PROBLEMAS DE BIOESTADÍSTICA

Asignatura: Bioestadística

Curso: 1^0 de Medicina

UNIVERSIDAD SAN PABLO CEU

FACULTAD DE MEDICINA

Santiago Angulo Díaz-Parreño   José Miguel Cárdenas Rebollo
Anselmo Romero Limón   Virginia Ruiz Morillo   Alfredo Sánchez Alberca

8 de febrero de 2018

Estadística Descriptiva

1. Se realizó una encuesta a 40 personas de más de 70 años sobre el número de medicamentos distintos que tomaban habitualmente. El resultado de dicha encuesta fue el siguiente:

$$3-1-2-2-0-1-4-2-3-5-1-3-2-3-1-4-2-4-3-2$$
$$3-5-0-1-2-0-2-3-0-1-1-5-3-4-2-3-0-1-2-3$$

Se pide:

a) Obtener la distribución de frecuencias de la muestra.
b) Dibujar el diagrama de barras y el polígono de frecuencias asociados.
c) Dibujar el diagrama de frecuencias acumuladas.
d) Calcular la media aritmética, la mediana y la moda.
e) Calcular la varianza y la desviación típica.
f) Calcular el coeficiente de variación de Pearson.

*2. El número de lesiones padecidas durante una temporada por cada jugador de un equipo de fútbol fue el siguiente:

$$0-1-2-1-3-0-1-0-1-2-0-1$$
$$1-1-2-0-1-3-2-1-2-1-0-1$$

Se pide:

a) Construir la tabla de frecuencias.
b) Dibujar el polígono de frecuencias.
c) Calcular los cuartiles y el rango intercuartílico e interpretarlo.
d) Calcular el coeficiente de asimetría e interpretarlo.

3. La siguiente tabla expresa la distribución de las puntuaciones obtenidas por un grupo de alumnos.

0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
7	8	13	6	7	6	6	5	6	2

Se pide:

a) Dibujar el histograma y polígono de frecuencias.
b) Calcular la media aritmética, la mediana y la moda.
c) Calcular el percentil 92.
d) Calcular la desviación típica.
e) Calcular el coeficiente de asimetría.
f) Calcular del coeficiente de curtosis.

4. Con el fin de realizar un estudio sobre el aprovechamiento de la energía solar, se han contabilizado las horas de sol registradas durante el mes de enero en las estaciones meteorológicas españolas. Los datos obtenidos son los siguientes:

Horas de Sol	N^o de estaciones
De 50 a 70	2
De 70 a 90	6
De 90 a 110	12
De 110 a 130	12
De 130 a 150	16
De 150 a 170	18
De 170 a 190	10
De 190 a 210	2
De 210 a 230	2
De 230 a 250	2

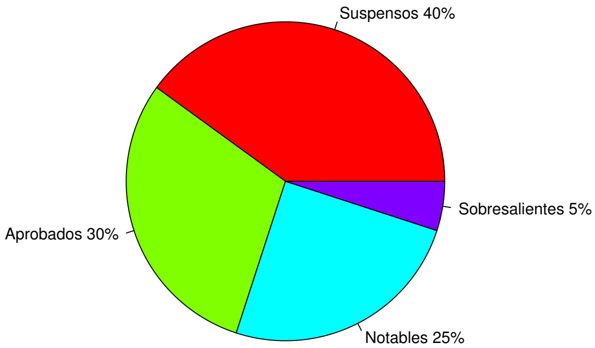
Hállese la media de horas de sol habidas en dicho mes, la desviación típica y el coeficiente de asimetría.

*5. En un estudio sobre el crecimiento se tomaron dos muestras, una de niños recién nacidos y otra de niños con un año de edad. Las estaturas observadas en cada muestra fueron:

Recién nacidos: 51-50-51-53-49-50-53-50-47-50.
Niños de un año: 62-65-69-71-65-66-68-69.

¿Según el coeficiente de variación, en cuál de las dos muestras es más representativa la media?

**6. El siguiente diagrama refleja el porcentaje de calificaciones obtenidas en un examen realizado a 80 alumnos:

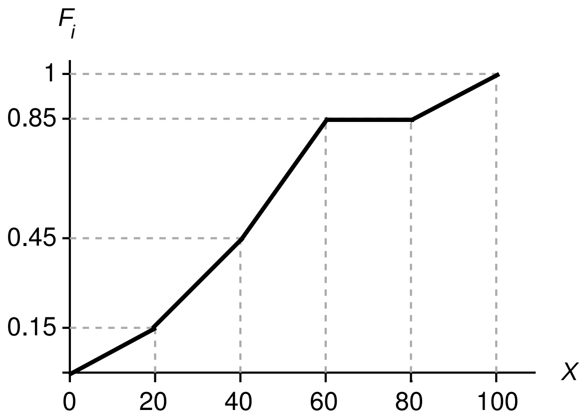


Se pide:

a) Construir la tabla de frecuencias para las calificaciones.
b) Dibujar el polígono de frecuencias acumuladas.
c) Calcular todos los estadísticos de tendencia central que sean posibles.
d) A partir de la variable calificación, construir la variable nota con los siguientes intervalos: Suspenso $[0,5)$, Aprobado $[5,7)$, Notable $[7,9)$ y Sobresaliente $[9,10]$, y calcular la nota media y estudiar su representatividad.

Nota: En los tres primeros apartados se debe trabajar con la variable calificación, mientras que en el último debe utilizarse la variable nota.

7. Dada la gráfica correspondiente a un polígono acumulativo de frecuencias relativas de una variable estadística agrupada en intervalos de una muestra de tamaño 20



se pide:

a) Construir la tabla de frecuencias.
b) Dibujar el histograma correspondiente.
c) Calcular la mediana y la moda.
d) Calcular la media aritmética y la desviación típica.

*8. Dada la siguiente tabla de frecuencias:

Intervalos	n_i	f_i	N_i	F_i
$[0,5)$	2			
$[5,10)$			8	
$[10,15)$				0.7
$[15,20)$	6			

a) Completar la tabla.

b) Calcular el coeficiente de variación y el rango intercuartílico e interpretar los resultados.

*9. Si a todos los datos de una muestra se les suma una misma cantidad positiva, ¿cómo se ve afectada la representatividad de la media? ¿Y si se multiplican por un mismo número distinto de 0? Razonar la respuesta.

*10. Se ha llevado a cabo un estudio sobre el número de radiografías realizadas durante el último año a un grupo de 200 personas, y la información se presenta en la siguiente tabla incompleta:

Radiografías	Personas	f_i	F_i
0		0,2	
1	84		
2			0,72
3			
4	24		
5		0,02	

a) Completar tabla.

b) Calcular media, mediana, desviación típica y coeficiente de variación e interpretar los resultados.

*11. En un estudio diseñado para investigar la efectividad de un nuevo producto anestésico local, la misma cantidad de producto fue suministrada a 20 pacientes, y se midió el tiempo transcurrido hasta lograr cierto grado de sensibilidad. Los resultados, en minutos, son los siguientes:

38, 43, 52, 64, 39, 54, 51, 47, 42, 58, 63, 36, 39, 47, 49, 46, 52, 44, 38, 57

a) Agrupar los datos desde 35 a 65 en 6 clases diferentes.

b) Una vez agrupados, calcular: Media, Desviación Típica y Coeficiente de Asimetría.

c) Teniendo en cuenta la distribución agrupada y suponiendo que todos aquellos datos que se encuentren por arriba del percentil 95 tienen un comportamiento anormal, ¿cuáles de los pacientes se puede considerar que han tenido un tiempo de insensibilidad anormal?.

**12. A continuación figura la distribución de edades de una muestra de 65 individuos sujetos a rehabilitación tras un infarto de miocardio:

Edad	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)
n_i	6	12	23	19	5

Por otra parte, sabemos que una distribución normal es simétrica y mesocúrtica, y, por tanto, una primera idea de si los datos muestrales provienen de una distribución normal nos la puede dar ver si tanto el coeficiente de asimetría como el de curtosis se encuentran en el intervalo [-2, 2] (en definitiva, lo suficientemente cercanos a 0 como para poder suponer que la distribución es simétrica y mesocúrtica).

a) ¿Podríamos suponer según esto que nuestros datos provienen de una distribución normal?.

b) ¿Calcular la edad, en esta muestra, por encima de la cuál se encuentra el 15 % de los individuos sujetos a rehabilitación tras un infarto de miocardio?.

**13. Para obtener información acerca del porcentaje de albúmina en el suero proteico de personas adultas, se analizaron muestras de 32 personas, con los siguientes resultados:

70,2	63,5	65,8	67,9	60,1	69,7	64,2	65,3
62,8	68,4	65,2	66,3	70,7	71,8	68,7	71,9
64,4	62,4	60,4	67,0	62,9	65,9	67,5	66,6
67,8	70,5	63,1	65,3	69,5	71,4	61,0	64,3

a) Agrupar la distribución de porcentajes de albúmina en 6 clases de igual amplitud, desde 60 hasta 72.

b) En la distribución agrupada calcular media, desviación típica, y cuartiles.

c) ¿Es representativa la media de la muestra de porcentajes de albúmina?.

d) Dibujar el diagrama de caja y bigotes de la distribución y determinar si hay o no algún dato atípico. Para ello, considerar que uno de los criterios más habituales para dibujar el gráfico de caja y bigotes es:

- Los extremos de la caja los marcan el primer cuartil C_1 y el tercer cuartil C_3 de la distribución.
- Dentro de la caja también se da la posición de la mediana mediante una línea recta.
- Para los bigotes b_1 y b_2, inicialmente se determina la posición de los puntos denominados vallas v_1 y v_2 restando y sumando respectivamente a primer y tercer cuartil $1,5$ veces el recorrido intercuartílico RI:

$$v_1 = C_1 - 1,5RI$$
$$v_2 = C_3 + 1,5RI$$

De tal forma que b_1 es el dato de la muestra más cercano a v_1 sin que su valor sea inferior a v_1, y b_2 es el dato de la muestra más cercano a v_2 sin que su valor sea superior a v_2.

**14. En un estudio estadístico realizado en la comunidad valenciana, aparece la siguiente tabla con los datos referidos al número de embarazos, abortos e hijos en una muestra de 999 mujeres:

n	Embarazos	Abortos	Hijos nacidos
0	61	751	67
1	64	183	80
2	328	51	400
3	301	10	300
4	122	2	90
5	81	2	62
6	29		
7	11		
8	2		
Total	999	999	999

a) ¿En qué variable de las tres estudiadas es más representativa la media?

b) Calcular la mediana del número de hijos nacidos.

c) ¿Qué valor es relativamente más alto, el de una mujer que ha tenido 4 abortos, o el de otra que ha tenido 7 embarazos?. Justificar adecuadamente la respuesta.

**15. Un médico de familia analiza el número de recetas que ha expedido entre sus abonados en los dos últimos meses. Teniendo en cuenta que atiende a 1000 abonados, la distribución del número de recetas es:

Recetas	Abonados
0	509
1	254
2	125
3	88
5	20
8	4

a) Calcular: media, desviación típica y coeficiente de variación del número de recetas. Interpretar el coeficiente de variación.

b) Calcular el coeficiente de asimetría de la distribución. Interpretarlo.

c) Teniendo en cuenta la definición dada más abajo, calcular le Media Recortada 5 % ($MR_{0,1}$) del número de recetas. ¿Cuándo crees que será conveniente la utilización de la media recortada en lugar de la media aritmética?.

Definición: Sea una muestra formada por n observaciones que se han ordenado de manera creciente. La Media Recortada una proporción p de casos en cada extremo de la distribución viene dada por medio de la siguiente expresión:

$$MR_p = \frac{\sum_{i=g+1}^{n-g} x_i}{n-2g}$$

donde g expresa el número de observaciones que deben ser eliminadas de cada extremo de la distribución, y que viene dado por $g = [p \cdot n]$, donde el símbolo [ ] implica que hay que tomar la parte entera del producto.

**16. En una ciudad española se está realizando un estudio de la edad de la madre en el momento del primer parto segmentando según si la madre es española o extranjera. Los datos obtenidos fueron:

Edad (años)	Frecuencia Españolas	Frecuencia Extranjeras
[15, 20)	8	6
[20, 25)	65	30
[25, 30)	253	21
[30, 35)	362	16
[35, 40)	125	4
[40, 45)	41	

a) ¿En qué colectivo es más representativa la media de edad en el momento del primer parto? Justificar adecuadamente la respuesta.

b) Calcular la mediana de la edad en el momento del primer parto en el grupo de españolas.

c) Calcular la media de edad global en el momento del primer parto considerando conjuntamente españolas y extranjeras.

d) Teniendo en cuenta sólo las españolas, ¿en qué percentil de edad se encuentra una mujer que ha tenido su primer parto a los 32 años?

