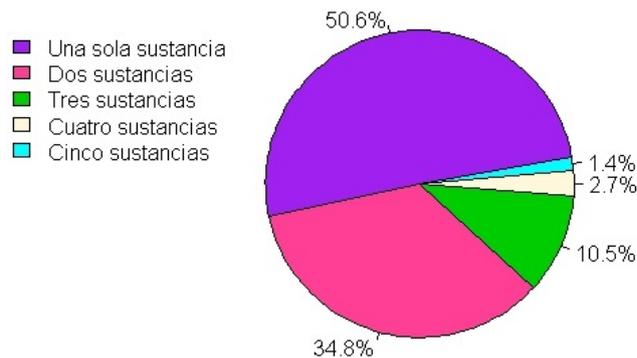


- a) ¿Qué gráficas aparecen representadas? ¿A qué tipo de variables hacen referencia?
- b) ¿Cuál es el tamaño muestral? ¿Cuántos encuestados tienen su perfil parcialmente privado? ¿Qué porcentaje de usuarios comparte la misma información con sus padres y amigos?
- c) Supongamos que les preguntamos a los encuestados cuántos amigos tienen en Facebook. ¿Qué tipo de gráfico crees que deberías utilizar para resumir esa información?

3. Se muestran a continuación los tiempos (en minutos) de 15 participantes en la última etapa de la Vuelta a España (2014), correspondiente a la contrarreloj celebrada en Santiago de Compostela el 14 de septiembre de 2014.

12.33, 12.34, 12.43, 12.66, 13.05, 11.95, 12.56,
12.15, 11.33, 11.96, 13.83, 13.56, 11.83, 13.08, 12.81.

- Contruye la tabla de frecuencias correspondiente.
 - Realiza una gráfica representativa e interprétala.
 - Calcula el tiempo medio de la etapa.
 - Representa el diagrama de cajas correspondiente.
4. De los datos recogidos en la "Encuesta domiciliaria sobre alcohol y drogas en España 2009/2010" (Delegación del Gobierno para el Plan Nacional Sobre Drogas. Ministerio De Sanidad, Política Social e Igualdad) se deduce que el policonsumo de drogas (legales e ilegales) es un patrón de consumo cada vez más prevalente en España y en Europa. Analiza el siguiente gráfico sobre el porcentaje de consumidores que han consumido una o más sustancias en el último año.



- ¿Qué variable se resume en el diagrama de sectores? ¿De qué tipo es?
 - Calcula la moda, la mediana y el cuantil 0.9.
 - ¿Cuál es el número medio de sustancias consumidas por individuo?
 - Supongamos que los datos de la gráfica han sido obtenidos a partir de una muestra formada por 1000 individuos. Si se añaden 500 consumidores más al estudio cuyo consumo medio es de 2.5 sustancias, ¿cuál es el número medio de sustancias consumidas para el total de 1500 personas?
5. Los siguientes datos corresponden a los pesos (en kg.) de 12 niñas de 5 años de edad.

15.0	17.3	18.3	21.9	13.8	20.8
17.5	19.7	15.1	26.7	20.4	16.4

- Calcula el peso medio y el peso mediano. Calcula la varianza y la desviación típica.
- Calcula Q_1 , Q_3 y el cuantil 0.9. Representa el boxplot de los datos.

Puedes ver los patrones de crecimiento infantil de la OMS en <http://www.who.int/childgrowth/es/>

6. El volumen corpuscular medio (VCM) es uno de los parámetros calculados en un examen de conteo sanguíneo completo. El VCM indica el tamaño de los glóbulos rojos y se mide en fentolitros. A continuación se muestran los valores de VCM de 19 pacientes que se han sometido a un examen de conteo sanguíneo completo.

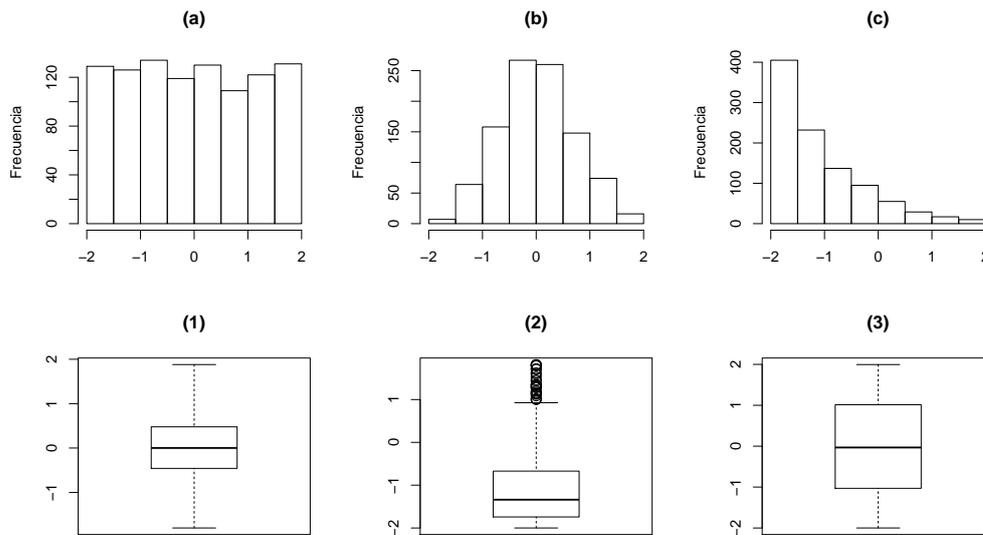
83, 77, 82, 84, 85, 92, 92, 93, 91, 86, 89, 109, 81, 79, 81, 88, 110, 90, 80.

- a) Construye la tabla de frecuencias y representa el histograma correspondiente.
 b) Dibuja el boxplot de estos datos.

7. En un centro de salud se registra el peso (en kg.) y la altura (en cm.) de los pacientes que han acudido a consulta el último mes. La siguiente tabla presenta un resumen de las respuestas proporcionadas por dicho grupo de pacientes. ¿Qué medidas presentan mayor variabilidad?

Variable	Media	Desviación típica
Peso	65.4	12.2
Altura	170.5	9.42

8. A continuación aparecen representados los histogramas y diagramas de cajas de tres conjuntos de datos distintos. Empareja cada histograma con el diagrama de cajas que le corresponde.



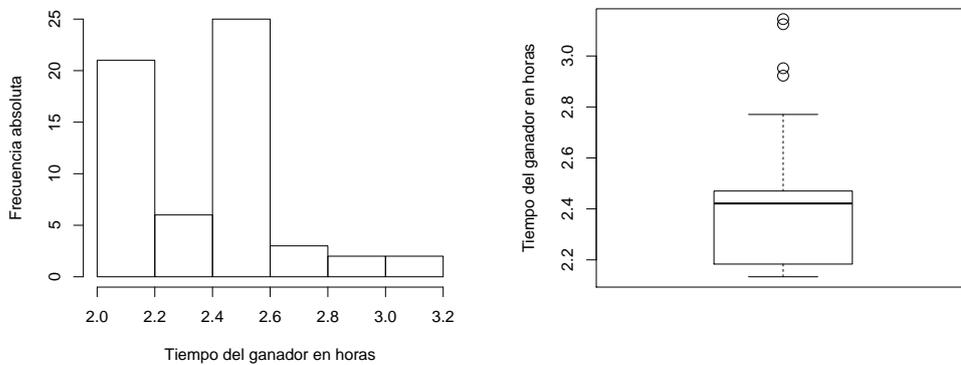
9. Se ha registrado la temperatura de 9 pacientes. La temperatura mínima registrada ha sido 36°C y la temperatura máxima registrada ha sido 39.2°C. La mediana es 37.2°C.

- a) Si eliminamos la observación correspondiente a la temperatura mínima registrada y la sustituimos por una nueva temperatura de 35.9°C, ¿cuál es la mediana del nuevo conjunto de datos?
 b) Si eliminamos la observación correspondiente a la temperatura máxima registrada y la sustituimos por una nueva temperatura, ¿cuál es la mediana del nuevo conjunto de datos?

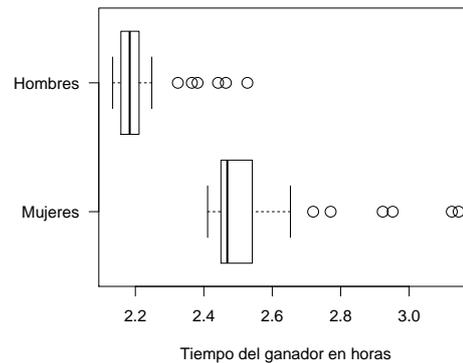
10. Para cada uno de los apartados a continuación, compara los conjuntos de datos (1) y (2) en términos de media y desviación típica. No es necesario que calcules los valores exactos, simplemente discute en qué casos son iguales o de qué manera difieren.

- a) (1) 0, 2, 4, 6, 8, 10.
(2) 20, 22, 24, 26, 28, 30.
- b) (1) 100, 200, 300, 400, 500.
(2) 0, 50, 300, 550, 600.
- c) (1) 0, 2, 4, 6, 8, 10.
(2) 0, 6, 12, 18, 24, 30.

11. El histograma y el diagrama de cajas que se muestran a continuación muestran la distribución de los tiempos de los ganadores (categoría masculina y femenina) del maratón de Nueva York entre 1970 y 1999.



- a) ¿Qué aspectos de la distribución de los tiempos de los ganadores se ven reflejados en el histograma pero son menos evidentes en el boxplot? ¿Qué aspectos de la distribución de los tiempos de los ganadores se deducen del boxplot pero son menos evidentes en el histograma? ¿Por qué la que la distribución de los tiempos es bimodal?
- b) A la vista del diagrama de cajas que aparece a continuación, compara la distribución de los tiempos de los ganadores en función del sexo.



En las primeras ediciones el maratón se celebraba dando varias vueltas alrededor de un circuito por Central Park, pero la carrera fue adquiriendo gran popularidad y en 1976 se cambió el trazado. Los datos atípicos corresponden a los tiempos de 1970 hasta 1975.

Capítulo 2: Probabilidad

1. Un hospital clasifica a cada paciente según disponga o no de seguro médico y según su estado de salud, que puede ser catalogado como bueno, aceptable, serio o crítico. El administrador registra primero un 0 si el paciente no tiene seguro y un 1 si lo tiene y después registra una de las letras b, a, s, c, según el estado en que se encuentre el paciente. Considera el experimento consistente en otorgar un código a un paciente nuevo.
 - a) ¿Cuál es el espacio muestral de este experimento?
 - b) Indica cuál es el suceso de que el paciente está en estado serio o crítico.
 - c) Indica cuál es el suceso de que el paciente está en estado serio o crítico y no tiene seguro.
 - d) Indica cuál es el suceso de que el paciente tiene seguro.