**17. En una ciudad se ha realizado un estudio sobre la edad media, en años, de los individuos en el momento de su muerte para el período de años que va desde 2004 a 2011. Además, se ha anotado también el número de defunciones en cada uno de esos años, obteniendo:

Año	Edad media al morir (años)	Defunciones
2004	79,4	95
2005	78,3	90
2006	80,2	101
2007	81,3	85
2008	83,1	115
2009	79,8	123
2010	84,3	130
2011	81,5	121

Considerando todo el período de años que va del 2004 al 2011, se pide:

a) Calcular la edad media global al morir en esa ciudad.

b) ¿Qué media es más representativa, la de la edad media al morir o la del número de defunciones? Justificar adecuadamente la respuesta.

c) ¿Cuánto vale la mediana de la edad media al morir?

d) ¿Cuánto vale la media recortada un 10 % (un 5 % en cada extremo) de la edad media al morir?

e) ¿Cuánto vale el apuntamiento de la edad media al morir? Interpretar el resultado obtenido.

**18. En un estudio para evaluar los efectos de la radioterapia se cuantificó la cantidad de tejido canceroso muerto en hígado, en gramos, después de aplicar a los pacientes una dosis de 5 grays. En total participaron 20 pacientes, 11 mujeres y 9 hombres, y los resultados obtenidos para el total de pacientes, X, y para los hombres, Y, aparecen recogidos en los siguientes sumatorios:

$$\sum x_i = 452; \quad \sum x_i^2 = 10312; \quad \sum (x_i - \bar{x})^3 = 55,440; \quad \sum (x_i - \bar{x})^4 = 947,264$$

$$\sum y_j = 190; \quad \sum y_j^2 = 4026; \quad \sum (y_j - \bar{y})^3 = 2,025; \quad \sum (y_j - \bar{y})^4 = 48,996$$

a) Calcular: media, desviación típica, coeficiente de variación, asimetría y curtosis, todos ellos en el total de datos.

b) ¿En qué muestra es más representativa la media, en la de hombres, en la de mujeres o en el total?

c) ¿En qué muestra hay mayor apuntamiento, en la de hombres o en el total?

d) ¿Qué valor es relativamente más alto, el de un hombre en el que mueren 20 gramos de tejido canceroso o el de una mujer en la que mueren 22 gramos? Justificar adecuadamente la respuesta.

Regresión y Correlación

19. Dada la siguiente tabla de correlación:

$X \setminus Y$	1	2	3
$[-2,2)$	3	6	1
$[2,6)$	4	7	3
$[6,10)$	5	3	0

Determinar:

a) Las distribuciones marginales. Media, Moda y Mediana.

b) Rectas de Regresión.

c) Coeficiente de correlación lineal. Interpretar el resultado.

20. Se ha realizado un estudio comparativo de las puntuaciones obtenidas por los alumnos en un test de ingreso en la universidad (X), y el número de asignaturas aprobadas en el primer curso (Y). Los resultados obtenidos se expresan en la siguiente tabla:

$X \setminus Y$	0	1	2	3	4
$[0,10)$	2	2	1	0	0
$[10,20)$	1	1	2	2	0
$[20,30)$	0	1	3	4	1
$[30,40)$	0	0	2	2	6

Se desea calcular:

a) Recta de regresión de X sobre Y.

b) Coeficiente de correlación e interpretación del mismo.

c) Si la universidad en cuestión sólo contara con alumnos que al menos logren aprobar dos asignaturas, ¿qué número de preguntas respondidas correctamente exigirá en el test?.

*21. En una población se ha realizado un estudio sobre el nivel de colesterol en sangre (X) y la tensión arterial máxima (Y) obteniendo. Para ello se ha tomado una muestra de 80 individuos que ha dado lugar a la siguiente tabla de frecuencias:

$X \setminus Y$	$[110,130)$	$[130,150)$	$[150,170)$	n_x
$[170,190)$		4		12
$[190,210)$	10	12	4	
$[210,230)$	7		8	
$[230,250)$	1			18
n_y		30	24	

Se pide:

a) Completar la tabla.

b) Recta de regresión del nivel de colesterol sobre la tensión.

c) Coeficiente de determinación e interpretación.

d) La tensión arterial máxima esperada para una persona cuyo nivel de colesterol es 270.

*22. En un centro dietético se está probando una nueva dieta de adelgazamiento en una muestra de 12 individuos. Para cada uno de ellos se ha medido el número de días que lleva con la dieta y el número de kilos perdidos desde entonces, obteniéndose los siguientes resultados:

(33 , 3.9), (51 , 5.9), (30 , 3.2), (55 , 6.0), (38 , 4.9), (62 , 6.2),
(35 , 4.5), (60 , 6.1), (44 , 5.6), (69 , 6.2), (47 , 5.8), (40 , 5.3)

Se pide:

a) Dibujar el diagrama de dispersión. Según la nube de puntos, ¿qué tipo de modelo explicaría mejor la relación entre los días de dieta y los kilos perdidos?

b) Calcular la recta de regresión de los kilos perdidos con respecto a los días de dieta.
Nota: Utilizar los datos muestrales sin agrupar.

c) Utilizar la recta anterior para predecir en número de kilos perdidos tras 40 días de dieta y tras 100 días. ¿Son fiables estas predicciones?

*23. Al realizar un estudio sobre la dosificación de un cierto medicamento, se trataron 6 pacientes con dosis diarias de 2 mg, 7 pacientes con 3 mg y otros 7 pacientes con 4 mg. De los pacientes tratados con 2 mg, 2 curaron al cabo de 5 días, y 4 al cabo de 6 días. De los pacientes tratados con 3 mg diarios, 2 curaron al cabo de 3 días, 4 al cabo de 5 días y 1 al cabo de 6 días. Y de los pacientes tratados con 4 mg diarios, 5 curaron al cabo de 3 días y 2 al cabo de 5 días. Se pide:

a) Dar el coeficiente de correlación e interpretación.

b) Determinar el tiempo esperado de curación para una dosis de 5 mg diarios.

24. Se consideran dos variables aleatorias X e Y tales que:

La recta de regresión de Y sobre X viene dada por la ecuación: $y - x - 2 = 0$.

La recta de regresión de X sobre Y viene dada por la ecuación: $y - 4x + 22 = 0$.

Calcular:

a) Valores de $\overline{x}$ e $\overline{y}$.

b) Coeficiente de correlación lineal.

25. En el ajuste rectilíneo a una distribución bidimensional se sabe que $\overline{x} = 2$, $\overline{y} = 1$, y el coeficiente de correlación lineal es 0 ($r = 0$).

a) Si $x = 10$, ¿cuál será el valor interpolado para y?.

b) Si $y = 5$, ¿cuál será el valor interpolado para x?.

c) Dibuja las rectas de regresión de Y sobre$X$, y la de X sobre Y.

*26. Se han medido dos variables S y T en 10 individuos, obteniéndose los siguientes resultados:

(-1.5 , 2.25), (0.8 , 0.64), (-0.2 , 0.04), (-0.8 , 0.64), (0.4 , 0.16),
(0.2 , 0.04), (-2.1 , 4.41), (-0.4 , 0.16), (1.5 , 2.25), (2.1 , 4.41).

Se pide:

a) Calcular la covarianza de S y T.

b) ¿Se puede afirmar que S y T son independientes? Justificar la respuesta.

c) ¿Qué valor predice la correspondiente recta de regresión para $t = 2$?

**27. Se realiza un estudio para establecer una ecuación mediante la cual se pueda utilizar la concentración de estrona en saliva para predecir la concentración del esteroide en plasma libre. Se extrajeron los siguientes datos de 10 varones sanos:

Estrona	1,4	7,5	8,5	9	9	11	13	14	14,5	16
Esteroide	30	25	31,5	27,5	39,5	38	43	49	55	48,5

a) Comprobar la idoneidad del modelo lineal de regresión. Si el modelo es apropiado, hallar la recta de regresión de la concentración de estrona en función de la concentración de esteroide.

b) Si un individuo presenta una concentración de estrona en saliva de 10, ¿qué concentración de esteroide en plasma libre predeciría el modelo de regresión lineal?

c) Para los dos primeros individuos, calcular los errores que se comenten al utilizar el modelo de regresión lineal para predecir la concentración de estrona. Razonar a que se deben estos errores.

**28. La tabla siguiente contiene los datos de las presiones sistólicas de 15 individuos en función de la edad de estos.

Edad(x)	20	30	40	50	60
Sistólica(y)	121	131	132	136	134
	130	125	129	128	142
	125	128	131	134	137

a) ¿Qué porcentaje de la varianza de la presión sistólica es explicada mediante un modelo de regresión lineal por la varianza de la edad?

b) ¿Qué edad le correspondería a un individuo que presenta una presión sistólica de 133? ¿Es fiable esta predicción? Razona la respuesta.

**29. En una análisis de niños sanos se deseaba establecer si existía relación lineal entre la edad (en años) del niño y el ángulo de Clarke (en grados), obteniéndose en una muestra de 7 niños los valores que aparecen a continuación:

Edad	3	4	5	6	7	8	9
Ángulo de Clarke	24	26	30	31	34	32	33

a) Calcular la ecuación de la recta de regresión del Ángulo de Clarke en función de la edad.

b) ¿Qué tanto por ciento de la variabilidad de la nube de puntos explicamos con el modelo lineal?. ¿Se puede considerar un modelo bueno?.

c) El coeficiente de correlación lineal, también llamado coeficiente de correlación de Pearson, o su correspondiente cuadrado (el coeficiente de determinación lineal), dan una medida del grado de asociación lineal entre variables pero siempre y cuando las mismas sean cuantitativas y con datos provenientes de distribuciones normales. Si no se cumplen los supuestos anteriores, muy a menudo se utiliza como medida de asociación lineal el *coeficiente de correlación de Spearman*, que se obtiene mediante la fórmula:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

donde:

- n es el número total de datos de la muestra.
- d_i es la llamada "distancia entre rangos"de cada uno de los puntos de la muestra. Se calcula ordenando por separado los valores de X y los de Y, y el denominado "rango"del valor, propio de cada valor, es simplemente su número de orden; mientras que la distancia entre rangos, propia de cada punto, se obtiene como la resta entre los rangos (números de orden) de la x y de la y del punto.

La interpretación del coeficiente de correlación lineal de Spearman es la misma que la del coeficiente de correlación lineal de Pearson (visto en teoría).

Para la muestra dada, calcular el coeficiente de correlación lineal de Spearman.

*30. En un experimento se ha medido el número de bacterias por unidad de volumen en un cultivo, cada hora transcurrida, obteniendo los siguientes resultados:

Horas	0	1	2	3	4	5	6	7	8
Nº de Bacterias	25	28	47	65	86	121	190	290	362

Se pide:

a) Dibujar el diagrama de dispersión. Según este diagrama, ¿qué tipo de modelo explicaría mejor la relación entre el numero de bacterias y las horas transcurridas?

b) Según el modelo anterior, ¿cuántas bacterias tendríamos al cabo de 3 horas y media? ¿Y al cabo de 10 horas? ¿Son fiables estas predicciones?

c) ¿Cuánto tiempo tendría que transcurrir para que en el cultivo hubiese 100 bacterias?

*31. La Actividad de una sustancia radiactiva en función del tiempo (en número de desintegraciones por segundo) viene dada por la siguiente tabla:

t (horas)	0	10	20	30	40	50	60	70
A (10^7 desintegraciones/s)	25,9	8,16	2,57	0,81	0,25	0,08	0,03	0,01

a) Representar los datos de la Actividad en función del tiempo. A la vista de la representación, ¿qué modelo de regresión explicaría mejor la relación entre la Actividad y el tiempo transcurrido?

b) Representar el logaritmo neperiano de la Actividad en función del tiempo. ¿Qué modelo de regresión se utilizaría para ajustar la nube de puntos obtenida?

c) Calcular la ecuación de la recta de regresión del logaritmo neperiaro de la Actividad en función del tiempo.

d) Teniendo en cuenta que, en teoría, la Actividad de una sustancia radiactiva en función del tiempo viene dada por la ecuación:

$$A(t) = A_0 e^{-\lambda t}$$

donde A_0 es la actividad inicial y λ es la llamada Constante de Desintegración, propia de cada sustancia radiactiva, utilizar la pendiente de la ecuación de la recta obtenida en el apartado anterior para calcular la constante de desintegración radiactiva de la sustancia con la que se han generado los datos.

*32. La concentración de un fármaco en sangre, C en mg/dl, es función del tiempo, t en horas, y viene dada por la siguiente tabla:

t	2	3	4	5	6	7	8
C	25	36	48	64	86	114	168

a) Según el modelo exponencial, ¿qué concentración de fármaco habría a las $4,8$ horas? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.

b) Según el modelo lineal, ¿qué tiempo tendría que transcurrir para que la concentración de fármaco fuese de 100 mg/de? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.

*33. En un estudio se pretende ver si existe o no relación entre la cantidad total inyectada de una determinada sustancia durante un mes, en cm^3, y el aumento de peso provocado en las personas sometidas al tratamiento, en kg. Para el estudio se tomaron paciente de unas características similares en edad, peso y altura, y los resultados obtenidos en una muestra de 6 personas fueron los siguientes:

Aumento Peso (kg)	1,9	2,4	2,8	3,1	3,3	3,4
Sustancia (cm^3)	15	20	25	30	35	40

a) Calcular el modelo logarítmico del Aumento de Peso en función de la Cantidad de Sustancia administrada.

b) ¿Es bueno o malo el modelo logarítmico calculado? Justificar adecuadamente la respuesta.

c) ¿Qué cantidad de sustancia tendríamos que administrar a una persona que queremos que aumente su peso en $3,2$ kg?

34. En un estudio en el que participaron las 8 universidades de una región se ha valorado la excelencia docente e investigadora, estableciendo los siguientes rankings (de mejor a peor):

Ranking Docencia	3	4	8	5	2	1	6	7
Ranking Investigación	6	5	4	3	7	8	1	2

¿Se puede decir que existe relación entre la excelencia docente y la investigadora? Justificar la respuesta.

**35. Se supone que la concentración de una sustancia en sangre Y depende de la concentración de otra sustancia X, ambas en microgramos por decilitro de sangre. Para probarlo, se han medido las concentraciones de siete individuos obteniendo:

X	2,1	4,9	9,8	11,7	5,9	8,4	9,2
Y	1,3	1,5	1,7	1,8	1,5	1,7	1,7

a) Calcular la ecuación del modelo lineal que sirve para dar Y como función de X.

b) ¿Es el modelo lineal adecuado para ajustar la nube de puntos? Justificar la respuesta.

c) Si los investigadores piensan que Y depende de X según el modelo $Y = aX^b$, ¿cuánto deberían valer a y b para que el ajuste fuese el más adecuado?

**36. En un grupo de personas se anota su edad y el número de veces que han acudido a consulta médica en el último año, obteniéndose los siguientes resultados:

Edad	36	58	72	84	65	44
Número	1	3	5	6	4	2

a) Calcular la ecuación del modelo lineal del número de consultas en función de la edad.

b) Calcular la ecuación del modelo exponencial del número de consultas en función de la edad.

c) Calcular la ecuación del modelo logarítmico del número de consultas en función de la edad.

d) ¿Qué porcentaje de la variabilidad del número de consultas queda explicado por cada uno de los modelos anteriores?

**37. En una ciudad se ha realizado un estudio sobre la edad media, en años, de los individuos en el momento de su muerte, desde 1930 hasta 1995, obteniendo los siguientes datos:

Año	Edad media al morir (años)
30	56,3
40	64,2
50	70,1
60	74,3
70	77,8
80	79,9
90	81,4
95	82,6

a) Calcular la ecuación del modelo logarítmico de la edad media al morir en función del año.

b) ¿Qué tanto por ciento de la variabilidad de la nube de puntos se explica con el modelo logarítmico? ¿Es mejor o peor que el modelo lineal? Justificar adecuadamente las respuestas.

c) Según el modelo logarítmico y suponiendo que todos los años tienen 365 días, ¿qué edad media al morir se espera que tengan los habitantes de esa ciudad en 2012? ¿En qué día de qué año se alcanzó en esa ciudad una media de edad al morir de 77,3 años? ¿Son buenas las predicciones realizadas? Justificar adecuadamente la respuesta.

**38. Supongamos dos variables biológicas, X e Y, que se piensa que están relacionadas. Para comprobarlo se miden las 2 variables en 10 individuos y se obtienen los siguientes sumatorios:

$$\sum_{i=1}^{10} x_i = 55\ ,\ \sum_{i=1}^{10} y_i = 478\ ,\ \sum_{i=1}^{10} \ln x_i = 15{,}104\ ,\ \sum_{i=1}^{10} \ln y_i = 36{,}028\ ,\ \sum_{i=1}^{10} x_i^2 = 385\ ,\ \sum_{i=1}^{10} y_i^2 = 33148$$

$$\sum_{i=1}^{10} (\ln x_i)^2 = 27{,}650\ ,\ \sum_{i=1}^{10} (\ln y_i)^2 = 135{,}667\ ,\ \sum_{i=1}^{10} x_i y_i = 3527\ ,\ \sum_{i=1}^{10} x_i \ln y_i = 220{,}055\ ,\ \sum_{i=1}^{10} y_i \ln x_i = 914{,}932$$

Con ello:

a) Calcular la ecuación de la recta de regresión de Y como función de X.

b) Calcular la ecuación del modelo exponencial de Y como función de X.

c) Calcular la ecuación del modelo logarítmico de Y como función de X.

d) ¿Qué tanto por ciento de la variabilidad de la nube de puntos se explica con cada uno de los modelos anteriores? ¿Cuál de ellos sería mejor para realizar el ajuste? Justificar la respuesta.

**39. Durante el primer año de vida de un niño se le pesó en varias ocasiones obteniéndose los siguientes resultados:

Edad (meses)	2	3	6	8	9	12
Peso (Kg)	5,2	6,1	7,4	8,6	8,9	9,8

a) Dar la ecuación del modelo lineal que mejor exprese el peso en función de la edad.

b) Según el modelo lineal, ¿Cuánto aumenta el peso por cada mes de vida?

c) Dar la ecuación del modelo potencial ($Y = aX^b$) que mejor exprese el peso en función de la edad.

d) Calcular con ambos modelos el peso esperado a los 4 meses de edad e indicar razonadamente cuál de las dos predicciones es más fiable.

**40. Para analizar la incidencia del síndrome de Down en una región, durante una año se han anotado el total de nacimientos según la edad de la madre (en años), y de ellos cuántos tenían el síndrome, obteniendo:

Edad	Nacimientos	Síndrome
[15 , 20)	1050	1
[20 , 25)	8933	10
[25 , 30)	15642	12
[30 , 35)	20384	26
[35 , 40)	6720	15
[40 , 45)	1356	9

a) ¿Qué edad media es superior, la de las madres que han tenido hijos sin síndrome de Down o la de las que los han tenido con el síndrome? Justificar numéricamente la respuesta.

b) Calcular el coeficiente de asimetría de la edad de las madres que han tenido hijos con síndrome de Down.

c) Calcular el percentil 70 de la edad de las madres que han tenido hijos sin el síndrome.

d) Calcular el tanto por mil de nacimientos con síndrome de Down en cada clase de la variable edad.

e) Dar la ecuación del modelo exponencial del tanto por mil de nacimientos con síndrome en función de la edad, considerando tan sólo los 6 pares de valores (la frecuencia de cada pareja de valores es la unidad).

f) Según el modelo anterior, ¿qué tanto por mil de síndrome de Down se espera en madres con 19 años? ¿Es buena la predicción anterior? Justificar numéricamente la respuesta.

**41. Cuatro alumnos, A, B, C y D calcularon las rectas de regresión de Y sobre X y de X sobre Y, obteniendo los siguientes resultados:

Alumno	Recta Y sobre X	Recta X sobre Y
A	$y = 2x + 3$	$y = 3x - 1$
B	$y = 3x + 1$	$y = -x + 2$
C	$y = 3$	$x = 2$
D	$y = 3x + 4$	$y = 2x + 1$

Indicar razonadamente cuáles de las respuestas no pueden ser correctas.

**42. En un estudio sobre radioterapia se cuantificó el efecto de la radiación en el hígado, tanto en tejido canceroso como en sano, midiendo la cantidad en gramos que moría en función de la dosis de radiación administrada, en grays (un gray es equivalente a una absorción de un julio de energía ionizante por cada Kg de tejido irradiado). Para ello, trabajaron con 7 pacientes en los que se anotó tanto la cantidad de radiación como de tejido canceroso y sano muertos. Los datos aparecen en la siguiente tabla:

Grays	Canceroso muerto (g)	Sano muerto (g)
1	3	1
2	8	3
3	19	6
4	25	8
5	29	14
6	31	21
7	32	34

a) Según el modelo logarítmico, ¿qué cantidad de tejido canceroso muerto se espera que se produzca con una dosis de 5,8 grays?

b) Según el modelo exponencial, ¿qué cantidad de tejido sano muerto se espera que se produzca con una dosis de 3,2 grays?

c) ¿Qué predicción es mejor, la del modelo exponencial o la del logarítmico? Justificar la respuesta.

d) ¿Qué dosis se necesitará para que haya 14 gramos de tejido canceroso muerto?

Cálculo de Probabilidades

43. En un laboratorio hay 10 frascos de ácido sulfúrico y 6 de ácido nítrico, y en otro hay 4 frascos de ácido sulfúrico y 14 de ácido nítrico. Se saca al azar un frasco de cada laboratorio. Hallar la probabilidad de que:

    a) Los dos frascos sean de ácido sulfúrico.

    b) Los dos sean de ácido nítrico.

    c) Uno sea de ácido sulfúrico y otro de ácido nítrico.

44. Sean A y B sucesos de un mismo espacio muestral tales que: $\mathrm{P}(A)$=3/8, $\mathrm{P}(B)$=1/2, $\mathrm{P}(A\cap B)$=1/4. Calcular:

    a) $\mathrm{P}(A\cup B)$.

    b) $\mathrm{P}(\overline{A})$ y $\mathrm{P}(\overline{B})$.

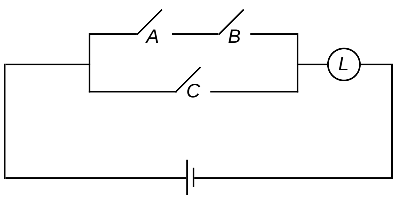
    c) $\mathrm{P}(\overline{A}\cap\overline{B})$.

    d) $\mathrm{P}(A\cap\overline{B})$.

    e) $\mathrm{P}(A/B)$.

    f) $\mathrm{P}(A/\overline{B})$.

45. Dado el siguiente circuito



    si la probabilidad de estar cerrado el interruptor A es 0.8, el B 0.9 y el C 0.7, ¿cuál es la probabilidad de que esté encendida la lámpara L?

46. La probabilidad de contraer hepatitis a partir de una unidad de sangre es 0'01. Un paciente recibe dos unidades de sangre durante su estancia en el hospital. ¿Cuál es la probabilidad de que contraiga hepatitis como consecuencia de ello?

47. Sean A y B sucesos de un mismo espacio muestral, tales que $\mathrm{P}(A)$=0'6 y $\mathrm{P}(A\cup B)$=0'9. Calcular $\mathrm{P}(B)$ si:

    a) A y B son independientes.

    b) A y B son incompatibles.

*48. En un estudio sobre el tabaco, se informa que el 40 % de los fumadores tienen padre fumador, el 25 % tienen madre fumadora, y el 52 % tiene al menos uno de los dos padres fumadores. Se elige una persona fumadora al azar. Calcular:

    a) Probabilidad de que la madre sea fumadora si lo es el padre.

    b) Probabilidad de que la madre sea fumadora si no lo es el padre.

    c) ¿Son independientes el tener padre fumador y el tener madre fumadora.

*49. Un equipo de atención primaria de salud realiza un estudio de la población, para evaluar la incidencia de hipertensión e hipercolesterolemia. Para ello analizan a 1000 personas de dicha población, seleccionadas aleatoriamente, encontrándose que 180 presentan hipertensión, 140 hipercolesterolemia y 800 ninguna de ambas. Se pide calcular la probabilidad de que una persona tomada al azar *a*) Presente ambas enfermedades.

b) Presente hipertensión si no presenta hipercolesterolemia.

50. A partir de una investigación realizada, se sabe que el 10 % de las personas de 50 años sufren un tipo particular de artritis. Se ha desarrollado un procedimiento para detectar esta enfermedad, y por las pruebas realizadas se observa que si se aplica el procedimiento a individuos que padecen la enfermedad, da positivo en el 85 % de los casos, mientras que si se aplica a individuos sanos, da positivo en el 4 % de los casos. Se pide:

    a) Calcular la probabilidad de que realizado el procedimiento a una persona, el resultado sea positivo.

    b) Si el resultado de aplicar el procedimiento a una persona ha sido positivo, ¿Cuál es la probabilidad de que padezca la enfermedad?

51. En un servicio clínico digestivo se sabe que, de cada 1000 pacientes con dolor de estómago, 700 presentan gastritis, 200 presentan úlcera y 100 presentan cáncer. En el análisis de la sintomatología gástrica, se ha comprobado que las probabilidades de presentar vómitos son 0'3 en el caso de gastritis, 0'6 en el caso de úlcera y 0'9 en el caso de cáncer. Llega un nuevo paciente con dolor de estómago que, además, presenta vómitos. ¿Qué diagnosticaríamos?

*52. Un test diseñado para diagnosticar el cáncer de cuello uterino da resultado positivo en el 10 % de los casos en los que no existe la enfermedad, y da negativo en el 5 % de los casos en los que sí que existe la enfermedad.