2. El técnico de un centro de transfusión determina el grupo sanguíneo del sistema AB0 y el factor Rh de las personas que acuden al centro. ¿Cuál es el espacio muestral de este experimento? Lista los elementos de los siguientes sucesos:
 - a) Un individuo es del grupo sanguíneo A.
 - b) Un individuo es Rh+.

3. Según datos del centro de transfusión de Galicia, el 85 % de las personas son Rh positivas. Además, la distribución de los grupos sanguíneos del sistema AB0 no es uniforme. En España, el 44 % de la población tiene el grupo 0, el 43 % tiene grupo A, el 10 % tiene grupo B y sólo el 3 % de la población tiene grupo AB.
 - a) Un individuo se acerca al centro de transfusión a donar sangre. ¿Cuál es la probabilidad de que sea del grupo A o del grupo B? (Sol: 0.53)
 - b) Teniendo en cuenta que el factor Rh es independiente del grupo sanguíneo del sistema AB0, ¿cuál es la probabilidad de que un individuo sea A+? (Sol: 0.365)
 - c) El centro de transfusión de Galicia realiza campañas informativas para concienciar al grupo de donantes 0- (donante universal). ¿Qué porcentaje de la población cuenta con este tipo de sangre? (Sol: 6.6 %)
 - d) Si conoces tu grupo y factor sanguíneo, calcula el porcentaje de la población con tu mismo tipo de sangre. ¿Qué porcentaje de la población es compatible con tu tipo de sangre?

Puedes consultar una tabla de compatibilidad entre donantes y receptores en
<http://www.donarsangre.org/donantes-de-sangre/grupos-sanguineos/>

4. Estudios sobre la depresión muestran que la aplicación de un determinado tratamiento mejora el estado del 72 % de aquellas personas sobre las que se aplica, no produce efecto alguno en un 10 % y empeora el estado en el resto. Se trata a un paciente que sufre depresión por estos medios. ¿Cuál es la probabilidad de que el tratamiento no vaya en detrimento de su estado? (Sol: 0.82)

5. Tratando a bebés prematuros, la cantidad de oxígeno recibido puede afectar a su visión. Se puede categorizar a cada niño tratado como de visión normal, de lesión media, de lesión moderada, de lesión grave o ciego. Un estudio muestra que la probabilidad de que ocurra cada uno de estos sucesos es de 0.80, 0.10, 0.06, 0.02 y 0.02, respectivamente.
- Determina el espacio muestral Ω para el experimento que consiste en la observación del estado de visión de un bebé prematuro.
 - Determinar la probabilidad de que un niño nazca con algún problema de visión. (Sol: 0.2)
 - ¿Cuál es la probabilidad de que, al seleccionar aleatoriamente a un bebé prematuro, nazca con visión normal? (Sol: 0.8)
6. El 4 % de las personas de una población son daltónicas, el 18 % son hipertensas y el 0.5 % son daltónicas e hipertensas. ¿Cuál es el porcentaje de personas que son daltónicas o hipertensas? (Sol: 21.5 %)
7. El 60 % de los individuos de una población están vacunados contra una cierta enfermedad. Durante una epidemia se sabe que el 20 % la ha contraído y que 2 de cada 100 individuos están vacunados y son enfermos. Calcular el porcentaje de vacunados que enferma y el de vacunados entre los que están enfermos. (Sol: 3.33 % y 10 %, respectivamente)
8. Se sabe que entre la población total de Estados Unidos, el 55 % padece de obesidad, el 20 % es hipertensa, y el 60 % es obesa o hipertensa. ¿Es independiente el que una persona sea obesa de que padezca hipertensión? (Sol: No)
9. Sninsky y otros realizaron un estudio para evaluar la eficacia y seguridad de una preparación de mesalamina oral recubierta de polímero sensible al pH en pacientes con actividad de leve a moderada de colitis ulcerosa. En la siguiente tabla se muestran los resultados del tratamiento al final de seis semanas, por tratamiento recibido:

Resultado	Grupo en tratamiento		
	Placebo	Mesalamina 1.6 g/día	Mesalamina 2.4 g/día
En remisión	2	6	6
Mejorado	8	13	15
Estable	12	11	14
Empeorado	22	14	8

- ¿Cuál es la probabilidad de que un paciente seleccionado aleatoriamente entre en remisión al final de seis semanas? (Sol: 14/131)
 - ¿Cuál es la probabilidad de que un paciente que recibe placebo logre la remisión al final de las seis semanas? (Sol: 2/44)
 - ¿Cuál es la probabilidad de que un paciente seleccionado aleatoriamente haya entrado en remisión y sea uno de los que recibió placebo? (Sol: 2/131)
 - ¿Cuál es la probabilidad de que un paciente seleccionado aleatoriamente sea uno de los que recibieron dosis de 2.4 g/día o esté en la lista de pacientes mejorados, o posea ambas condiciones? (Sol: 64/131)
10. A un congreso internacional de pediatría asisten cien personas, de las cuales 60 hablan sólo inglés, 30 sólo francés y los 10 restantes ambos idiomas. Calcular la probabilidad de que se entiendan dos congresistas elegidos al azar. (Sol: 0.636)

11. Según datos de un estudio sobre la calidad del sistema sanitario a nivel mundial, en un determinado país el 61 % de las personas recibe asistencia sanitaria pública, el 24 % de las personas contrata asistencia sanitaria privada, y el 8 % comparten asistencia pública y privada.
- Calcula el porcentaje de personas que tienen cobertura sanitaria de algún tipo. (Sol: 77 %)
 - ¿Cuál es la probabilidad de que un habitante de dicho país reciba asistencia pública si sabemos que está pagando asistencia sanitaria privada? (Sol: 0.33)
 - ¿Son independientes los sucesos “recibir asistencia sanitaria pública” y “contratar asistencia sanitaria privada”? (Sol: No)
 - ¿Cuál es la probabilidad de que un habitante de dicho país contrate asistencia privada sabiendo que no recibe asistencia sanitaria pública? (Sol: 0.41)
12. Se considera una célula en el instante $t = 0$. En el instante $t = 1$ la célula puede reproducirse dividiéndose en dos, con probabilidad $3/4$; o bien morir, con probabilidad $1/4$. Si la célula se divide, entonces, en el tiempo $t = 2$, cada uno de sus dos descendientes puede también subdividirse o morir, independientemente uno de otro y con las mismas probabilidades de antes. ¿Cuántas células puede haber en el tiempo $t = 2$? ¿Con qué probabilidad? (Sol: Puede haber 0, 2 ó 4 células con probabilidades 0.2968, 0.28125, y 0.4218, respectivamente)
13. Elegido un individuo al azar y observado por rayos X, se diagnosticó que estaba tuberculoso. La probabilidad de que en la población de la que se eligió el individuo, uno de ellos sea tuberculoso es de 0.01. La sensibilidad de la prueba es de 0.97 y la probabilidad de falso positivo es 0.001. ¿Cuál es la probabilidad de que el individuo sea tuberculoso, habiéndolo diagnosticado como tal por rayos X? (Sol: 0.9073)
14. Una enfermedad puede estar producida por tres virus A, B y C. en el laboratorio hay 3 tubos de ensayo con el virus A, 2 tubos con el virus B y 5 tubos con el virus C. La probabilidad de que el virus A produzca la enfermedad es de $1/3$, que la produzca B es de $2/3$ y que la produzca el virus C es de $1/7$. Se inocula un virus a un animal y contrae la enfermedad, ¿Cuál es la probabilidad de que el virus que se inocule sea el C? (Sol: 0.2343)
15. Los estudios epidemiológicos indican que el 20 % de los ancianos sufre un deterioro neuropsicológico. Sabemos que la tomografía axial computerizada (TAC) es capaz de detectar este trastorno en el 80 % de los que lo sufren, pero también da un 3 % de falsos positivos entre las personas sanas. Si tomamos un anciano al azar y da positivo en el TAC, ¿cuál es la probabilidad de que esté realmente enfermo? (Sol: 0.8695)
16. Una ambulancia en la plaza Roja, al trasladarse hacia el hospital, puede hacerlo por la calle Fray Rosendo Salvado, República del Salvador o San Pedro de Mezonzo, con probabilidades 0.2 , 0.7 y 0.1, respectivamente. La probabilidad de que la ambulancia sufra un atasco por la calle Fray Rosendo Salvado es 0.5, por la calle República del Salvador es 0.6 y por la calle San Pedro de Mezonzo es 0.4.
- Calcula la probabilidad de que la ambulancia quede atrapada en un atasco. (Sol: 0.56)
 - Si la ambulancia ha llegado al hospital sin sufrir ningún atasco, ¿cuál es la probabilidad de que haya elegido circular por la calle Fray Rosendo Salvado? (Sol: 0.227)
17. Con el objeto de diagnosticar la colelitiasis se usan ultrasonidos. Tal técnica tiene una sensibilidad del 91 % y una especificidad del 98 %. En la población que nos ocupa la probabilidad de colelitiasis es de 0.2.