Se sabe que en una cierta población de mujeres, el 4 % padece dicha enfermedad. Si una mujer elegida aleatoriamente se somete al test, y da positivo, ¿qué probabilidad hay de que padezca la enfermedad?

*53. En un estudio se han probado tres tipos de tratamientos A, B y C contra una determinada enfermedad. De los pacientes participantes en el estudio, el 50 % fueron tratados con el tratamiento A, el 30 % con el B y el 20 % con el C. Posteriormente se observaron los pacientes que sanaron y los que tuvieron algún efecto secundario, según se muestra en la siguiente tabla:

Tratamiento	Sanados	Con efectos secundarios
A	86 %	12 %
B	92 %	14 %
C	81 %	6 %

Se pide:

a) Si se selecciona un enfermo al azar, ¿cuál es la probabilidad de que haya sanado? ¿Y de que haya tenido algún efecto secundario?

b) Si un enfermo ha sanado, ¿qué tratamiento es más probable que haya recibido? ¿Y si en vez de decirnos que ha sanado nos dicen que no ha tenido efectos secundarios?

c) Si en total hay un 8 % pacientes que no sanaron pero que tampoco tuvieron efectos secundarios, ¿cuál es la probabilidad de que un enfermo se haya curado sin tener efectos secundarios?

**54. Para comprobar la eficacia de un test diagnóstico se lleva a cabo una experiencia cuyos resultados se recogen en la siguiente tabla:

	Test +	Test −
Enfermos	4680	120
No Enfermos	80	2020

Calcular para dicho test:

a) Las probabilidades de Verdadero Negativo, Verdadero Positivo, Falso Negativo y Falso Positivo.

b) Los Valores Predictivos, tanto el Positivo como el Negativo.

c) La probabilidad de Diagnóstico Acertado.

**55. El dolor intenso sin derrame en una zona concreta de la articulación de la rodilla es síntoma de esguince en el Ligamento Lateral Externo de la misma (L.L.E.). Si los esguinces en dicho ligamento se clasifican como: de grado 1, cuando hay simple distensión, que se presenta en un 60 % de los casos; de grado 2, cuando hay ruptura parcial, que se presenta en un 30 % de los casos; y de grado 3, cuando hay ruptura total, que se presenta en un 10 %. Y teniendo en cuenta que el síntoma se presenta en un 80 % de los que tienen el esguince de grado 1, en un 90 % de los de grado 2, y en un 100 % de los de grado 3:

a) Si una persona se produce un esguince de L.L.E., ¿cuál es la probabilidad total de que padezca dolor intenso sin derrame?.

b) Si una persona llega a una consulta con dolor intenso sin derrame en la zona adecuada de la rodilla, ¿cuál sería el diagnóstico?.

c) De un total de 10000 personas analizadas con dolor intenso sin derrame en la zona adecuada de la rodilla, ¿cuántas se espera que hayan sufrido un esguince de grado 1? ¿Y de grado 2? ¿Y de grado 3?.

d) Si mantenemos iguales el resto de probabilidades dadas como dato, ¿cuáles deben ser las probabilidades de esguince de grado 2 y de grado 3 para que la probabilidad de esguince de grado 2 si se padece el síntoma sea igual a la de grado 3 si se padece el síntoma?

**56. Supongamos dos test diagnóstico, A y B, completamente independientes, que se utilizan para diagnosticar una misma enfermedad. Si la prevalencia de la enfermedad en una población es de un 2 %, la sensibilidad de A es de un 95 %, la sensibilidad de B es de un 97 %, la especificidad de A es de un 90 %, y la de B de un 85 %, calcular:

a) El valor predictivo positivo del test A.

b) La probabilidad de que, aplicados ambos a un individuo cualquiera de la población, alguno de los test dé positivo.

c) La probabilidad de que, aplicados ambos a un individuo cualquiera de la población, los dos den diagnóstico erróneo.

**57. Los estudios epidemiológicos indican que el 20 % de los ancianos sufren un deterioro neuropsicológico. Sabemos que, para este tipo de lesiones, la tomografía axial computerizada (TAC) presenta una sensibilidad del 80 %, pero que también da un 3 % de falsos positivos. Por lo tanto, para este tipo de lesiones:

a) ¿Cuáles son los valores predictivo positivo y predictivo negativo del TAC?

b) ¿Cuál es la probabilidad de diagnóstico acertado con el TAC?. Dejando igual el resto de probabilidades dadas como dato en el problema, ¿cuánto debería valer la especificidad del TAC para que la probabilidad de diagnóstico acertado fuese del 95 %?

c) En las condiciones iniciales del problema (es decir, sin el cambio introducido en la segunda pregunta del apartado anterior), si además del TAC aplicamos a un anciano un nuevo test diagnóstico que actúa de forma independiente, tal que su sensibilidad es del 98 % y su especificidad del 95 %, ¿cuál será la probabilidad de que alguno de los test se equivoque? ¿Y la de que acierten los dos?.

**58. Si suponemos una enfermedad con una prevalencia del 5 % para cuyo diagnóstico se utilizan 2 test, A y B, tal que la sensibilidad de A es del 99 % y su especificidad es del 95 %; y sabiendo que el test B da positivo en un 90 % de los individuos en los que previamente ha dado positivo A, y da negativo en un 95 % de los individuos en los que previamente ha dado negativo A, se pide:

a) Los valores predictivos del test A.

b) Si los dos test se aplican a 10000 individuos: ¿En cuántos se espera que dé positivo el test B? ¿En cuántos se espera que los dos test den positivo? ¿En cuántos se espera que alguno de los dos test dé positivo?

c) ¿Cuál debería ser la prevalencia de la enfermedad para que la probabilidad de diagnóstico acertado con A fuese del 97 %?

**59. Según la clasificación de la New York Heart Association, el grado funcional de insuficiencia cardíaca se clasifica en 4 categorías dependiendo del esfuerzo físico para que se produzca disnea (dificultad respiratoria o falta de aire):

- A la categoría A pertenecen los pacientes en los que la disnea se produce sólo en niveles de esfuerzo altos.
- A la categoría B pertenecen los que la disnea se produce en niveles de esfuerzo medianos.
- A la categoría C pertenecen los que la disnea se produce en niveles de esfuerzo pequeños.
- A la categoría D pertenecen los que la disnea se produce incluso en reposo.

En un hospital se está investigando la evolución en el grado funcional de insuficiencia cardíaca como consecuencia de un tipo determinado de intervención en el corazón. Para los pacientes en los que se procedería a realizar la intervención, se observó que el 10 % pertenecían a la categoría A, el 20 % a la B, el 30 % a la C y el 40 % a la D. Después de la intervención todos los pacientes de la categoría A siguieron en A; el 50 % de los de B pasó a A y el otro 50 % siguió en B; el 30 % de los de C pasó a A, el 40 % de los C pasó a B y el resto se quedó en C; mientras que sólo un 10 % de los que inicialmente estaban en D pasó a A, el 30 % pasó a B, el 40 % a C y el resto siguió en D.

Con ello:

a) Tomando al azar un paciente de dicho hospital que cumple los criterios para la intervención, ¿cuál es la probabilidad de que después de la misma esté en la categoría C?

b) Si sabemos que un paciente después de intervenido pertenece a la categoría B, ¿cuál es la categoría de la que resulta más probable que proceda? Justificar adecuadamente la respuesta.

c) Si el hospital trabaja con un total de 10000 pacientes intervenidos, ¿cuántos en ningún caso han pertenecido a la categoría C, ya sea antes o después de la intervención? ¿Y cuántos han pertenecido a la categoría A ya sea antes o después de la intervención?

**60. Para el diagnóstico de una enfermedad se utilizan dos test diagnóstico diferentes, A y B, de los que se sabe que:

- El test A da positivo en un 2 % de la población, mientras que el B da positivo en un 2,5 % de la población.
- Alguno de los dos test da positivo en el 2,8 % de la población.
- El valor predictivo positivo del test A es del 98 %, mientras que el valor predictivo positivo del B es del 95 %.
- El valor predictivo negativo del tes A es del 95 %.

Se pide:

a) ¿Cuánto vale la prevalencia de la enfermedad diagnosticada?

b) ¿Cuánto valen la sensibilidad y la especificidad, tanto de A como de B?

c) ¿Cuánto vale la probabilidad de que el test B dé positivo si previamente ha dado positivo A?

**61. La pioderma canina (infección bacteriana de la piel del perro que produce heridas y caída del pelo) es provocada por 4 tipos diferentes de bacterias: A) Staphylococcus intermedius en un 70 % de los casos, B) Staphylococcus scheleiferi en un 15 %, C) Staphylococcus pseudointermedius en un 10 %, y D) flora Gram Negativa en un 5 %. Para su tratamiento se utilizan dos tipos de antibióticos: 1) la cefalexina que elimina la infección en un 90 % de los casos del tipo A, en un 95 % del B, en un 100 % de C y en un 80 % de D, y 2) la amoxicilina que elimina la infección en un 80 % de A, en un 90 % de B, en un 95 % de C y en un 100 % de D. Justificando numéricamente todas las respuestas, se pide:

a) Si tenemos un perro con pioderma canina, ¿con qué antibiótico es más probable que se cure?

b) Si un estudio estadístico revela que la amoxicilina se aplica en un 70 % de los casos de pioderma canina y en el 30 % restante se aplica cefalexina, ¿qué probabilidad hay de que un perro con la enfermedad se cure al ser tratado?

c) Si sabemos que un perro infectado tratado con cefalexina no ha curado, ¿qué tipo de bacteria es más probable que tenga?

d) Si sabemos que un perro infectado tratado no ha curado, ¿qué tipo de tratamiento es más probable que haya recibido?

e) Si suponemos que los antibióticos actúan de forma completamente independiente y tenemos un perro infectado, ¿qué tipo de bacteria es más probable que tenga si es tratado consecutivamente con los dos antibióticos y no cura?

**62. Supongamos A y B factores de riesgo de una enfermedad. Si entre los que padecen la enfermedad: el 45 % tienen los dos factores de riesgo, el 20 % tienen A pero no se tiene B, el 25 % tienen B pero no A, y el 10 % no tienen ninguno de los dos; mientras que entre los que no padecen la enfermedad: el 10 % tienen los dos factores, el 30 % tienen A pero no B, el 20 % tiene B pero no A, y el 40 % no tienen ninguno. Si además sabemos que esa enfermedad la padecen el 10 % de los individuos de la población:

a) ¿Qué probabilidad hay de que se esté enfermo si se tienen los dos factores de riesgo?

b) ¿Qué probabilidad hay de que se tenga el factor A?

c) ¿Qué probabilidad hay de se esté enfermo si se tiene el factor A?

d) A efectos de padecer la enfermedad, ¿qué factor de riesgo es más peligroso? Justificar adecuadamente la respuesta.

**63. Al aplicar un test diagnóstico, cuya especificidad es 0,85, para detectar una enfermedad en una población, se sabe que el valor predictivo positivo es $0,40$ y la probabilidad de diagnóstico acertado es $0,84$. Calcular la prevalencia de la enfermedad.

Variables Aleatorias

64. Sea X una variable aleatoria discreta cuya ley de probabilidad es

X	4	5	6	7	8
$P(X=x)$	0,15	0,35	0,10	0,25	0,15

a) Calcular y representar gráficamente la función de distribución.

b) Obtener:

1) $P(X < 7{,}5)$.
2) $P(X > 8)$.
3) $P(4 \leq X \leq 6{,}5)$.
4) $P(5 < X < 6)$.

65. Sea la variable aleatoria X con la siguiente función de distribución:

$$F(x) = \begin{cases} 0 & \text{si} & x < 1 \\ 1/5 & \text{si} & 1 \leq x < 4 \\ 3/4 & \text{si} & 4 \leq x < 6 \\ 1 & \text{si} & x \geq 6 \end{cases}$$

Se pide:

a) Distribución de probabilidad.

b) Obtener:

1) $P(X = 6)$.
2) $P(X = 5)$.
3) $P(2 < X < 5{,}5)$.
4) $P(0 \leq X < 4)$.

*66. Se realiza un experimento aleatorio consistente en inyectar un virus a tres tipos de ratas y observar si sobreviven o no. Se comprueba que las probabilidades asociadas a los elementos del espacio muestral son:

E	VVV	VVM	VMV	VMM	MVV	MVM	MMV	MMM
P	0,1	0,1	0,1	0,15	0,15	0,1	0,1	0,2

donde V es vivir y M es morir. Se pide:

a) Construir la variable aleatoria que mida el número de ratas vivas y su función de probabilidad.

b) Calcular la función de distribución.

c) Calcular $P(X \leq 1)$, $P(X \geq 2)$ y $P(X = 1{,}5)$.

d) Calcular la media y la desviación típica.

67. La probabilidad de curación de un paciente al ser sometido a un determinado tratamiento es 0,85. Calcular la probabilidad de que en un grupo de 6 enfermos sometidos a tratamiento:

    a) se curen la mitad.

    b) se curen al menos 4.

68. 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 0'10.

    a) ¿Qué probabilidad hay de que ninguno se contagie?

    b) ¿Qué probabilidad hay de que al menos dos se contagien?

    c) ¿Cuántos se espera que contraigan la enfermedad?

69. La probabilidad de que al administrar una vacuna dé una determinada reacción es 0'001. Si se vacunan 2000 personas, ¿Cuál es la probabilidad de que aparezca una reacción adversa?

*70. El número medio de llamadas por minuto que llegan a una centralita telefónica es igual a 120. Hallar las probabilidades de los sucesos siguientes:

    a) A={durante 2 segundos lleguen a la centralita menos de 4 llamadas}

    b) B={durante 3 segundos lleguen a la centralita 3 llamadas como mínimo}

71. Un examen de tipo test consta de 10 preguntas con tres respuestas posibles para cada una de ellas. Se obtiene un punto por cada respuesta acertada y se pierde medio punto por cada pregunta fallada. Un alumno sabe tres de las preguntas del test y las contesta correctamente, pero no sabe las otras siete y las contesta al azar. ¿Qué probabilidad tiene de aprobar el examen?

*72. Se ha comprobado experimentalmente que una de cada 20 billones de células expuestas a un determinado tipo de radiación muta volviéndose cancerígena. Sabiendo que el cuerpo humano tiene aproximadamente 1 billón de células por kilogramo de tejido, calcular la probabilidad de que una persona de 60 kg expuesta a dicha radiación desarrolle cáncer. Si la radiación ha afectado a 3 personas de 60 kg, ¿cuál es la probabilidad de que desarrolle el cáncer más de una?.

*73. En un servicio de urgencias de cierto hospital se sabe que, en media, llegan 2 pacientes a la hora. Calcular:

    a) Si los turnos en urgencias son de 8 horas, ¿cuál será la probabilidad de que en un turno lleguen más de 5 pacientes?.

    b) Si el servicio de urgencias tiene capacidad para atender adecuadamente como mucho a 4 pacientes a la hora, ¿cuál es la probabilidad de que a lo largo de un turno de 8 horas el servicio de urgencias se vea desbordado en alguna de las horas del turno?.

**74. Sabiendo que la prevalencia de la isquemia cardíaca es del 1 %, y que la aplicación de un test diagnóstico para detectar la isquemia cardíaca tiene una sensibilidad del 90 %, y una especificidad del 95 %. Calcular:

    a) Los valores predictivos, tanto el positivo como el negativo.

    b) La probabilidad de diagnóstico acertado.

    c) Si tenemos un grupo de 10 enfermos de isquemia cardíaca, ¿cuál es la probabilidad de que diagnostiquemos la enfermedad a menos de 8?.

**75. Supongamos que para la detección de una enfermedad se utiliza un test diagnóstico tal que la probabilidad de positivo con dicho test vale 0,01, y sus valores predictivos positivo y negativo valen, respectivamente, 0,95 y 0,98. Se pide:

    a) ¿Cuál es la prevalencia de la enfermedad?.

    b) ¿Cuánto valen la sensibilidad y la especificidad del test?.

    c) Si aplicamos el test a 12 individuos enfermos, ¿qué probabilidad hay de que se equivoque en alguno de ellos?.

    d) Si aplicamos el test a 12 individuos, ¿que probabilidad hay de que acierte en todos?.

**76. Recientes estudios sobre la incidencia del cáncer en la población española afirman que un tercio de los individuos acabarán padeciendo cáncer en algún momento de su vida, aunque sólo la cuarta parte de los que lo padezcan acabarán falleciendo como consecuencia del mismo.

a) Si disponemos de un grupo de 10 españoles, ¿cuál es la probabilidad de que acaben muriendo por cáncer más de 2?

b) Si tenemos un grupo de 8 españoles con cáncer, ¿cuál es la probabilidad de que acaben muriendo por otras causas menos de 7?

c) Si entre los hombres la probabilidad de padecer cáncer es 2/3 de la probabilidad de las mujeres, y teniendo en cuenta que en la población española hay un 55 % de mujeres, si consideramos un grupo de 15 mujeres: ¿qué probabilidad habría de que terminasen muriendo por cáncer más de 3?

**77. Supongamos una enfermedad que se piensa que es causada por un único gen con dos alelos posibles: A y B; que se combinan de 4 formas diferentes: AA, AB, BA y BB. Estas formas diferentes de combinación de los alelos reciben el nombre de genotipos, y en este caso hay dos genotipos homocigóticos, formados por dos alelos iguales: AA y BB, y uno heterocigótico, formado por las dos combinaciones de alelos diferentes, AB y BA. Supongamos además que la probabilidad del alelo A en la población vale 0,95, la del B 0.05, y que se heredan de forma completamente independiente. También se sabe que la probabilidad de tener la enfermedad si se tiene un doble alelo B en el genotipo es del 90 %, del 50 % si se tiene un único alelo B y sólo del 10 % si no se tiene ningún alelo B. Con todo ello:

a) ¿Cuál es la probabilidad de cada uno de los 3 genotipos diferentes?

b) ¿Cuál es la probabilidad de tener la enfermedad en la población?

c) Si se sabe que un individuo tiene la enfermedad, ¿cuál es el genotipo más probable?

d) Si tenemos una pareja con 8 hijos y tal que el hombre es AA y la mujer la AB, ¿cuál es la probabilidad de que estén enfermos al menos 2 hijos?

**78. Se vacuna a 8 personas y se sabe que la probabilidad de que algunos de ellos sufra una reacción es de 0,48678:

a) Calcular la probabilidad de que una persona vacunada sufra reacción.

b) En el grupo de 8 personas vacunadas, ¿Cuál es la probabilidad de que sufran reacción más de 2?

c) Si en un centro de salud se vacunan 80 personas un día y tienen 6 dosis de corticoides para administrar a personas que hayan sufrido reacción, ¿qué probabilidad hay de que les falten dosis de corticoides ese día?

d) ¿Cuál sería el máximo número de personas que se podría vacunar para que la probabilidad de que hubiera alguna reacción fuera menor de 0,8?

Nota: Los tres últimos apartados se pueden contestar suponiendo que la probabilidad de que una persona vacunada sufra reacción es de 0,06

**79. Se sabe que en una población hay un 45 % de hombres, que el 12 % de los hombres terminan padeciendo alzheimer mientras que en las mujeres sólo lo padecen el 8 %. Además, los factores genéticos están presentes en un 30 % de los hombres que padecen alzheimer mientras que en las mujeres ese porcentaje se reduce al 20 %. Con todo ello, se pide:

a) ¿Cuál es la probabilidad de padecer alzheimer en esa población?

b) ¿Cuál es la probabilidad de que un individuo sea hombre si sabemos que ha padecido alzheimer? ¿Y la de que sea mujer si no lo ha padecido?

c) ¿Cuál es la probabilidad de padecer alzheimer por factores genéticos tanto en los hombres como en las mujeres como en la población?

d) ¿Cuál es la probabilidad de que extraído un individuo de la población sea un hombre sin alzheimer o una mujer con alzheimer debido a factores genéticos?

e) Si tomamos 200 individuos al azar de esa población, ¿qué probabilidad habría de que más de 2 sean mujeres con alzheimer causado por factores genéticos?

80. Una variable aleatoria continua X tiene una función de densidad dada por:

$$f(x) = \begin{cases} k(6-3x) & \text{si} \quad 0 \le x \le 2 \\ 0 & \text{si} \quad x < 0 \text{ ó } x > 2 \end{cases}$$

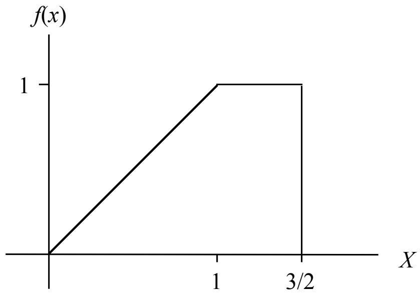
a) Determinar el valor de k.

b) Hallar $P(X \leq 1)$; $P(X > 2)$; $P(X = 1/4)$; $P(1/3 \leq X \leq 2/3)$.

c) Calcular μ y σ.

d) Hallar la función de distribución $F(x)$.

*81. Dada la función de densidad dada por la siguiente gráfica,



calcular:

a) $P(X < 1)$, $P(X > 0)$, $P(X = 1/4)$, $P(1/2 \leq X \leq 3/2)$.

b) Media y desviación típica.

**82. La variable aleatoria X con función de densidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x > 0; \\ 0, & \text{si } x \leq 0. \end{cases}$$

se dice que tiene una distribución exponencial de parámetro $\lambda > 0$ y se utiliza para medir tiempos de espera y de vida, y teniendo en cuenta que:

$$\int a \cdot e^{-ax} dx = -e^{-ax} + Cte$$

Y suponiendo que tenemos dos microorganismos diferentes cuyos tiempos de vida siguen distribuciones exponenciales, el primero de parámetro $\lambda = 0{,}02$ dias^{-1}, y el segundo con $\lambda = 0{,}03$ dias^{-1}, se pide:

a) ¿Cuál es la probabilidad de que el primero de los microorganismos viva entre 30 y 60 días?.

b) ¿Cuál es la probabilidad de que ambos microorganismos vivan menos 40 días?

c) ¿Cuál es la probabilidad de que alguno de los microorganismos viva menos de 40 días?.

83. Sea Z una variable aleatoria que sigue una distribución $N(0,1)$. Determinar el valor de t en cada uno de los siguientes casos:

a) El área entre 0 y t es 0,4783.

b) El área a la izquierda de t es 0,6406.

c) El área entre $-1{,}5$ y t es 0,2313.

84. Hallar las siguientes probabilidades:

a) $P(-2{,}4 \leq Z \leq -1{,}2)$ si Z es $N(0,1)$.

b) $P(|Z| > 1{,}2)$ si Z es $N(0,1)$.

c) $P(1{,}3 \leq X \leq 3{,}3)$ si X es $N(2,1)$.

d) $P(|X - 3| > 2)$ si Xes $N(3,4)$.

85. Entre los diabéticos, el nivel de glucosa en la sangre en ayunas, puede suponerse de distribución aproximadamente normal, con media 106 mg/100 ml y desviación típica 8 mg/100 ml.

a) Hallar $P(X \leq$120 mg/ 100 ml).

b) ¿Qué porcentaje de diabéticos tendrá niveles entre 90 y 120 mg/100 ml?

c) Encontrar un valor que tenga la propiedad de que el 25 % de los diabéticos tenga un nivel de glucosa X por debajo de dicho valor.

86. Se sabe que el nivel de colesterol en varones de más de 30 años sigue una distribución normal, de media 220 y desviación típica 30. Realizando un estudio sobre 20000 varones mayores de 30 años,

    a) ¿Cuántos se espera que tengan su nivel de colesterol entre 210 y 240?

    b) ¿Cuántos se espera que tengan su nivel de colesterol por encima de 250?

    c) ¿Cuál será el nivel de colesterol, por encima del cual se espera que esté el 20 % de la población?

*87. En una población con 40000 personas, se sabe que 2276 tienen entre 0.80 y 0.84 miligramos de bilirrubina por decilitro de sangre, y que 11508 tienen más de 0.84. Suponiendo que la concentración de bilirrubina en sangre sigue una distribución normal, se pide:

    a) Calcular su media y su desviación típica.
    Nota: En caso de no conseguir calcular la media y la desviación típica, tomar los valores $\mu = 0{,}8$ y $\sigma = 0{,}1$ y continuar con el ejercicio.

    b) Calcular el número de personas con más de 1 miligramo de bilirrubina por decilitro de sangre.

*88. Se supone que la tensión arterial de los habitantes de una población de 20000 habitantes sigue una distribución normal, cuya media es 13 y su rango intercuartílico 4. Se pide:

    a) ¿Cuántas personas tienen una tensión por encima de 16?.

    b) ¿Cuánto tendrá que disminuir la tensión de una persona que tiene 16 para situarse en el 40 % de la población con tensión más baja?.

**89. La probabilidad de que en un grupo de 5 individuos mayores de 70 años todos padezcan arterioesclerosis cerebral es de 12,5 por mil.

    a) ¿Cuál es la probabilidad de padecer la enfermedad entre los mayores de 70 años?.

    b) En un grupo de 1000 personas, ¿cuál es la probabilidad de que padezcan la enfermedad más de 450?.