- a) Si a un individuo de tal población se le aplican los ultrasonidos y dan positivos, ¿cuál es la probabilidad de que sufra colelitiasis? (Sol: 0.9191)
- b) Si el resultado fuese negativo, ¿cuál sería la probabilidad de que no tenga la enfermedad? (Sol: 0.9775)
18. Un equipo de investigación médica pretende evaluar la utilidad de cierto síntoma (llamado S) para el diagnóstico de determinada enfermedad. En una muestra aleatoria independiente de 775 pacientes con esa enfermedad, 744 pacientes presentaron el síntoma. En una muestra aleatoria independiente de 1380 individuos sin la enfermedad, 21 presentaron el síntoma.
- a) Para el contexto de este ejercicio, ¿qué es un falso positivo? ¿Qué es un falso negativo?
- b) Calcular la sensibilidad y la especificidad del síntoma (Sol: 0.96 y 0.984, respectivamente)
- c) Supongamos que la tasa de la enfermedad en la población general es 0.001, ¿cuál es el valor que predice la positividad del síntoma? ¿Cuál es el valor que predice la negatividad del síntoma? (Sol: 0.059 y 0.99, respectivamente)
- d) Calcular los valores que predicen la positividad y la negatividad del síntoma para las siguientes tasas hipotéticas: 0.0001, 0.01, 0.10.
19. Una encuesta del proyecto *Pew Internet and American Life Project* llevada a cabo en 2010 determinaba que el 17 % de los encuestados que poseían un teléfono móvil utilizaban su teléfono para consultar información sobre salud. En el año 2012 se llevó a cabo una nueva encuesta de la que se obtienen los siguientes datos: el 85 % de los encuestados posee teléfono móvil. De ellos, el 53 % utiliza un smartphone. El 52 % de los poseedores de un smartphone afirman usar su teléfono para consultar información sobre salud, mientras que solo el 6 % de los poseedores de un móvil convencional afirma usar su teléfono para consultar información sobre salud. ¿Qué porcentaje de encuestados en 2012 con móvil (convencional o smartphone) utiliza su teléfono para consultar información sobre salud? Compara el resultado con el dato de 2010. (Sol: 30.38 %)
20. En un examen mamográfico se diagnostica un positivo. La mujer en cuestión pregunta a su médico: “¿es seguro que sufro cáncer? Si no es así, ¿cuál es la probabilidad de que lo tenga?”. Se sabe que la probabilidad de que una mujer sufra cáncer de mama es del 1 %. Si una mujer sufre cáncer de mama, la probabilidad de que la prueba ofrezca un resultado positivo es del 90 %. Si una mujer no padece cáncer de mama, la probabilidad de que aun así la prueba de un resultado positivo es del 9 %. De las siguientes respuestas, ¿cuál sería la más acertada?
- a) La probabilidad de que sufra cáncer ronda el 81 %.
- b) Aproximadamente 9 de cada 10 mujeres con una mamografía positiva, sufren cáncer.
- c) Aproximadamente 1 de cada 10 mujeres con una mamografía positiva, sufren cáncer.
- d) La probabilidad de que sufra cáncer ronda el 1 %.

El artículo “Helping Doctors and Patients Make Sense of Health Statistics” de Gigerenzer et al. (2008) analiza la interpretación que hacen los profesionales médicos de los estudios de salud. Se planteó este mismo problema a 160 ginecólogos. El 60 % contestó que la respuesta correcta era la a) o la b). Sólo el 21 % de los encuestados contestó correctamente la c).

21. En una determinada población, la probabilidad de que un varón elegido al azar sufra algún problema cardiovascular es 0.25. Los varones con problemas cardiovasculares presentan el doble de probabilidad de ser fumadores que aquellos que no sufren problemas cardiovasculares. Elegido un varón al azar, ¿cuál es la probabilidad de que sufra algún problema cardiovascular si sabemos que es fumador? (Sol: 0.4)

Capítulo 3: Variables aleatorias

1. Sea X la variable aleatoria que expresa el número de pacientes con enfermedades articulares en centros de salud con las siguientes probabilidades:

x_i	0	1	2	3	4	5	6	7
p_i	0.230	0.322	0.177	0.155	0.067	0.024	0.015	0.01

Comprueba que se trata efectivamente de una distribución de probabilidad y represéntala. Calcula y representa la función de distribución. ¿Cuál es el número medio de pacientes con enfermedades articulares? (Sol: $\mu=1.689$)

2. En el grupo de adultos (> 16 años) la probabilidad de sobrevivir al trasplante de médula ósea en talasemia es 0.6. Un centro hospitalario planea realizar trasplantes de médula ósea a 3 pacientes adultos.
- Escribe el espacio muestral correspondiente al posible resultado de las 3 operaciones de trasplante.
 - Considera la variable X =Número de pacientes que sobreviven al trasplante. Calcula y representa la función de masa y la función de distribución de la variable X .
 - ¿Cuál es la probabilidad de que sobrevivan exactamente 2 pacientes al trasplante de médula? (Sol: 0.432)
3. Supongamos que el 40 % de los enfermos de una determinada dolencia se recuperan. Si en un centro hospitalario hay 4 pacientes internados que sufren de esa dolencia,
- ¿Cuál es la probabilidad de que 2 se recuperen? (Sol: 0.3456)
 - ¿Cuál es la probabilidad de que todos se recuperen? (Sol: 0.0256)
 - ¿Cuál es la probabilidad de que al menos 2 se recuperen? (Sol: 0.5248)
4. La probabilidad de que un paciente que acude a una consulta de atención primaria sea derivado a otra consulta es 0.2. Si a una consulta de atención primaria acuden 5 pacientes calcula:
- La probabilidad de que sean derivados exactamente 3 pacientes. (Sol: 0.0512)
 - La probabilidad de que sean derivados exactamente 5 pacientes. (Sol: 0.00032)
 - La probabilidad de que sean derivados menos de 5 pacientes. (Sol: 0.99968)
 - Calcula el número medio de pacientes derivados a otra consulta, la varianza y la desviación típica. (Sol: $\mu = 1$, $\sigma^2 = 0.8$, $\sigma = 0.894$)
5. En un hospital, el número medio de pancreatitis agudas atendidas al día es 0.9. Calcula la probabilidad de que un día determinado sean atendidas 3 pancreatitis agudas en dicho hospital. (Sol: 0.04939)

6. Se estima que la probabilidad de que haya complicaciones graves en pacientes con fallos coronarios ingresados en la UCI es 0.05. Si en la UCI de un determinado hospital hay ingresados 60 pacientes con fallos coronarios, ¿cuál es la probabilidad de que ninguno de ellos sufra complicaciones graves? (Sol: Utilizando la binomial, la probabilidad es 0.04606. Utilizando la aproximación por la Poisson, 0.049787)
7. En un hospital, el número medio de ingresos por día en la unidad de quemados es 8.4. Calcula:
- La probabilidad de que una semana haya entre 50 y 55 ingresos en la unidad de quemados. (Sol: 0.229)
 - La probabilidad de que un día haya exactamente dos ingresos en la unidad de quemados. (Sol: 0.0079)
 - La probabilidad de que un día haya al menos un ingreso en la unidad de quemados. (Sol: 0.9997)
8. Un estudio sobre salud laboral establece que el 9 % de los profesores que imparten clase en centros de Primaria y Secundaria se da de baja por sufrir alguna patología psiquiátrica, siendo la más común la depresión, aunque también hay casos de estrés o neurosis.
- Supongamos que un determinado centro de primaria cuenta con 7 docentes. ¿cuál es la probabilidad de que ninguno de ellos solicite la baja por alguna patología psiquiátrica? (Sol: 0.5167)
 - ¿cuál es la probabilidad que ningún docente solicite la baja por alguna patología psiquiátrica en un centro con 60 docentes? (Sol: Utilizando la binomial, la probabilidad es 0.00348. Utilizando la aproximación por la Poisson, 0.00451)
9. El gerente de un centro de atención primaria sabe, por experiencia, que el 20 % de las personas que solicitan cita previa no asisten a la consulta. Si el centro da 10 citas pero solo puede atender a 8 pacientes, ¿cuál es la probabilidad de que todas las personas que acuden con cita previa a la consulta sean atendidas? (Sol: 0.6241)
10. Diez individuos entran en contacto con un portador de tuberculosis. La probabilidad de que la enfermedad se contagie del portador a un sujeto cualquiera es de 0.1. ¿Cuántos individuos se espera que contraigan la enfermedad? (Sol: $\mu = 1$)
11. Sea X una variable con distribución binomial, con media 2 y varianza $4/3$.
- Determina la función de distribución de X y represéntala gráficamente.
 - Calcula la media y varianza de $Y=4X+3$. (Sol: Utilizando las propiedades de media y varianza, $E(Y) = E(4X + 3) = 4E(X) + 3 = 11$, $Var(Y) = Var(4X + 3) = 16Var(X) = 64/3$)
12. Se está realizando un estudio de cribado de cáncer de colon en una determinada población. Se estudiará a sujetos de 50 a 75 años con la prueba Hemoccult, en la que se comprueba la presencia de sangre en una muestra de heces. La prueba Hemoccult tiene una sensibilidad del 88 % y una especificidad del 96 %. La población de estudio tiene una prevalencia de cáncer de colon del 1.2 %. Se realiza la prueba de Hemoccult a cinco sujetos. ¿Cuál es la probabilidad de que la prueba no resulte positiva en ninguno de ellos? (Sol: 0.77)
13. En la consulta del Dr. Martín se registra una media de 2 pacientes cada hora. De manera independiente, en la consulta del Dr. González se registra una media de 3 pacientes cada hora. Supuesto que el modelo de Poisson es adecuado para describir el número de pacientes que llegan en un cierto intervalo de tiempo:

- a) Calcula la probabilidad de que entre las nueve y las nueve y media de la mañana, acuda algún paciente a la consulta del Dr. Martín. (Sol: 0.632)
- b) Calcula la probabilidad de que entre las nueve y las nueve y media de la mañana, ambas consultas tengan algún paciente. (Sol: 0.491)
- c) Calcula la probabilidad de que Dr. González no tenga que atender pacientes durante 40 minutos. (Sol: 0.135)
14. Un médico aplica una prueba a 6 adultos para detectar una enfermedad cuya prevalencia es del 10 %. La sensibilidad de la prueba es del 95 % y la especificidad es del 88 %.
- a) ¿Cuál es la probabilidad de que exactamente a 4 personas les de un resultado negativo? (Sol: 0.249)
- b) Si se realiza la prueba a 11 pacientes, ¿cuántos se espera que den positivo? (Sol: $\mu = 2.23$)
15. Se lanza una moneda 3 veces. Consideramos los eventos $A =$ “Obtener dos o más cruces” y $B =$ “Obtener una o dos caras”. ¿Son A y B independientes? ¿Y si la moneda se lanza 4 veces? (Sol: Con 3 lanzamientos A y B son independientes. Con 4 lanzamientos A y B no son independientes.)
16. Un aparato de registro de movimientos oculares presenta una probabilidad de error de 0.2. Se realizan 8 mediciones independientes con dicho aparato. Sea X el número de errores cometidos por el aparato en dichas mediciones:
- a) Calcula la función de masa de X y su media. (Sol: $\mu = 1.6$)
- b) ¿Cuál es el número de errores más probable? (Sol: 1 error)
- c) Sabiendo que se ha producido al menos un error en las mediciones, calcula la probabilidad de que se hayan producido exactamente 2 errores. (Sol: 0.3528)
17. Se lanza una moneda 3 veces. Consideramos los eventos $A =$ “Obtener dos o más cruces” y $B =$ “Obtener una o dos caras”. ¿Son A y B independientes? ¿Y si la moneda se lanza 4 veces? (Sol: Con 3 lanzamientos A y B son independientes. Con 4 lanzamientos A y B no son independientes.)
18. La líneas telefónicas de un centro de salud comunican el 60 % del tiempo.
- a) Si llamas al centro de salud, ¿con que probabilidad te contestarán en el primer intento? ¿y en el segundo? ¿y en el tercer intento? (Sol: 0.4, 0.24 y 0.144, respectivamente)
- b) Un paciente ha llamado al centro de salud 5 veces a lo largo del día. ¿Cuál es la probabilidad de que atendiesen exactamente 2 llamadas? ¿Cuál es la probabilidad de que en la quinta llamada sea la segunda vez que le atienden? (Sol: 0.3456 y 0.1382, respectivamente)