**90. Si sabemos, por estudios previos, que las cepas que provocarán la gripe del siguiente otoño-invierno afectarán a un 20 % de la población:

    a) ¿Cuál es la probabilidad de que en una población de 10000 habitantes queden infectados menos de 1900?.

    b) Suponiendo que se vacunan los 10000 habitantes y sabiendo, por estudios previos, que la vacuna inmuniza al 98 % de los vacunados, ¿Cuál es la probabilidad de que queden sin inmunizar menos de 180?.

    c) De nuevo, suponiendo que se han vacunado los 10000 habitantes y teniendo en cuenta que, por estudios previos, la vacuna produce reacciones alérgicas en uno de cada 5000 casos, ¿cuál es la probabilidad de que se produzca alguna reacción alérgica en dicha población?.

**91. Una solución contiene virus bacteriófagos T_4 en una concentración de $4 \cdot 10^6$ por mm^3. En la misma solución hay $2 \cdot 10^6$ bacterias por mm^3. Suponiendo que todos los virus infectan bacterias y que se distribuyen al azar entre las mismas, se pide:

    a) ¿Cuál es el porcentaje de bacterias que no están infectadas por el virus?.

    b) ¿Qué porcentaje de bacterias tendrá al menos 2 virus fijados sobre ellas?.

    c) Si tomamos un volumen pequeño de dicha solución en el que hay 4 bacterias, ¿cuál es la probabilidad de que alguna esté infectada?.

    d) Si tomamos un volumen en el que hay 10000 bacterias, ¿cuál es la probabilidad de que estén infectadas al menos 8600?.

**92. En un estudio estadístico se pretende ver si están relacionados el peso del recién nacido, PRN en kg, con el peso de la madre justo después del nacimiento, PM en kg. Para ello, con una muestra de tamaño 10, los datos obtenidos fueron:

PM	65,1	70,2	62,4	58,8	56,3	84,2	91,3	63,4	70,3	61,6
PRN	3,8	2,7	3,5	2,9	3,2	3,6	3,1	2,7	3,4	3,0

a) Calcular la recta de regresión del peso del recién nacido en función del peso de la madre.

b) ¿Es el modelo lineal adecuado para ajustar la nube de puntos?. Justificar adecuadamente la respuesta.

c) Según el modelo lineal, ¿qué peso esperamos que tenga una madre cuyo hijo ha pesado 2,4 Kg al nacer?

d) ¿Qué media es más representativa, la de pesos de las madres o la de los hijos?

e) ¿Qué peso sería relativamente más alto, el de una madre de 70 Kg o el de un hijo de 3,3 Kg?

f) Suponiendo que la variable peso del recién nacido sigue una distribución normal de media y desviación típica las calculadas en apartados anteriores, ¿qué probabilidad hay de que un recién nacido tenga un peso entre 3 y 3,5 Kg? ¿Cuánto vale el percentil 90 de la distribución?

**93. De una determinada sustancia producida por el organismo humano se sabe que el percentil 80 de su concentración en sangre vale $3,4$ microgramos por decilitro y su percentil 30 vale $2,3$ microgramos por decilitro. Suponiendo que la concentración en sangre sigue una distribución normal:

a) Calcular la media y la desviación típica de la distribución.

b) Si se considera que dicha sustancia también puede utilizarse como dopante si se inyecta por vía intravenosa, y se pone como valor límite para considerar que un individuo va dopado 4 microgramos por decilitro de sangre, entonces después de analizar un grupo de 10000 individuos no dopados, ¿cuántos se concluirá que sí que lo están?

c) Si la concentración de la sustancia en los individuos dopados sigue una distribución normal de media $4,5$ y desviación típica $0,3$ microgramos por decilitro y se sabe que el 10 % de los individuos analizados se dopan, ¿cuál será la probabilidad de que se diagnostique adecuadamente a una persona como dopada o no dopada al considerar lo 4 microgramos por decilitro como punto de corte?

**94. En un laboratorio se está probando el efecto que sobre el tiempo de vida de los ratones tiene una dieta baja en grasas. Para ello, se dispone de dos grupos: uno que sigue la dieta tradicional y el otro con la dieta baja en grasas, y se supone que el tiempo de vida en ambos grupos sigue distribuciones normales de diferente media pero igual desviación típica. Si el 20 % de los ratones con la dieta normal vive más de 12 meses y el 5 % menos de 8 meses, mientras que el 85 % de los que siguen la dieta baja en grasas vive más de 11 meses, se pide:

a) ¿Cuánto vale la media y la desviación típica del tiempo de vida de los ratones que siguen la dieta baja en grasas?

b) Si en laboratorio hay un 40 % de ratones con la dieta normal y un 60 % con la dieta baja en grasas, ¿cuál es la probabilidad de que un ratón tomado al azar muera antes de 9 meses?

**95. Por estudios realizados se sabe que la probabilidad de padecer cáncer de pulmón si se ha fumado durante una cantidad de años x viene dada por la función logística (también llamada curva sigmoidea):

$$P(x) = \frac{e^{-5+0,2\cdot x}}{1+e^{-5+0,2\cdot x}}$$

a) Si tenemos un grupo de 5 fumadores tal que todos ellos llevan fumando 20 años, ¿cuál es la probabilidad de que haya desarrollado cáncer más de uno?

b) Si tenemos un grupo de 1000 fumadores que llevan fumando 30 años, ¿cuál es la probabilidad de que hayan desarrollado cáncer menos de 710?

c) Si tenemos un grupo de 1000 fumadores que llevan fumando 1 mes, ¿cuál es la probabilidad de que hayan desarrollado cáncer menos de 2?

d) Si tomamos una persona al azar que comienza a fumar en el momento de empezar el seguimiento, ¿cuánto tiempo debe seguir fumando para que la probabilidad de que desarrolle cáncer sea igual a 0,5?

**96. El ácido úrico está presente en la sangre de los individuos sanos siguiendo una distribución normal de media 5,4 mg/dl y desviación típica 0,6 mg/dl. Sin embargo, en individuos que padecen gota (un tipo de artritis que ocurre cuando el ácido úrico se acumula en la sangre y causa inflamación articular) la distribución también es normal pero de media 7,0 mg/dl y desviación típica 0,3 mg/dl.

a) Supongamos que para diagnosticar la enfermedad se utiliza un simple análisis de sangre y que se concluye que el individuo padece gota siempre que su nivel de ácido úrico sea superior a 6,5 mg/dl. ¿Qué sensibilidad y qué especificidad presentaría dicho test diagnóstico?

b) Supongamos una población con un 90 % de sanos y que para diagnosticar la enfermedad se utiliza un simple análisis de sangre y se concluye que el individuo padece gota siempre que su nivel de ácido úrico esté por arriba del percentil 95 de los sanos. ¿Cuánto valdrían los valores predictivos, el positivo y el negativo del test?

c) Supongamos que se sabe que en los sanos el ácido úrico sigue la distribución comentada en el enunciado, mientras que en los enfermos la media es 7,0 mg/dl pero la desviación típica es desconocida. ¿Cuánto vale la desviación típica si se sabe que el percentil 90 de los sanos coincide con el percentil 15 de los enfermos?

**97. Suponiendo que el dolor que experimenta una persona en una sesión de rehabilitación de una lesión de codo, recogido en una escala VAS (escala analógica visual), sigue una distribución normal de media 6,2 y desviación típica desconocida, se pide:

a) Calcular la desviación típica sabiendo que el 90 % de los pacientes tiene un dolor por debajo de 7,3.

Nota: para los siguientes apartados se puede utilizar la desviación típica igual a 0,5.

b) Si se han tratado 900 pacientes, ¿cuántos habrán tenido un nivel de dolor entre 5,5 y 6,5?

c) ¿Cuánto valen los percentiles 5 y 95 del nivel de dolor?

d) Si una semana se atiende a 100 pacientes en rehabilitación de esa lesión de codo, ¿cuál es la probabilidad de que menos de 2 tengan un dolor por arriba de 7,5?

**98. En los hombres, el tiempo de recuperación de una fractura de ligamento cruzado anterior de la rodilla sigue una distribución normal de media 6 meses y desviación típica desconocida. En las mujeres, también sigue una normal pero de media 5 meses y desviación típica desconocida.

a) Sabiendo que el 70 % de las mujeres y el 10 % de los hombres se recuperan antes de 5,3 meses, ¿cuánto vale la desviación típica de las mujeres? ¿Y la de los hombres?

Nota: para los siguientes apartados se puede tomar la desviación típica de las mujeres igual a 0,3 meses y la de los hombres 0,6 meses.

b) Si en total en la población hay un 55 % de mujeres, ¿cuál será la probabilidad de que un individuo cualquiera de esa población tarde en recuperarse de la fractura entre 5,5 y 6 meses?

c) El percentil 90 de las mujeres, ¿a qué percentil correspondería en los hombres?

**99. En un experimento se ha comprobado que los individuos sanos responden en promedio a un 80 % de los estímulos de un determinado tipo aplicados en una zona concreta de la piel, mientras que los pacientes con una patología neurológica sólo responden en promedio a un 40 % de esos estímulos.

a) Si tenemos un grupo de 6 pacientes sanos, ¿qué probabilidad hay de que respondan al estímulo al menos 4?

b) Si tenemos 4000 pacientes sanos, ¿qué probabilidad hay de que respondan al estímulo al menos 3240?

c) Si para clasificar a un individuo como sano o enfermo lo sometemos a 10 estímulos y decimos que está sano si responde a 6 o más, ¿cuál es la sensibilidad y la especificidad del test?

100. Calcular:

a) $P(T \leq 1{,}476)$ si $T \sim T(5)$.

b) $P(T \geq 0{,}69)$ si $T \sim T(16)$.

c) El valor t_0 tal que $P(T < t_0) = 0{,}995$, con $T \sim T(12)$.

d) El valor t_0 tal que $P(T > t_0) = 0{,}01$, con $T \sim T(8)$.

101. Calcular:

a) $P(X \leq 5{,}23)$ si $X \sim \chi^2(12)$.

b) $P(X \geq 1{,}65)$ si $X \sim \chi^2(8)$.

c) El valor x_0 tal que $P(X < x_0) = 0{,}995$, con $X \sim \chi^2(18)$.

d) El valor x_0 tal que $P(X > x_0) = 0{,}25$, con $X \sim \chi^2(7)$.

102. Calcular:

a) El valor f_0 tal que $P(F < f_0) = 0{,}9$, con $F \sim F(12, 8)$.

b) El valor f_0 tal que $P(F > f_0) = 0{,}025$, con $F \sim F(5, 7)$.

Intervalos de Confianza

103. Una muestra aleatoria de tamaño 81 extraída de una población normal con $\sigma^2 = 64$, tiene una $\overline{x} = 78$. Calcular el intervalo de confianza del 95 % para μ.

104. Para determinar si un pescado es o no apto para el consumo por su contenido en Hg (mercurio), se realizan 15 valoraciones obteniendo una media de 0,44 ppm (partes por millón) de Hg, y una desviación típica de 0,08 ppm. Calcular los límites de confianza para la media, a un nivel de significación $\alpha = 0{,}1$.

105. Se obtuvieron cinco determinaciones del pH de una solución con los siguientes resultados: 7.90, 7.85, 7.89, 7.86, 7.87. Hallar unos límites de confianza de la media de todas las determinaciones del pH de la misma solución, al nivel de significación $\alpha = 0{,}01$.

106. Se desea saber cuál debe ser el tamaño muestral mínimo de una muestra para poder realizar la estimación de la tasa media de glucosa plasmática de una determinada población, con un nivel de confianza 0'95 y pretendiendo una amplitud de 2'5 mg.

NOTA: En una muestra previa de tamaño 10 se obtuvo una desviación típica de 10 mg.

107. Para que un fármaco sea efectivo, la concentración de un determinado principio activo debe ser 20 mg/mm^3. Se recibe un lote de dicho fármaco y se analizan 10 para medir la concentración del principio activo, obteniendo los resultados siguientes:

$$17{,}6 - 19{,}2 - 21{,}3 - 15{,}1 - 17{,}6 - 18{,}9 - 16{,}2 - 18{,}3 - 19 - 16{,}4.$$

En vista de los resultados, ¿podremos rechazar el lote con una confianza 0'95 de no equivocarnos?

*108. En un estudio sobre el consumo anual de litros de cerveza entre la población de menores de 18 años de una ciudad se obtuvo la siguiente muestra:

$$42, 16, 60, 29, 7, 20, 30, 25, 38, 5.$$

Se pide:

a) Calcular el intervalo de confianza del 95 % para la media. Si se considera que un consumo medio por encima de 40 litros es peligroso, ¿existen pruebas significativas para afirmar que la población de partida no está en peligro?

b) ¿Qué tamaño muestral mínimo hubiese sido necesario para conseguir un intervalo de confianza de amplitud 5?

109. Se realizó un estudio sobre el contenido de principio activo de un determinado fármaco a partir de una muestra, determinándose los siguientes resultados en mg/cm^3:

$$46{,}4 - 46{,}1 - 45{,}8 - 47{,}0 - 46{,}1 - 45{,}9 - 45{,}8 - 46{,}9 - 45{,}2 - 46{,}0.$$

Obtener un intervalo de confianza del 95 % para la varianza del contenido de principio activo de dicho fármaco, suponiendo que sigue una distribución normal.