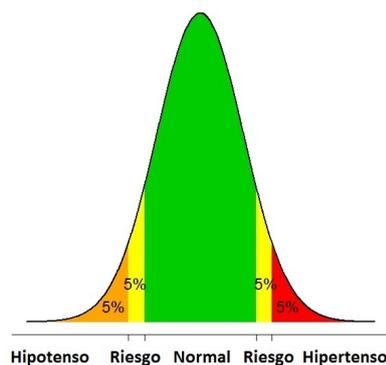
Capítulo 4: Variables aleatorias continuas

1. Comprueba que en una normal estándar $N(0, 1)$:
 - a) Aproximadamente el 68 % del área encerrada bajo la función de densidad está contenida entre -1 y $+1$.
 - b) Aproximadamente el 95 % del área encerrada bajo la función de densidad está contenida entre -2 y $+2$.
 - c) Aproximadamente el 99 % del área encerrada bajo la función de densidad está contenida entre -3 y $+3$.
2. Sea Z una variable aleatoria normal estándar. Calcula:
 - a) El área encerrada por la función de densidad entre $z = 0$ y $z = 1.35$.
 - b) $P(Z \leq 2)$
 - c) $P(-0.5 \leq Z \leq 2.65)$
 - d) El valor de z ($z > 0$) de manera que el área encerrada entre 0 y z sea 0.2 .
 - e) El valor de z tal que la probabilidad de obtener un valor mayor que z sea 0.1 .
3. Los errores en el peso proporcionado por la báscula de un ambulatorio son normales de media 0 y desviación 1 kg. Calcula la probabilidad de que la diferencia entre el peso real de un paciente y el proporcionado por la báscula no supere los 500 gr. (bien por exceso o bien por defecto).
4. La capacidad vital es la cantidad de aire que es posible expulsar de los pulmones después de haber inspirado completamente. Hemos calculado la capacidad vital estandarizada CVE en una población infantil (la CVE se calcula como la capacidad vital corregida adecuadamente mediante la media y desviación típica). Se asume que la capacidad vital estandarizada se distribuye como una normal $N(0, 1)$.
 - a) Si se considera que la salud pulmonar de un niño es débil cuando su capacidad vital estandarizada es menor que -1.5 , ¿qué porcentaje de la población estudiada presenta una salud pulmonar débil?
 - b) Un niño tiene un crecimiento pulmonar normal si su capacidad vital estandarizada está entre -1.5 y 1.5 . ¿Qué porcentaje de la población presenta un crecimiento pulmonar normal?
 - c) Completa las siguientes frases:
 - 1) Según el modelo, el 90 % de la población infantil tiene una capacidad vital estandarizada menor que aproximadamente _____.
 - 2) Según el modelo, el 20 % de la población infantil con mayor capacidad pulmonar estandarizada tiene una CVE mayor que aproximadamente _____.
5. Los valores de sodio sérico en adultos sanos se distribuye como una variable normal de media 141 mM y varianza 9 mM². Asumiendo dicha distribución:

- a) ¿Qué porcentaje de adultos tienen un nivel de sodio sérico inferior a 137mM?
 b) ¿Qué porcentaje de adultos tienen un nivel de sodio sérico de entre 137 y 145mM?
6. Dada una variable $X \in N(\mu, \sigma)$
- a) ¿Qué porcentaje del área bajo la función de densidad está contenida entre $\mu - \sigma$ y $\mu + \sigma$?
 b) ¿Qué porcentaje del área bajo la función de densidad está contenida entre $\mu - 2\sigma$ y $\mu + 2\sigma$?
 c) ¿Qué porcentaje del área bajo la función de densidad está contenida entre $\mu - 3\sigma$ y $\mu + 3\sigma$?
7. La presión arterial sistólica corresponde al valor máximo de la tensión arterial en sístole. Se asume que la presión sistólica se distribuye como una variable normal, donde el valor medio y la desviación típica dependen de la edad. Se muestra a continuación la media y desviación típica para 3 grupos de edad.

	Presión sistólica (mmHg)	
	Media	Desviación típica
20-24 años	123.9	13.74
40-44 años	129.0	15.07
55-64 años	139.8	19.99

- a) ¿Qué porcentaje de la población de entre 20 y 24 años presenta una presión sistólica inferior a 150 mmHg?
 b) ¿Qué porcentaje de la población de entre 40 y 44 años presenta una presión sistólica superior a 100 mmHg?
 c) ¿Qué porcentaje de la población de entre 55 y 64 años presenta una presión sistólica de entre 130 y 145 mmHg?
 d) Además, un modelo de hipertensión-hipotensión aceptado es el que se muestra a continuación. Calcula, para cada grupo de edad, los límites de presión sistólica que clasifican a un paciente como hipotenso, hipertenso, en riesgo o con presión sanguínea normal.



8. El nivel de colesterol en la sangre se mide de acuerdo a un índice llamado LDL. Para el caso de personas adultas, la distribución del colesterol en la sangre es aproximadamente normal y en el caso de los hombres tiene una media de 4.8 unidades LDL con una desviación estándar igual a 0.6 unidades. El nivel normal (o riesgo normal) de colesterol se considera aquel que

queda entre los límites $\mu \pm \sigma$ en unidades LDL. Una persona con más de $\mu + \sigma$ pero menos de $\mu + 2\sigma$ unidades LDL tiene un nivel de riesgo moderado. Si tiene un nivel de $\mu + 2\sigma$ o superior se considera de alto riesgo y se hace propenso a sufrir un infarto. Por otra parte, si el nivel de colesterol en la sangre de un adulto está por debajo de $\mu - \sigma$ unidades, se considera de riesgo bajo.

- a) ¿Cuáles son los porcentajes de población de hombres adultos que están incluidos en cada uno de los 4 niveles de riesgo descritos?
 - b) ¿A partir de qué nivel de colesterol se encuentra el 10 % de la población de hombres adultos con mayor riesgo?
9. Para ayudar a la evaluación del pronóstico de pacientes con una determinada enfermedad pulmonar se calculan dos índices, independientes entre sí. Se asume que el primero de los índices se distribuye según una normal $N(120, 10)$ y que el segundo se distribuye según una normal $N(15, 3)$. Se consideran susceptibles de una revisión más profunda aquellos pacientes que en el primer índice superen el valor 142. También son susceptibles de una revisión más profunda aquellos pacientes que en el segundo índice presenten un valor inferior a 8. ¿Qué porcentaje de pacientes son susceptibles de una revisión más profunda?
10. Una determinada prueba mide los niveles de las tres células sanguíneas básicas: glóbulos rojos, blancos y plaquetas. Se asume que el nivel de glóbulos blancos se distribuye según una normal de media 14 y desviación típica 3.6. Además una persona se clasifica en un grupo de riesgo de infección cuando su nivel de glóbulos blancos es inferior a 10.
- a) ¿Cuál es la probabilidad de que un paciente sea clasificado en el grupo de riesgo de infección?
 - b) Si se realiza la prueba en 9 pacientes de manera independiente:
 - 1) ¿Cuál es la probabilidad de que al menos 2 de ellos sean clasificados en el grupo de riesgo de infección?
 - 2) ¿Cuál es el número esperado de pacientes en el grupo de riesgo?
11. Un estudio realizado en un hospital de EEUU determinó que el peso al nacer se distribuye como una normal de media 109 onzas y desviación típica 13 onzas. Sabiendo que una onza equivale a 28.35 gramos:
- a) ¿Cuál es el peso medio al nacer en gramos?
 - b) Si X denota el peso al nacer en gramos. ¿Cuál es la varianza de X ?
 - c) Si Y denota el peso al nacer en kilos, ¿Cuál es la desviación típica de Y ?
 - d) ¿Cuál es la probabilidad de que un niño al nacer peso menos de 3200 gramos?
12. El coeficiente intelectual de una población sigue una distribución normal de media 100 y desviación típica 15. ¿Cuál de las siguientes afirmaciones es correcta?
- a) El 95 % central de los individuos de la población estarán situados entre 85 y 115.
 - b) El 95 % central de los individuos de la población estarán situados entre 50 y 150.
 - c) El 95 % central de los individuos de la población estarán situados entre 70 y 130.

13. Queremos estudiar la capacidad diagnóstica de una prueba de glucemia. En base a la experiencia se determina que el nivel de glucosa en sangre de pacientes sanos se distribuye como una normal de media 80 mg/dl y desviación típica 10 mg/dl. También se sabe que el nivel de glucosa en sangre de pacientes diabéticos se distribuye como una normal de media 160 mg/dl y desviación típica 31.4 mg/dl.
- Si la prueba de glucemia establece que un individuo está enfermo cuando su nivel de glucosa es superior a 100 mg/dl, ¿Cuál es la sensibilidad y especificidad de la prueba diagnóstica?
 - ¿Cuál es la sensibilidad y especificidad de la prueba diagnóstica si el punto de corte se establece en 90 mg/dl.?
14. El nivel de hemoglobina en hombres sigue una distribución normal de media 15.5 g/dL y desviación típica 1.25 g/dL. El nivel de hemoglobina en mujeres sigue una distribución normal de media 14 g/dL. Se sabe además que el 90 % de las mujeres tienen un nivel de hemoglobina superior a 12.7 g/dL. Calcula:
- El porcentaje de hombres con un nivel de hemoglobina entre 13.8 g/dL y 16 g/dL.
 - El porcentaje de mujeres con un nivel de hemoglobina inferior a 15.75 g/dL.
 - El 60 % de los habitantes de una población son mujeres. Si se elige una persona al azar, ¿cuál es la probabilidad de que su nivel de hemoglobina sea inferior a 14 g/dL?

Soluciones

- $P(-1 \leq Z \leq 1) = 0.6826$
 - $P(-2 \leq Z \leq 2) = 0.9545$
 - $P(-3 \leq Z \leq 3) = 0.9973$
- 0.4115
 - 0.97725
 - 0.687475
 - Aproximadamente 0.53
 - Aproximadamente 1.28
- La probabilidad es 0.383
- 6.68 %
 - 86.66 %
 - 1) 1.28
2) 0.84
- 9.17 %
 - 81.64 %
- 0.6826
 - 0.9545
 - 0.9973
- 97.06 %
 - 97.25 %
 - 29.05 %

- d) Los límites de presión sistólica que clasifican a un paciente como hipotenso, hipertenso, en riesgo o con presión sanguínea normal son:

	Presión sistólica (mmHg)				
	Hipotenso	Riesgo	Normal	Riesgo	Hipertenso
20-24 años	< 101.3	101.3-106.3	106.3-141.4	141.4-146.4	> 146.4
40-44 años	< 104.2	104.2-109.7	109.7-148.2	148.2-153.7	> 153.7
55-64 años	< 107.0	107.0-114.2	114.2-165.3	165.3-172.5	> 172.5

8. a) Nivel normal: 68.26 %, riesgo moderado: 13.59 %, alto riesgo: 2.27 %, riesgo bajo: 15.85 %
 b) A partir de 5.56 unidades de LDL
9. El 2.36 % de los pacientes son susceptibles de una revisión más profunda
10. a) 0.1335
 b) 1) 0.3427
 2) $\mu = 1.2015$
11. a) $\mu = 3090.15\text{gr}$
 b) $\sigma^2 = 135829.1\text{gr}^2$
 c) $\sigma = 0.36855\text{kg}$
 d) 0.6141
12. La afirmación c) es la correcta
13. a) La sensibilidad es 0.97193 y la especificidad es 0.97725
 b) La sensibilidad es 0.98679 y la especificidad es 0.8413
14. a) 56.85 %
 b) 95.72 %
 c) 0.346

Capítulos 5 y 6: Estimación puntual e Intervalos de confianza

1. El peso al nacer recogido durante un largo período de tiempo en un hospital muestra una media de 3100 gramos y una desviación típica de 370 gramos.
 - a) Calcular la probabilidad de que, en una muestra de 25 niños, el peso medio al nacer esté entre 2950 y 3120 gramos.
 - b) ¿Y si la muestra es de 50 niños?
2. M&M's determina la distribución de los colores de sus bolitas de chocolate en base a estudios de mercado sobre las preferencias de los consumidores. Así, en la producción de M&M's la proporción de bolitas de color azul es del 24 %, la de naranjas es del 20 %, la de verdes es del 16 %, la de rojas es del 13 % y la de marrones es del 13 %.
 - a) En una muestra aleatoria de 100 M&M's, ¿cuál es la proporción esperada de bolitas de color azul? ¿qué probabilidad tienes de que toquen entre un 20 % y un 25 % de bolitas azules? ¿qué probabilidad tienes de que toquen más de un 25 % de bolitas azules?
 - b) En una muestra aleatoria de 1000 M&M's, ¿cuál es la proporción esperada de bolitas de color azul? ¿qué probabilidad tienes de que toquen entre un 20 % y un 25 % de bolitas azules? ¿qué probabilidad tienes de que toquen más de un 25 % de bolitas azules?
3. Supongamos que la proporción real de fumadores en una determinada comunidad se conoce y es igual a 0.4. Si queremos estimar dicha proporción de fumadores a partir de una muestra de tamaño $n = 100$, ¿cuál es la probabilidad de que la proporción estimada sea correcta salvo un error de $\pm 3\%$? ¿Y si se realiza la estimación con una muestra de tamaño $n = 200$?
4. Cierta empresa se ha propuesto comercializar un aparato para analizar la concentración de glucosa en sangre. Los fabricantes son conocedores de que su método presenta un error de medición cuya desviación típica es de 2.4 mg/l. Sin embargo, dado que desconocen la media, se han decidido a tomar una muestra que les permita estimarla. A continuación consta tal muestra de los errores de medición (en mg/l):

0.51	-2.75	1.83	2.97	-0.82	2.32	-0.69	-2.19
1.47	-1.54	0.30	-1.25	0.18	-0.21	-1.95	-3.67

Elabora una estimación del error medio y construye un intervalo de confianza a un nivel del 99 %, suponiendo que los errores siguen una distribución normal.

5. Una empresa de tecnología que elabora productos para el cuidado de la salud comercializa electrodos adhesivos redondos. Estamos interesados en determinar el diámetro medio de dichos electrodos. Se sabe que el proceso de producción sigue una distribución normal y padece una desviación típica de 0.1 cm. Construye un intervalo de confianza para el diámetro medio de los electrodos al 95 % utilizando que en una muestra de 25 electrodos fabricados por la empresa el diámetro medio fue de 3.5 cm.

6. La exostosis auditiva externa (EAE) es una anomalía ósea del canal auditivo externo. Esta lesión está asociada a una prolongada inmersión en agua fría y aparece con frecuencia en individuos que participan en actividades acuáticas. Un estudio publicado en una revista especializada en Otorrinolaringología pretende determinar la prevalencia de EAE en una población de surfistas. Para ello se sometió a 307 surfistas profesionales a un cuestionario.

- a) De los 307 surfistas encuestados, 132 afirmaron haber necesitado tratamiento médico para infecciones de oído en alguna ocasión. ¿Cómo estimarías la proporción de surfistas que sufren infecciones de oído en base a la muestra? Calcula el intervalo de confianza para la proporción de surfistas que sufren algún tipo de infección de oído con un nivel de confianza del 95 %. Calcula el intervalo de confianza para la proporción de surfistas que sufren algún tipo de infección de oído con un nivel de confianza del 90 %.
- b) Los surfistas encuestados en este estudio surfean fundamentalmente en aguas frías (por debajo de 12°C). Se cree que la prevalencia de EAE es distinta en surfistas de aguas templadas. Supongamos que de los 307 surfistas examinados 230 fueron diagnosticados de EAE y que en otro estudio realizado a 75 surfistas de aguas templadas, 30 fueron diagnosticados de EAE. Construye un intervalo de confianza al 90 % para la diferencia de prevalencias de EAE entre surfistas de aguas frías y surfistas de aguas templadas.

Los datos del apartado a) están tomados del artículo "Prevalence of external auditory canal exostoses in surfers". Arch Otolaryngol Head Neck Surg. 1999

7. Se pretende conocer la media y la varianza del tiempo de eliminación de un medicamento. Para ello se han observado los tiempos en una muestra de pacientes obteniéndose los siguientes datos (en horas):

5.64	7.83	6.92	5.31	8.85	7.94	6.04	5.19
7.33	8.24	7.68	6.47	6.09	8.75	5.87	7.28

Supón que los datos proceden de una distribución normal y, en base a ello, confecciona estimaciones para la media y la varianza. Calcula el intervalo de confianza a un nivel del 90 % para el tiempo medio de eliminación del medicamento.

8. Para estudiar si la presión ejercida en la parte superior del brazo aumenta o no el tiempo de hemorragia, 29 personas fueron sometidas a una presión de 40 mmHg y a continuación se les practicó una punción digital obteniéndose un tiempo medio de hemorragia de 2.192 minutos con una desviación estándar de 0.765 minutos. Otras 33 personas actuaron como controles, no se les aplicó presión y su tiempo medio de hemorragia al pincharles un dedo fue de 1.407 minutos con una desviación estándar de 0.588 minutos. Determina con un nivel de confianza de 95 % un intervalo de confianza para la diferencia de los tiempos medios de hemorragia entre los tratados y los controles. Se asume que los tiempos de hemorragia en ambos grupos son normales y con la misma varianza.
9. En un estudio sobre nutrición se analizó la ganancia de peso de 16 niños sometidos a una dieta especial durante un periodo de 3 meses. Se observó una ganancia media de peso 2.49 kg. Un grupo de control formado por 16 niños de constitución física similar fue sometido a una dieta normal durante el mismo periodo de tiempo, observándose una ganancia media de peso de 2.05 kg. Se supone que la desviación estándar para la ganancia de peso es 2 kg.

- a) Determina con un nivel de confianza de 95 % un intervalo de confianza para la diferencia en la ganancia media de peso entre niños tratados y los controles.
- b) ¿Cuál sería el intervalo de confianza si suponemos que tanto el grupo control como el de tratamiento estaba formado por 50 niños? Compáralo con el intervalo calculado en el apartado anterior.
10. Un investigador está planeando hacer un estudio sobre el nivel medio de presión sistólica en pacientes con hipertensión. Algunos resultados previos indican que la presión sistólica es aproximadamente normal con una desviación típica de 15 mmHg.
- a) Si el investigador desea obtener un intervalo de confianza para el nivel medio de presión sistólica de longitud 4 mmHg con una confianza de 95 %, ¿cuántos pacientes hipertensos tendrían que ser incluidos en el estudio?
- b) Si el investigador decide mantener el nivel de confianza en el 95 % pero desea que el intervalo obtenido para el nivel medio de presión sistólica sea más pequeño, ¿tendrá que aumentar o disminuir el tamaño de la muestra? Calcula el número de pacientes que debe considerar para tener un intervalo de longitud 3mmHg con confianza 95 %.
- c) ¿Cuál será la longitud del intervalo de confianza al 95 % para el nivel medio de presión sistólica si hace el estudio sobre 100 pacientes? ¿Qué pasará con la longitud del intervalo si reduce el estudio a 50 pacientes manteniendo el nivel de confianza? ¿Y si hace el estudio sobre 100 pacientes pero construye el intervalo de confianza al 99 %?
11. Según reconoce Sanidad, es cada vez más elevada la prevalencia de pacientes polimedicados (que toman 6 o más medicamentos) en el nivel asistencial. Esto hace necesario reforzar las estrategias para optimizar los recursos. Un centro de salud ha llevado a cabo un estudio para conocer la prevalencia de polimedicados. Se han seleccionado 649 pacientes de los cuales 149 están sometidos a tratamientos que superan los 6 medicamentos. Calcula un intervalo de confianza para la prevalencia de polimedicados con un nivel de confianza del 95 %.
12. Para estudiar el efecto del ejercicio físico sobre el nivel de triglicérido, se ha realizado el siguiente experimento con 11 individuos: previo al ejercicio se tomaron muestras de sangre para determinar el nivel de triglicérido por 100 mililitros de sangre de cada sujeto. Después los individuos fueron sometidos a un programa de ejercicios que se centraba diariamente en carreras y marchas. Al final del periodo de ejercicios, se tomaron nuevamente muestras de sangre y se obtuvo una segunda lectura del nivel de triglicérido. De este modo, se dispone de dos conjuntos de observaciones del nivel de triglicérido por 100 mililitros de sangre de los sujetos:

Sujeto	1	2	3	4	5	6	7	8	9	10	11
Previo	68	77	94	73	37	131	77	24	99	629	116
Posterior	95	90	86	58	47	121	136	65	131	630	104

Suponiendo normalidad en el nivel de triglicérido, construye un intervalo de confianza de nivel 95 % para la diferencia entre el nivel medio de triglicérido antes y después del programa de ejercicios.