*110. Para determinar el nivel medio de colesterol en la sangre de una población, se realizaron análisis sobre una muestra de 8 personas, obteniéndose los siguientes resultados:

$$196 - 212 - 188 - 206 - 203 - 210 - 201 - 198$$

Hallar intervalos de confianza para la media y la varianza de nivel de colesterol con un nivel de significación 0.1, suponiendo que el nivel de colesterol en la población sigue una distribución normal.

111. Para determinar la concentración media de albúmina en la sangre se realizaron mediciones sobre un grupo experimental obteniéndose los siguientes resultados, expresados en g/l: 38-42-46-37-49-42-40-36. Obtener un intervalo de confianza para la varianza de la población con un nivel de significación 0.05.

112. Leemos en una revista médica que la cuarta parte de los cancerosos de cierto tumor de estómago presentan vómitos, con una precisión o tolerancia del 10 % y con una confianza del 99 %. ¿Con cuántos pacientes se ha realizado el estudio?

*113. Un país está siendo afectado por una epidemia de un virus. Para valorar la gravedad de la situación se tomaron 40 personas al azar y se comprobó que 12 de ellas tenían el virus. Determinar el intervalo de confianza para el porcentaje de infectados con un nivel de significación 0.05.

114. Se desea obtener un intervalo de confianza del 95 % para la diferencia de marcas obtenidas por chicos y chicas en una prueba física. Se toma una muestra de 50 chicas y 75 chicos, obteniendo las chicas una marca media de 76 y los chicos de 82. Además, se conocen las desviaciones típicas de las marcas obtenidas en las poblaciones de chicas y chicos, que son 6 y 8 respectivamente.

115. Se está ensayando un nuevo procedimiento de rehabilitación para una cierta lesión. Para ello se trataron nueve pacientes con el procedimiento tradicional y otros nueve con el nuevo, y se midieron los días que tardaron en recuperase, obteniéndose los siguientes resultados:

Método tradicional: 32-37-35-28-41-44-35-31-34
Método nuevo: 35-31-29-25-34-40-27-32-31

Se desea obtener un intervalo de confianza del 95 % para la diferencia de las medias del tiempo de recuperación obtenido con ambos procedimientos. Se supone que los tiempos de recuperación siguen una distribución normal, y que las varianzas son aproximadamente iguales para los dos procedimientos.

116. En un hospital pediátrico se comprobó que de 200 niños con un determinado síndrome, 48 murieron antes de cumplir un año de edad, mientras que sólo 25 de 125 niñas con el mismo síndrome murieron. ¿Se puede afirmar con cierta seguridad que el síndrome es más letal en los niños que en las niñas?

117. Se ha realizado un estudio para investigar el efecto del ejercicio físico en el nivel de colesterol en la sangre. En el estudio participaron once personas, a las que se les midió el nivel de colesterol antes y después de desarrollar un programa de ejercicios. Los resultados obtenidos fueron los siguientes

Persona	Nivel previo	Nivel posterior
1	182	198
2	232	210
3	191	194
4	200	220
5	148	138
6	249	220
7	276	219
8	213	161
9	241	210
10	280	213
11	262	226

Hallar un intervalo de confianza del 90 % para la diferencia del nivel medio de colesterol antes y después del ejercicio.

118. Dos químicos A y B realizan 14 y 16 determinaciones, respectivamente, de plutonio. Los resultados obtenidos se muestran en la siguiente tabla

A		B	
263.36	254.68	286.53	254.54
248.64	276.32	284.55	286.30
243.64	256.42	272.52	282.90
272.68	261.10	283.85	253.75
287.33	268.41	252.01	245.26
287.26	282.65	275.08	266.08
250.97	284.27	267.53	252.05
		253.82	269.81

Se pide:

a) Calcular intervalos de confianza del 95 % de confianza para cada caso.

b) ¿Se puede decir que existen diferencias significativas en la media?

*119. Un equipo de investigación está interesado en ver si una droga reduce el colesterol en la sangre. Con tal fin toma una muestra de 10 pacientes y determina el contenido de colesterol antes y después del tratamiento. Los resultados expresados en miligramos por cada 100 mililitros son los siguientes:

Paciente	1	2	3	4	5	6	7	8	9	10
Antes	217	252	229	200	209	213	215	260	232	216
Después	209	241	230	208	206	211	209	228	224	203

Se pide:

a) Construir la variable Diferencia que recoja la diferencia entre los niveles de colesterol antes y después del tratamiento, y calcular el intervalo de confianza con $1-\alpha = 0{,}95$ para dicha variable.

b) A la vista del intervalo anterior, ¿se concluye que la aplicación de la droga ha disminuido el nivel de colesterol en la sangre?

*120. Se está ensayando un nuevo procedimiento de rehabilitación para una cierta lesión. Se sabe que de 80 deportistas tratados con el procedimiento tradicional, se recuperaron perfectamente 26, mientras que de los 20 tratados con el nuevo procedimiento se han recuperado 11. ¿Se puede afirmar con una confianza del 95 % que el nuevo procedimiento es mejor que el tradicional?

*121. En una muestra aleatoria de 200 personas, 114 están a favor de la fluoración de las aguas. Se pide:

a) Hallar el intervalo de confianza del 96 % para la fracción de la población que está a favor de la fluoración de las aguas.

b) ¿Qué tamaño mínimo de muestras habría que tomar para tener una confianza del 96 % de que la proporción muestral difiere menos de 0.02 de la proporción real de la población?

*122. Para ver si una campaña de publicidad sobre un fármaco ha influido en sus ventas, se tomó una muestra de 8 farmacias y se midió el número de fármacos vendidos durante un mes, antes y después de la campaña, obteniéndose los siguientes resultados:

Antes	147	163	121	205	132	190	176	147
Después	150	171	132	208	141	184	182	145

Obtener la variable diferencia y construir un intervalo de confianza para la media de la diferencia con un nivel de significación 0.05. ¿Existen pruebas suficientes para afirmar con un 95 % de confianza que la campaña de publicidad ha aumentado las ventas?

*123. Para comparar la eficacia de dos tratamientos A y B en la prevención de repeticiones de infarto de miocardio, se aplicó el tratamiento A a 80 pacientes y el B a 60. Al cabo de dos años se observó que habían sufrido un nuevo infarto 14 pacientes de los sometidos al tratamiento A y 15 de los del B. Se pide:

a) Construir un intervalo de confianza del 95 % para la diferencia entre las proporciones de personas sometidas a los tratamientos A y B que no vuelven a sufrir un infarto.

b) A la vista del resultado obtenido, razonar si con ese nivel de confianza puede afirmarse que uno de los tratamientos es más eficaz que el otro.

**124. Se quiere probar si la cirrosis hepática hace variar el índice de colinesterasa en suero. Se eligen 2 muestras aleatorias e independientes, una primera de 60 individuos normales, con media $1,6$ y desviación típica $0,3$, y la segunda de 50 individuos cirróticos, con media 1,1 y desviación típica $0,4$. ¿Podemos concluir que existen diferencias significativas, con un 99 % de confianza, entre las medias de la colinesterasa en individuos normales e individuos cirróticos?.

**125. En un análisis de obesidad dependiendo del hábitat en niños menores de 5 años, se obtienen los siguientes resultados:

	Casos analizados	Casos con sobrepeso
Hábitat rural	1150	480
Hábitat urbano	1460	660

Se pide:

a) Construir un intervalo de confianza, con un nivel de significación $0,01$, para la proporción de niños menores de 5 años con sobrepeso en el hábitat rural. Igualmente para el hábitat urbano.

b) Construir un intervalo de confianza, con un nivel de confianza del 95 %, para la diferencia de proporciones de niños menores de 5 años con sobrepeso entre el hábitat rural y el urbano. A la vista del resultado obtenido, ¿se puede concluir, con un 95 % de confianza, que la proporción de niños menores de 5 años con sobrepeso depende del hábitat?.

**126. Un grupo de investigadores obtuvo datos acerca de las concentraciones de amilasa en el suero de muestras de individuos sanos y de individuos hospitalizados, con el objetivo de determinar si la concentración media es, o no, diferente en ambas poblaciones. Las concentraciones, en unidades/ml, en 10 individuos sanos fueron:

100 103 96 93 91 104 93 99 88 91

Y en 12 individuos enfermos fueron:

118 115 101 104 116 114 112 113 117 123 119 121

Suponiendo que la concentración de amilasa en suero sigue una distribución normal, tanto en individuos sanos como hospitalizados, y que las varianzas son desconocidas pero iguales, se pide:

a) Calcular el intervalo de confianza para la diferencia de medias con un nivel de confianza del 95 %.

b) ¿A qué conclusión deben llegar los investigadores sobre la igualdad o no de la concentración de amilasa?. Justificar la respuesta.

**127. Se ha realizado un estudio con 1000 mujeres que han dado a luz recientemente, elegidas al azar entre los registros de los diferentes hospitales de la comunidad de Madrid, para saber si un nuevo protocolo (visitas al médico y consumo de ciertos fármacos) resulta más efectivo para prevenir las infecciones (ya sean pre, intra o postparto). Del total, 750 han seguido el protocolo habitual, entre las cuales 35 han sufrido algún tipo de infección; mientras que 250 han seguido el protocolo nuevo y 9 de ellas han padecido alguna infección. ¿Se puede afirmar, con un 95 % de confianza, que la proporción de mujeres que ha tenido algún tipo de infección ha sido diferente según el protocolo utilizado?.

**128. Se supone que el tiempo, en años, de incubación del virus que provoca el SIDA depende linealmente de la cantidad suministrada de un cierto fármaco retrovírico, en mg/día. Para ello, se realizó el seguimiento a 10 pacientes desde que se les confirmó la presencia del virus hasta que se produjeron los primeros síntomas de inmunodeficiencia, anotándose tanto el tiempo, X, como la dosis del fármaco que se le había suministrado, Y:

X	3,2	4,6	1,1	5,2	0,4	6,3	1,8	3,4	4,2	7,3
Y	30	60	20	30	10	60	40	40	40	60

a) Calcular la recta de regresión del tiempo de incubación en función de la cantidad de fármaco suministrada.

b) Calcular el coeficiente de determinación lineal e interpretarlo.

c) Si utilizamos la notación $y = a + bx$ para la recta de regresión, donde a es la ordenada en el origen y b es la pendiente de la recta, hay que tener en cuenta que tanto a como b son estadísticos muestrales, es decir, su valor será diferente dependiendo de la muestra con la que trabajemos, y lo que habitualmente damos es una estimación puntual de los mismos. Para una muestra concreta, si llamamos a las estimaciones puntuales: $\widehat{a}$ y $\widehat{b}$, la ecuación de la recta de regresión es: $y = \widehat{a} + \widehat{b}x$, y, como ya sucede con el resto de estadísticos muestrales, también podríamos dar una estimación por intervalo con nivel de significación α, que en el caso de la pendiente tiene la forma:

$$\left(\hat{b} - t(n-2)_{\alpha/2} \frac{\sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \hat{a}\sum_{i=1}^{n} y_i - \hat{b}\sum_{i=1}^{n} x_i y_i}{n-2}}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}, \hat{b} + t(n-2)_{\alpha/2} \frac{\sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \hat{a}\sum_{i=1}^{n} y_i - \hat{b}\sum_{i=1}^{n} x_i y_i}{n-2}}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \right)$$

Utilizando la fórmula anterior, calcular el intervalo, con un 95 % de confianza, para la pendiente de la recta de regresión obtenida en el primera apartado del problema.

**129. El número de muertos en accidentes de carretera durante el 2005 en España fue el siguiente

Mes	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Muertos	272	269	293	274	308	314	374	354	310	318	269	297

Se pide:

a) Calcular el intervalo de confianza del 99 % para el número medio de muertos al mes.

b) ¿Existen pruebas significativas, con un 99 % de confianza, para afirmar que el número medio de muertes en los meses de verano (junio, julio, agosto y septiembre) es superior al del resto del año? Suponer que las varianzas poblacionales son iguales.

**130. Los fabricantes de un producto energético a base de glúcidos de asimilación rápida, especialmente diseñado para corredores de fondo, afirman que si se toma adecuadamente:

- Aumenta la proporción de corredores que terminan el maratón.
- En los que lo terminan, disminuye el tiempo medio para concluirlo.
- En los que no lo terminan, aumenta la distancia final recorrida hasta el momento en que se detienen.