Soluciones

- a) La probabilidad de que el peso medio al nacer calculado a partir de una muestra de 25 niños esté entre 2950 y 3120 gramos es $P(2950 < \bar{X} < 3120) = 0.58471$.
 - b) Si $n = 50$, $P(2950 < \bar{X} < 3120) = 0.6458$.
- a) Para $n = 100$, $\mathbb{E}(\hat{p}) = 0.24$, $P(0.2 < \hat{p} < 0.25) = 0.418$ y $P(\hat{p} > 0.25) = 0.407$.
 - b) Para $n = 1000$, $\mathbb{E}(\hat{p}) = 0.24$, $P(0.2 < \hat{p} < 0.25) = 0.768$ y $P(\hat{p} > 0.25) = 0.229$.
- Para $n = 100$, la probabilidad de que la proporción estimada sea correcta salvo un error de $\pm 3\%$ es 0.4582. Si $n = 200$, la probabilidad es 0.6212.
- El IC para μ al 99 % es $(-1.8911, 1.2048)$.
- El IC para μ al 95 % es $(3.4608, 3.5392)$.
- a) $\hat{p} = 0.429$. El IC para p al 95 % es $(0.3745, 0.4853)$. El IC para p al 90 % es $(0.3836, 0.4763)$.
 - b) Siendo p_1 la prevalencia de EAE en surfistas de aguas frías y p_2 la prevalencia de EAE en surfistas de aguas templadas, el IC para $p_1 - p_2$ al 90 % es $(0.2479, 0.4504)$.
- $\bar{x} = 6.9643$, $s^2 = 1.4346$. El IC para μ al 90 % es $(6.4403, 7.4884)$.
- Siendo μ_1 el tiempo medio de hemorragia de los tratados y μ_2 el tiempo medio de hemorragia de los controles, el IC para $\mu_1 - \mu_2$ al 95 % es $(0.4406, 1.1293)$.
- a) El IC para la diferencia en la ganancia media de peso entre niños tratados y los controles al 95 % es $(-0.945, 1.8259)$.
 - b) Para $n = 50$, el IC para la diferencia en la ganancia media de peso entre niños tratados y los controles al 95 % es $(-0.344, 1.224)$.
- a) Debe incluir $n = 217$ pacientes.
 - b) Debe incluir $n = 385$ pacientes.
 - c) La longitud del intervalo de confianza al 95 % si se hace el estudio sobre 100 pacientes sería 5.88. Si se reduce el estudio a 50 pacientes la longitud sería 8.31. Si se hace el estudio sobre 100 pacientes pero se construye el intervalo de confianza al 99 %, la longitud sería 7.74.
- El IC para p al 95 % es $(0.1972, 0.26194)$.
- El IC de nivel 95 % para la diferencia entre el nivel medio de triglicérido antes y después del programa de ejercicios es $(-28.996, 3.9059)$.

Capítulos 7 y 8: Contrastes de hipótesis

1. Según fuentes estadísticas, en la actualidad la edad media de las madres primerizas en España es de 29.3 años.

a) Tomamos una muestra de 36 madres primerizas gallegas y observamos que la edad media de dichas mujeres es 30.5 años. Asumimos que la edad de las madres primerizas en Galicia sigue una distribución normal con una desviación típica de 2 años. Para una significación del 5 %, ¿podemos concluir que la edad media de las madres primerizas en Galicia difiere de la de España?

b) Se considera ahora una muestra de 10 madres primerizas de Portugal. Sus edades son:

30 28 27 28 28 28 24 23 31 30

Asumimos que la edad de las madres primerizas en Portugal también sigue una distribución normal con una desviación típica de 2 años.

- 1) Para una significación del 5 %, ¿podemos concluir que la edad media de las madres primerizas en Portugal difiere de la de España?
- 2) Calcula el p -valor del contraste.
- 3) Para una significación del 1 %, ¿podemos concluir que la edad media de las madres primerizas en Portugal difiere de la de España?

2. Según datos de 2003, el 62.68 % de los jóvenes españoles de entre 18 y 29 años afirman utilizar preservativo siempre que mantienen relaciones sexuales con parejas ocasionales. Tras una campaña preventiva sobre el uso del preservativo llevada a cabo en los últimos años, se realizó encuesta a 3150 jóvenes de entre 18 y 29 años. De ellos, 2047 afirmaron utilizar preservativo siempre que mantienen relaciones sexuales con parejas ocasionales. ¿Se puede concluir que la campaña preventiva ha sido efectiva para una significación del 5 %?

Puedes encontrar datos sobre salud en España en la web del Instituto Nacional de Estadística:

http://www.ine.es/inebmenu/mnu_salud.htm

3. Cierta empresa se ha propuesto comercializar un aparato para analizar la concentración de glucosa en sangre. Los fabricantes son conocedores de que su método presenta un error de medición cuya desviación típica es de 2.4 mg/l. Sin embargo, dado que desconocen la media, se han decidido a tomar una muestra que les permita estimarla. A continuación consta tal muestra de los errores de medición (en mg/l):

0.51 -2.75 1.83 2.97 -0.82 2.32 -0.69 -2.19
1.47 -1.54 0.30 -1.25 0.18 -0.21 -1.95 -3.67.

a) ¿Es el error medio significativamente distinto de cero para una significación 0.1? ¿Cómo contestarías a la pregunta utilizando el intervalo de confianza construido en el boletín 6?

- b) ¿Es el error medio significativamente distinto de cero para una significación 0.05?
- c) Supongamos ahora que las observaciones provienen de un aparato cuyo error de medición presenta una desviación típica de 0.67 mg/l. ¿Dirías ahora que el error medio es significativamente distinto de cero para una significación 0.05? Calcula e interpreta el p-valor.
4. Para conocer el uso que hombres y mujeres hacen de los servicios sanitarios es necesario realizar estudios que permitan conocer mejor los factores que intervienen en sus decisiones y en las del personal sanitario. Según la Encuesta Nacional de Salud de 2003, un 17 % de las mujeres acuden a consulta médica con frecuencia. Se lleva a cabo un estudio en el que participan 2150 hombres. Del total de los hombres, 275 afirman haber acudido a consulta médica durante las dos semanas anteriores al momento de la encuesta. ¿Se puede concluir que el uso de los servicios sanitarios por parte de los hombres es menor que el de las mujeres?

En la Encuesta Nacional de Salud se analizan las diferencias entre hombres y mujeres en el estado de salud o en los hábitos de consulta. Algunos estudios apuntan, para explicar el diferente uso de los servicios, a una mayor medicalización de la salud de las mujeres o la práctica más frecuente de conductas de riesgo por parte de los hombres, sobre todo en edades tempranas.

http://www.msps.es/organizacion/sns/planCalidadSNS/e02_t05.htm

5. Para estudiar si la presión ejercida en la parte superior del brazo aumenta o no el tiempo de hemorragia, 29 personas fueron sometidas a una presión de 40 mmHg y a continuación se les practicó una punción digital obteniéndose un tiempo medio de hemorragia de 2.192 minutos con una desviación estándar de 0.765 minutos. Otras 33 personas actuaron como controles, no se les aplicó presión y su tiempo medio de hemorragia al pincharles un dedo fue de 1.407 minutos con una desviación estándar de 0.588 minutos. Se asume que los tiempos de hemorragia en ambos grupos son normales y con la misma varianza.
- a) ¿Se puede concluir que el tiempo medio de hemorragia es significativamente distinto al ejercer presión en la parte superior del brazo que al no ejercer presión? (significación 0.05).
- b) ¿Se puede concluir que el tiempo medio de hemorragia es significativamente mayor al ejercer presión en la parte superior del brazo? (significación 0.05).
6. Para estudiar el efecto del ejercicio físico sobre el nivel de triglicérido, se ha realizado el siguiente experimento con 11 individuos: previo al ejercicio se tomaron muestras de sangre para determinar el nivel de triglicérido por 100 mililitros de sangre de cada sujeto. Después los individuos fueron sometidos a un programa de ejercicios que se centraba diariamente en carreras y marchas. Al final del periodo de ejercicios, se tomaron nuevamente muestras de sangre y se obtuvo una segunda lectura del nivel de triglicérido. De este modo, se dispone de dos conjuntos de observaciones del nivel de triglicérido por 100 mililitros de sangre de los sujetos:

Sujeto	1	2	3	4	5	6	7	8	9	10	11
Previo	68	77	94	73	37	131	77	24	99	629	116
Posterior	95	90	86	58	47	121	136	65	131	630	104

Suponiendo normalidad en el nivel de triglicérido, ¿hay pruebas suficientes para afirmar que el ejercicio físico produce cambios en el nivel de triglicérido?

7. En un estudio sobre nutrición se analizó la ganancia de peso de 16 niños sometidos a una dieta especial durante un periodo de 3 meses. Se observó una ganancia media de peso 3.05 kg. Un

grupo de control formado por 16 niños de constitución física similar fue sometido a una dieta normal durante el mismo periodo de tiempo, observándose una ganancia media de peso de 2.05 kg. Se supone que la desviación estándar para la ganancia de peso es 2 kg.

- a) ¿Se puede concluir que la ganancia media de peso es significativamente mayor en los niños sometidos a la dieta especial? (significación 0.05)
- b) Calcula el p -valor del contraste.

8. Se trata de estudiar el efecto de un tratamiento dirigido a elevar el colesterol HDL. Para ello se ha medido el colesterol HDL de 10 pacientes. A continuación se les ha sometido al tratamiento y se ha vuelto a medir el colesterol HDL. Los datos se muestran a continuación. Determinar si

Caso Nº	HDL pre-tratamiento	HDL post-tratamiento
1	81	85
2	37	38
3	35	37
4	64	72
5	46	51
6	37	45
7	45	38
8	43	58
9	21	25
10	51	61

hay suficiente evidencia estadística, a nivel 0.01, para afirmar que el tratamiento es efectivo.

9. Una compañía farmacéutica afirma que cierto medicamento elimina el dolor de cabeza en un cuarto de hora en el 90 % de los casos. Tomada una muestra de 200 pacientes a los que se les administró el medicamento, se observó la desaparición del dolor en 170 de ellos. Contrastar la hipótesis de la compañía para un nivel de significación del 5 %.
10. Registramos los niveles en plasma de determinado ácido graso en 30 pacientes de Retinitis Pigmentosa (RP) y en 32 voluntarios sanos (S), y los resultados fueron los siguientes:

RP	$n = 30$	Media =35.8	Desviación típica=20.5
S	$n = 32$	Media =45.8	Desviación típica=30.1

- a) Suponiendo que las poblaciones son normales y a la vista de los resultados obtenidos, ¿podemos concluir que la media es significativamente más baja en los pacientes de RP?
 - b) Estimar mediante un intervalo de confianza del 95 % el valor medio en personas sanas.
11. A un grupo de 10 enfermos se les suministra un antidepresivo. Mediante pruebas adecuadas se valora en 4 el valor inicial de ese tipo de enfermos. Después de la administración del medicamento, el estado del paciente tuvo las siguientes puntuaciones:

3 5 4.5 7 6 6.5 4 5.5 7 7

A la vista de los datos, ¿puede decirse que los enfermos han mejorado significativamente? (Existe mejoría si la puntuación es mayor de 4. Utilizar un nivel de significación de 0.01.)

Soluciones

1. a) El valor del estadístico de contraste es 3.6. Rechazamos H_0 , es decir, existen diferencias significativas.
b) 1) El valor del estadístico de contraste es -2.53 . Rechazamos H_0 .
2) El p -valor del contraste es 0.0114.
3) No. Para una significación del 1%, no rechazamos H_0 .
2. El valor del estadístico de contraste es 2.6737. Rechazamos H_0 , es decir, hay evidencia estadística de que la campaña es efectiva.
3. a) El valor del estadístico de contraste es -0.5718 . No rechazamos H_0 .
El IC para μ al 90% es $(-1.32, 0.64)$. Como $\mu_0 = 0$ pertenece al IC, no rechazamos H_0 .
b) No.
c) El valor del estadístico de contraste es -2.04 . Para una significación del 5%, rechazamos H_0 . El p -valor es 0.0413 (por lo tanto, para un nivel de significación mayor que 0.0413, rechazamos H_0).
4. El valor del estadístico de contraste es -5.19 . Para una significación del 5%, rechazamos H_0 , es decir, hay evidencia estadística de que el uso de los servicios sanitarios por parte de los hombres es menor que el de las mujeres.
5. a) El valor del estadístico de contraste es 4.55. Para una significación del 5%, rechazamos H_0 .
b) El valor del estadístico de contraste es 4.55. Para una significación del 5%, rechazamos H_0 , es decir, el tiempo medio de hemorragia es significativamente mayor al ejercer presión en la parte superior del brazo.
6. El valor del estadístico de contraste es -1.7005 . Para una significación del 5%, no rechazamos H_0 , es decir, no hay evidencia de que el ejercicio afecte al nivel de triglicéridos.
7. a) El valor del estadístico de contraste es 1.4142. Para una significación del 5%, no rechazamos H_0 , es decir, no podemos concluir que la ganancia media de peso es significativamente mayor en los niños sometidos a la dieta especial.
b) El p -valor del contraste es 0.078.
8. El valor del estadístico de contraste es 2.68. No rechazamos H_0 (no hay suficiente evidencia a nivel 0.01 para afirmar que el tratamiento es efectivo).
9. El valor del estadístico de contraste es -2.35 . Para una significación del 5%, no rechazamos H_0 (no hay evidencia estadística de que la afirmación de la compañía es cierta).
10. a) El valor del estadístico de contraste es -1.518 . No rechazamos H_0 para una significación del 5%.
b) El intervalo de confianza del 95% para el valor medio en personas sanas es $(34.94, 56.65)$.
11. El valor del estadístico de contraste es 3.5. Para una significación del 1%, rechazamos H_0 , es decir, los enfermos han mejorado significativamente.

Capítulo 9: Contrastes para datos categóricos

1. La siguiente tabla muestra la clasificación de 1343 niños según el grado de cumplimiento de su calendario vacunal y el nivel socio-cultural de sus padres. Determina si existe una asociación significativa entre el grado de cumplimiento del calendario vacunal de los niños y el nivel socio-cultural de sus padres.

	Cumplimiento calendario vacunal		
Nivel socio-cultural	Bajo	Medio	Alto
Bajo	114	229	228
Medio bajo	7	134	277
Medio alto	7	63	150
Alto	2	38	94

2. Para evaluar el efecto de la exposición a asbesto sobre el riesgo de fallecer por cáncer de pulmón, un estudio comparó un grupo de 6.245 trabajadores expuestos a este agente con otro grupo de 7.895 trabajadores sin exposición a este factor. A lo largo de 22 años de seguimiento, en el primer grupo se presentaron 76 defunciones por cáncer en el aparato respiratorio, en tanto que en el grupo no expuesto el número de defunciones por esta causa fue 28. Construye la tabla de contingencia correspondiente y determina si existe una asociación significativa entre la exposición a asbesto y el riesgo de fallecer por cáncer de pulmón.

El asbesto es un grupo de minerales naturales fibrosos. Se ha venido utilizando en el aislamiento de los edificios, como componente de diversos productos (tejas, tuberías de agua, mantas ignífugas y envases médicos), como aditivo de los plásticos y en la industria automovilística.

<http://www.who.int/mediacentre/factsheets/fs343/es/index.html>

3. Un estudio transversal para conocer la prevalencia de osteoporosis y su relación con algunos factores de riesgo potenciales (ver web de Investigación e Innovación Sanitaria de la Consellería de Sanidade de la Xunta de Galicia) incluyó a 400 mujeres con edades entre 50 y 54 años. A cada una se le realizó una densitometría de columna y en cada caso se completó un cuestionario de antecedentes. Se pretende determinar si existe una asociación significativa entre la prevalencia de osteoporosis y antecedentes de dieta pobre en calcio. De las 80 pacientes que presentaban osteoporosis 58 presentaban antecedentes de dieta pobre en calcio, en tanto que entre las 320 que no tenían osteoporosis, el número de mujeres con este antecedente era de 62.
 - a) Construye la tabla de contingencia correspondiente y determina, para una nivel de significación del 1 %, si existe una asociación significativa entre la prevalencia de osteoporosis y antecedentes de dieta pobre en calcio.
 - b) Calcula el estadístico Chi-cuadrado corregido (corrección de Yates) y determina en base a ese estadístico si, para un nivel de significación del 5 %, existe una asociación significativa entre la prevalencia de osteoporosis y antecedentes de dieta pobre en calcio.

4. Supongamos que se quiere estudiar la posible asociación entre el hecho de que una gestante fume durante el embarazo y que el niño presente bajo peso al nacer. Para responder a esta pregunta se realiza un estudio de seguimiento sobre una cohorte de 2000 gestantes, a las que se interroga sobre su hábito tabáquico durante la gestación y se determina además el peso del recién nacido. Los resultados de este estudio se muestran en la siguiente tabla:

Gestante	Recién nacido de bajo peso	
	Sí	No
Fumadora	43	204
No fumadora	105	1645

- ¿Se puede concluir que existe una relación estadísticamente significativa entre el hecho de que una gestante fume durante el embarazo y que el niño presente bajo peso al nacer?
- Determina el riesgo relativo.
- Calcula el estadístico Chi-cuadrado corregido (corrección de Yates) y determina si existe una relación estadísticamente significativa entre el hecho de que una gestante fume durante el embarazo y que el niño presente bajo peso al nacer.

Ejemplo tomado de <http://www.fisterra.com/mbe/investiga/chi/chi.asp#ji>

5. En un estudio sobre VIH se pretende determinar si existe asociación significativa entre la edad del paciente y el nivel de linfocitos CD4. Para ello se determina el nivel de linfocitos CD4 (<200, 200-500, >500) en pacientes de 3 grupos de edad. ¿Se puede concluir que existe una relación estadísticamente significativa entre el nivel de linfocitos y la edad del paciente?

Nivel de linfocitos	Edad		
	≤ 30 años	31 – 40 años	≥ 41 años
<200	6	30	6
200-500	20	72	21
>500	19	49	12

6. Se quiere estudiar la posible asociación entre la presencia de infección postoperatoria (IPO) y la diabetes (DIAB) en una población de operados. En una muestra de 1337 personas de edad < 65 años y en otra de 892 de edad ≥ 65 años se obtuvieron los siguientes resultados. ¿Existe asociación significativa entre IPO y diabetes en cada grupo de edad?

DIAB	< 65 años	
	IPO	NO IPO
Sí	15	29
No	190	1103

DIAB	≥ 65 años	
	IPO	NO IPO
Sí	28	65
No	215	584

7. Se realizó un estudio de seguimiento para detectar la posible asociación entre enfermedades cardiovasculares y el exceso de peso. Se eligieron 1990 hombres con edades entre 55 y 59 años de estatura similar. Tras 5 años de seguimiento se observaron los datos resumidos en la tabla. ¿Se puede admitir que el exceso de peso se asocia con el infarto de miocardio?

Infarto	Peso				
	55 – 64 kg.	65 – 74 kg.	75 – 84 kg.	85 – 94 kg.	> 95 kg.
Sí	8	18	48	93	23
No	290	680	550	205	75

Soluciones

1. Se rechaza la hipótesis de independencia al 5% (hay relación significativa).

Nivel		Cumplimiento			
		Bajo	Medio	Alto	
Bajo	Observed	114	229	228	571
	Expected	55,27	197,28	318,45	
	Cell Chi-Sq	62,40	5,10	25,69	
Mediobajo	Obs	7	134	277	418
	Expected	40,46	144,42	233,12	
	Cell Chi-Sq	27,67	0,75	8,26	
Medioalto	Obs	7	63	150	220
	Expected	21,30	76,01	122,70	
	Cell Chi-Sq	9,60	2,23	6,08	
Alto	Observed	2	38	94	134
	Expected	12,97	46,30	74,73	
	Cell Chi-Sq	9,28	1,49	4,97	
		130	464	749	1343
Overall Chi-Square		163,51			
P-value		0,0000			
Degrees of Freedom		6			

2. Se rechaza la hipótesis de independencia al 5% (hay relación significativa).

Abesto		Cancer		
		Si	No	
Si	Observed	76	6169	6245
	Expected	45,93	6199,07	
	Cell Chi-Sq	19,68	0,15	
No	Observed	28	7867	7895
	Expected	58,07	7836,93	
	Cell Chi-Sq	15,57	0,12	
		104	14036	14140
Overall Chi-Square		35,51		
P-value		0,0000		
Degrees of Freedom		1		

3. a) El valor del estadístico Chi-cuadrado es 86.01. Se rechaza la hipótesis de independencia al 1% (hay relación significativa)
- b) La corrección de Yates es 83.5. Se rechaza la hipótesis de independencia al 5% (hay relación significativa)

4. Los valores observados y esperados son:

Fumador		Bajopeso		
		Si	No	
Si	Observed	43	204	247
	Expected	18,31	228,69	
	Cell Chi-Sq	33,31	2,67	
No	Observed	105	1645	1750
	Expected	129,69	1620,31	
	Cell Chi-Sq	4,70	0,38	
		148	1849	1997
Overall Chi-Square		41,06		
P-value		0,0000		
Degrees of Freedom		1		

a) Se rechaza la hipótesis de independencia al 5% (hay relación significativa)

b) El riesgo relativo es

$$RR = \frac{43/247}{105/1750} = 2.9$$

Las mujeres expuestas al tabaco durante la gestación tienen 2.9 veces más probabilidades de tener niños de bajo peso que las no expuestas al tabaco durante la gestación

c) La corrección de Yates es 39.41 (Se rechaza la hipótesis de independencia al 5%)

5. No se rechaza la hipótesis de independencia al 5%

Linfocito		Edad			
		<30	31-40	>41	
<200	Observed	6	30	6	42
	Expected	8,04	26,99	6,97	
	Cell Chi-Sq	0,52	0,34	0,14	
200-500	Obs	20	72	21	113
	Expected	21,64	72,61	18,75	
	Cell Chi-Sq	0,12	0,01	0,27	
>500	Observed	19	49	12	80
	Expected	15,32	51,40	13,28	
	Cell Chi-Sq	0,88	0,11	0,12	
		45	151	39	235
Overall Chi-Square		2,51			
P-value		0,6432			
Degrees of Freedom		4			

6. Para menores de 65 años se rechaza la hipótesis de independencia al 5% (hay relación significativa)

		IPO		
DIAB		Si	No	
Si	Observed	15	29	44
	Expected	6,75	37,25	
	Cell Chi-Sq	10,10	1,83	
No	Observed	190	1103	1293
	Expected	198,25	1094,75	
	Cell Chi-Sq	0,34	0,06	
		205	1132	1337

Overall Chi-Square 12,33
P-value 0,0004
Degrees of Freedom 1

Para mayores de 65 años no se rechaza la hipótesis de independencia al 5%.

		IPO		
DIAB		Si	No	
Si	Observed	28	65	93
	Expected	25,34	67,66	
	Cell Chi-Sq	0,28	0,10	
No	Observed	215	584	799
	Expected	217,66	581,34	
	Cell Chi-Sq	0,03	0,01	
		243	649	892

Overall Chi-Square 0,43
P-value 0,5120
Degrees of Freedom 1

7. Se rechaza la hipótesis de independencia al 5%.

		Peso					
Infarto		55-64	65-74	75-84	85-94	>95	
Si	Observed	8	18	48	93	23	190
	Expected	28,45	66,64	57,10	28,45	9,36	
	Cell Chi-Sq	14,70	35,50	1,45	146,44	19,89	
No	Observed	290	680	550	205	75	1800
	Expected	269,55	631,36	540,90	269,55	88,64	
	Cell Chi-Sq	1,55	3,75	0,15	15,46	2,10	
		298	698	598	298	98	1990

Overall Chi-Square 240,99
P-value 0,0000
Degrees of Freedom 4

Capítulo 10: Regresión y correlación

1. Se lleva a cabo un estudio, por medio de detectores radioactivos, sobre la capacidad corporal para absorber hierro y plomo. En el estudio participaron 6 personas y después de 10 días se obtuvieron los siguientes resultados.

Hierro	1.7	2.2	3.5	4.3	8.0	6.0
Plomo	2.1	3.0	1.8	2.5	4.2	4.0

- Representa el diagrama de dispersión de los datos. ¿Te parece adecuado considerar un modelo de regresión lineal para explicar el valor del plomo en función del hierro?
 - Calcula y representa la recta de regresión del valor del plomo sobre el valor del hierro.
 - ¿Cuál es el coeficiente de correlación lineal?
 - ¿Qué valor de plomo cabe esperar para una persona con un nivel de hierro igual a 2.2?
 - Calcula el porcentaje de explicación de la recta.
2. Para tener valores comparables del gasto cardíaco entre distintos sujetos se utiliza un determinado índice cardíaco. Se ha medido dicho índice cardíaco (Y) en 7 pacientes de diferentes edades.

$X = \text{Edad}$	15	20	30	40	50	60	70
$Y = \text{Índice cardíaco}$	6.5	5.6	5.4	6	4.6	1.4	0.1

- Calcula la recta de regresión de Y sobre X .
 - ¿Cuál es el coeficiente de correlación lineal? ¿Y el de determinación?
3. Se ha llevado a cabo un estudio sobre un total de 6 pacientes. Se ha determinado en cada uno de ellos la concentración de una determinada sustancia A en sangre (X) y la concentración de una determinada sustancia B en sangre (Y). Ambas variables se miden en mg/100ml:

X	8	50	81	102	140	181
Y	0.12	0.71	1.09	1.38	1.95	2.5

- Obtén la ecuación de la recta de regresión que explique Y en función de X por el método de mínimos cuadrados.
- Estudia el grado de asociación lineal de la muestra anterior.
- Supongamos que sabemos que un nuevo paciente tiene una concentración en sangre de la sustancia A igual a 95, pero hemos extraviado su correspondiente medida de la concentración de la sustancia B. Haz una predicción de dicha concentración.

4. De una variable estadística bidimensional (X, Y) sabemos que:

- La recta de regresión de Y sobre X es $Y = 2 + 0.5X$.
- La recta de regresión de X sobre Y es $X = -4 + 2Y$.
- $s_x = 3$.

Halla la covarianza entre X e Y y la varianza de Y .

5. Registramos la evolución del nivel de creatinina en pacientes tratados con Captopril después de ser sometidos a diálisis.

Días transcurridos	1	5	10	15	20	25	35
Creatinina (mg/dl)	5.7	5.2	4.8	4.5	4.2	4	3.8

- a) ¿Cuál es la covarianza entre ambas variables?
 - b) Calcula y representa la recta de regresión que exprese el nivel de creatinina en función de los días de tratamiento.
 - c) Calcula la variabilidad no explicada (suma de cuadrados no explicada) y la variabilidad explicada (suma de cuadrados explicada por la recta de regresión).
 - d) ¿Cuál es la variabilidad total?
 - e) Calcula e interpreta el coeficiente de determinación de la recta de regresión.
6. Se han estudiado el cociente intelectual de 100 niños (X) y sus calificaciones en Matemáticas (Y) obteniéndose los siguientes resultados:

$$\bar{x} = 110 \quad \bar{y} = 2.5 \quad s_x = 10 \quad s_y = 0.5$$

Además se sabe que el coeficiente de correlación entre ambas variables es de 0.85.

- a) ¿Qué nota se puede predecir para un niño con un cociente intelectual de 125?
- b) ¿Cuál es la ecuación de la recta de regresión de X sobre Y ?

Soluciones

1. Ajustamos un modelo de regresión con Statistix y obtenemos los siguientes resultados.

Least Squares Linear Regression of plomo

Predictor

Variables	Coefficient	Std Error	T	P
Constant	1.51000	0.60124	2,51	0,0660
hierro	0.33230	0.12515	2,66	0,0567

R-Squared	0,6380	Resid. Mean Square (MSE)	0.44467
Adjusted R-Squared	0,5475	Standard Deviation	0.66683
AICc	10.705		
PRESS	3.1510		

Source	DF	SS	MS	F	P
Regression	1	3.13466	3.13466	7,05	0,0567
Residual	4	1.77867	0.44467		
Total	5	4.91333			

- Denotamos por X el hierro y por Y el plomo. La recta de regresión ajustada es $Y = 1.51 + 0.33X$.
- El coeficiente de correlación lineal es $r_{xy} = 0.7987$.
- Para una persona con un nivel de hierro igual a 2.2 cabe esperar un valor de plomo de 2.236.
- El porcentaje de explicación de la recta es 63.8%.

2. Ajustamos un modelo de regresión con Statistix y obtenemos los siguientes resultados.

Least Squares Linear Regression of IC

Predictor					
Variables	Coefficient	Std Error	T	P	
Constant	8.63102	1.07747	8,01	0,0005	
edad	-0.10813	0.02399	-4,51	0,0064	

R-Squared	0,8025	Resid. Mean Square (MSE)	1.45067		
Adjusted R-Squared	0,7631	Standard Deviation	1.20444		
AICc	14.249				
PRESS	13.591				

Source	DF	SS	MS	F	P
Regression	1	29.4810	29.4810	20,32	0,0064
Residual	5	7.2533	1.4507		
Total	6	36.7343			

3. Ajustamos un modelo de regresión con Statistix y obtenemos los siguientes resultados.

Least Squares Linear Regression of Y

Predictor					
Variables	Coefficient	Std Error	T	P	
Constant	0.00173	0.01906	0,09	0,9320	
X	0.01377	1.741E-04	79,09	0,0000	

R-Squared	0,9994	Resid. Mean Square (MSE)	5.837E-04		
Adjusted R-Squared	0,9992	Standard Deviation	0.02416		
AICc	-29.110				
PRESS	4.10E-03				

Source	DF	SS	MS	F	P
Regression	1	3.65075	3.65075	6254,49	0,0000
Residual	4	0.00233	0.00058		
Total	5	3.65308			

4. $s_{xy} = 4.5$, $s_y^2 = 2.25$.

5. Ajustamos un modelo de regresión con Statistix y obtenemos los siguientes resultados.

Least Squares Linear Regression of creat

Predictor					
Variables	Coefficient	Std Error	T	P	
Constant	5.47503	0.13930	39,30	0,0000	
dias	-0.05518	0.00723	-7,64	0,0006	

R-Squared	0,9210	Resid. Mean Square (MSE)	0.04391		
Adjusted R-Squared	0,9052	Standard Deviation	0.20955		
AICc	-10.234				
PRESS	0.7023				

Source	DF	SS	MS	F	P
Regression	1	2.56043	2.56043	58,31	0,0006
Residual	5	0.21957	0.04391		
Total	6	2.78000			

- a) Denotamos por X los días transcurridos y por Y la creatinina. La covarianza es $s_{xy} = 140.143$.
- b) La recta de regresión ajustada es $Y = 5.47503 - 0.05518X$.
- c) $VNE = 0.21957$, $VE = 2.56043$.
- d) $VT = 2.78$.
- e) $R^2 = 0.921$.
6. a) Para un niño con un cociente intelectual de 125, predecimos una nota de 3.1375.
- b) La ecuación de la recta de regresión de X sobre Y es $X = 67.5 + 17Y$.