Para comprobar dichas afirmaciones, se ha realizado un estudio en el reciente maratón popular de Madrid en el que se ha trabajado con 100 corredores divididos en 2 grupos de forma completamente aleatoria: 50 toman un placebo y otros 50 el producto energético. Los resultados obtenidos han sido:

- Entre los que tomaron el placebo no terminaron el maratón un 36 %, mientras que entre los que tomaron el producto no lo terminaron un 20 %.
- Entre los que terminaron, el tiempo medio de los que tomaron el placebo fue de 3 horas 47 minutos con una desviación típica de 23 minutos mientras que el tiempo medio de los que tomaron el producto fue de 3 horas 36 minutos con una desviación típica de 18 minutos.
- Entre los que no lo terminaron, la distancia final recorrida en media por los que tomaron el placebo fue de $32,45$ Km con una desviación típica de $2,30$ Km mientras que los que tomaron el producto recorrieron una distancia media de $31,68$ Km con una desviación típica de $3,10$.

Teniendo en cuenta los datos anteriores, trabajando con un 95 % de confianza y considerando distribuciones normales con varianzas poblacionales que pueden suponerse iguales, ¿se puede concluir que los fabricantes del producto tienen razón en alguna de sus afirmaciones? Justificar adecuadamente la respuesta.

**131. Se ha determinado el antígeno prostático específico (APE), en nanogramos por decilitro de sangre, en pacientes que tenían cáncer de próstata, tanto en los que tuvieron una progresión rápida de la enfermedad (grupo A) como en los que no progresó el cáncer (grupo B). Los resultados obtenidos fueron los siguientes:

APE Grupo A	7,5	5,6	5,3	9,3	6,2	4,5	6,0	8,3	2,1	5,8
APE Grupo B	5,9	1,5	12,3	5,4	9,2	7,6	7,4	7,7	9,5	11,5

Suponiendo que los datos siguen distribuciones normales aunque con diferentes varianzas poblacionales, y considerando un nivel de confianza del 90 %, ¿existen evidencias para poder afirmar que el APE tiene un valor pronóstico significativo para determinar si el cáncer de próstata progresará o no? Justificar adecuadamente la respuesta.

**132. En el servicio de urgencias de un hospital se atendieron un día a unas personas cuyas edades se indican en la siguiente tabla:

Edad	Hombres	Mujeres
[0, 20)	5	4
[20, 40)	2	3
[40, 60)	4	5
[60, 80)	12	17
[80, 100)	9	21

a) Calcular las edades medias de los hombres y de las mujeres que acudieron a urgencias ese día e indicar cuál de ellas es más representativa.

b) ¿Cuánto vale el tercer cuartil de edad en los hombres?

c) Una mujer de 90 años que acudió a urgencias ese día, ¿qué percentil de edad le correspondería dentro del grupo de mujeres?

d) Considerando que todas las edades de cada clase son iguales a sus marcas de clase y que los datos de ese día constituyen una muestra de las edades de las personas que acuden al servicio de urgencias del hospital, calcular el intervalo de confianza para la edad media de los hombres que acuden al servicio de urgencias de ese hospital con un nivel de significación 0,08.

e) En las mismas condiciones del apartado anterior, calcular el intervalo de confianza para la diferencia entre las edades medias de hombres y mujeres que acuden al servicio de urgencias de ese hospital, con un nivel de significación 0,08, y a partir de él razonar si se puede rechazar que las edades medias de los hombres y las mujeres que acuden al servicio de urgencias pueden ser iguales.

**133. Se dispone de 2 grupos de personas, unos sanos y otros enfermos de gota, y se les realiza un análisis de sangre para cuantificar el ácido úrico en mg/dl, obteniendo los resultados que aparecen en la siguiente tabla:

Grupo	n	$\bar{x}$	s
Sanos	15	5,1	0,9
Con gota	27	6,2	0,5

a) Suponiendo que el contenido de ácido úrico en personas sanas sigue una distribución normal y trabajando con un 99 % de confianza, ¿cuánto vale la media poblacional del contenido de ácido úrico en sangre en las personas sanas? ¿Se puede concluir que dicho contenido es inferior a 5,8 mg/dl?

b) Con un 95 % de confianza, ¿hay diferencias significativas entre las medias de sanos y enfermos de gota? Suponer que el ácido úrico en sangre sigue distribuciones normales en los dos colectivos y que las varianzas poblaciones, aunque desconocidas, pueden suponerse iguales.

c) Suponiendo invariables el resto de datos de la tabla del enunciado y de nuevo suponiendo normalidad e igualdad de varianzas poblacionales, ¿qué desviación típica muestral mínima debería haber en el grupo de sanos para que no hubiese diferencias significativas con un 95 % de confianza?

**134. El nivel de hemoglobina A1c en sangre (HgbA1c) nos da una idea de si los niveles de glucosa en sangre durante los últimos 3 meses han podido ser elevados. Por ello, la prueba de la hemoglobina A1c se aplica para saber si los diabéticos han cumplido o no con los protocolos de administración de insulina que se les han aconsejado. En la siguiente tabla aparecen los resultados de la cantidad de hemoglobina A1c (en % del total de hemoglobina) de 10 individuos diabéticos, 5 hombres y 5 mujeres, en el momento inicial del tratamiento con insulina (HgbA1c a) y 3 meses después del mismo (HgbA1c b):

Sexo	H	H	H	H	H	M	M	M	M	M
HgbA1c a	6,7	7,4	9,2	9,6	7,4	8,1	10,8	7,1	7,9	10,8
HgbA1c b	7,0	7,4	8,6	8,1	6,8	7,0	8,5	7,7	9,7	7,7

a) ¿En qué muestra es más representativa la media de hemoglobina Ac1 al comienzo del estudio, en la de hombres o en la de mujeres? Justificar adecuadamente la respuesta.

b) Independientemente del sexo, ¿cuánto vale el coeficiente de apuntamiento de la variable diferencia (HgbA1c a - HgbA1c b)? Interpretar el resultado.

c) Suponiendo que el nivel de hemoglobina A1c al comienzo del estudio sigue distribuciones normales tanto en hombres como en mujeres, y que las varianzas poblacionales, aún siendo desconocidas, pueden suponerse iguales, ¿hay diferencias significativas entre hombres y mujeres en el nivel de hemoglobina A1c al comienzo del estudio con un 95 % de confianza? Justificar adecuadamente la respuesta.

d) Con un 99 % de confianza e independientemente del sexo, ¿hay diferencias significativas entre los niveles de hemoglobina A1c al comienzo y a los 3 meses del estudio? Justificar adecuadamente la respuesta.

**135. Para comprobar el posible efecto que el uso (y abuso) del móvil puede tener en la nota final obtenida en la asignatura de Bioestadística, un profesor divide al azar a sus alumnos en 2 grupos diferentes en los que aplica la misma metodología docente, pero en el primero prohíbe terminantemente el uso del móvil y en el segundo lo permite para lo que el alumno quiera tenerlo (excepto en los exámenes).

Los resultados obtenidos fueron:

- Grupo sin móvil: de los 40 alumnos aprobaron 25; la nota media global del grupo fue de 6,0 con desviación típica 1,2; la nota media de los aprobados fue de 7,1 con desviación típica 0,7.
- Grupo con móvil: de los 42 alumnos aprobaron 17; la nota media global del grupo fue de 5,2 con desviación típica 1,3; la nota media de los aprobados fue de 6,8 con desviación típica 0,6.

Se pide determinar con un 90 % de confianza si hay un efecto significativo del uso del móvil en:

a) La proporción de aprobados.

b) En la nota media global.

c) En la nota media de los que aprueban. Para este apartado, considerar que las varianzas poblacionales, aunque desconocidas, pueden suponerse iguales.

**136. En el servicio de cardiología de un hospital se está ensayando una nueva técnica para el tratamiento de las arritmias y se quiere comprobar si hay o no diferencias significativas con la técnica convencional. Para ello, se ha trabajado con 160 pacientes, 80 con cada técnica y se ha comprobado que:

- A los 6 meses de tratamiento no se habían reproducido las arritmias en 70 pacientes de la técnica nueva y en 60 de la convencional.
- Entre los que sí que se habían reproducido las arritmias a los 6 meses, el tiempo transcurrido hasta el primer episodio de arritmias tuvo una media de 60 días y una desviación típica de 10 días en los tratados con la convencional, y una media de 80 días y una desviación típica de 15 días en los tratados con la nueva.

Se pide:

a) Con un 99 % de confianza, ¿hay diferencias significativas en la proporción de pacientes en los que se reprodujeron las arritmias antes de 6 meses?

b) Considerando los pacientes en los que sí que se reprodujeron las arritmias antes de los 6 meses de tratamiento, con un 95 % de confianza, ¿hay diferencias significativas en el tiempo medio transcurrido hasta el primer episodio? Considerar que las desviaciones típicas poblacionales, aunque desconocidas, pueden suponerse iguales.

c) Con un 90 % de confianza y considerando los pacientes tratados con la técnica convencional en los que se reprodujeron las arritmias antes de los 6 meses, ¿se puede considerar que el tiempo medio hasta la primera arritmia fue distinto de 70 días?

**137. Un investigador piensa que el tiempo de recuperación de una fractura de ligamento lateral de la rodilla puede depender del sexo del individuo. Para comprobarlo tomó pacientes de edades similares y anotó el tiempo de recuperación en meses y su sexo, obteniendo:

Sexo	H	H	H	H	H	H	M	M	M	M	M	M
Tiempo	5,6	6,7	6,2	7,1	5,9	6,5	6,3	5,4	5,3	6,1	5,1	5,6

a) ¿Se puede concluir con un 80 % de confianza que la media del tiempo de recuperación en los hombres es distinta de 6,5 meses?

b) ¿Hay diferencias significativas con un 50 % de confianza entre las medias de los hombres y las mujeres? Suponer que la varianzas poblacionales son desconocidas pero iguales.

c) Calcular el intervalo con un 90 % de confianza para la proporción de pacientes cuyo tiempo de recuperación es superior a 6 meses.

Contrastes de Hipótesis

138. Se sabe que una vacuna que se está utilizando al cabo de dos años sólo protege al 60 % de las personas a las que se administró.

Se desarrolla una nueva vacuna, y se quiere saber si al cabo de dos años protege a más personas que la primera. Para ello se seleccionan 10 personas al azar y se les inyecta la nueva vacuna. Establecemos que si más de 8 de los vacunados conservan la protección al cabo de dos años, entonces consideraremos la nueva vacuna mejor que la antigua. Se pide:

a) Calcular la probabilidad de cometer un error de tipo I.

b) Si la nueva vacuna protegiera a un 80 % de las personas vacunadas al cabo de 2 años, ¿Cuál será la probabilidad de cometer un error de tipo II?

139. Hacer el ejercicio anterior estableciendo que si más de 7 de los vacunados conservan la protección al cabo de 2 años consideraremos la nueva vacuna mejor que la antigua.

140. Repetir el ejercicio seleccionando a 100 personas y estableciendo que si más de 85 de los vacunados conservan la protección al cabo de 2 años, consideraremos la nueva vacuna mejor que la antigua.

NOTA: Aproximar la distribución binomial mediante una distribución normal.

141. Hacer el ejercicio anterior estableciendo que si más de 75 de los vacunados conservan la protección al cabo de dos años, consideraremos que la nueva vacuna es mejor que la antigua.

142. Un fisioterapeuta afirma que con un nuevo procedimiento de rehabilitación que él aplica, determinada lesión tiene un tiempo de recuperación medio no mayor de 15 días. Se seleccionan al azar 36 personas que sufren dicho tipo de lesión para verificar su afirmación, y se obtiene un tiempo medio de recuperación de 17 días y una cuasivarianza de 9. ¿Contradice lo observado en la muestra la afirmación del fisioterapeuta para un $\alpha = 0{,}05$?

143. Se decide retirar una cierta vacuna si produce más de un 10 % de reacciones alérgicas. Se consideran 100 pacientes sometidos a la vacuna y se observan 15 reacciones alérgicas. ¿Debe retirarse la vacuna? (Utilizar un $\alpha = 0{,}01$).

144. Se utiliza un grupo de 150 pacientes para comprobar la teoría de que la vitamina C tiene alguna influencia en el tratamiento del cáncer. Los 150 pacientes fueron divididos en dos grupos de 75. Un grupo recibió 10 gramos de vitamina C y el otro un placebo cada día, además de la medicación habitual. De los que recibieron la vitamina C, 47 presentaban alguna mejoría al cabo de cuatro semanas, mientras que de los que recibieron el placebo, 43 experimentaron mejoría. Contrastar esta hipótesis.

145. Se realizó en dos hospitales una encuesta entre los pacientes sobre la satisfacción con la atención recibida, calificándola de 0 a 100. En el hospital A rellenaron la encuesta 12 pacientes, obteniéndose una media de 85 y una cuasivarianza de 16, mientras que en el hospital B rellenaron la encuesta 10 pacientes, obteniéndose una media de 81 y una cuasivarianza de 25. ¿Puede concluirse que el nivel de satisfacción en el hospital A es mayor que en el B?

NOTA: Hacer previamente un contraste de igualdad de varianzas.

*146. Se compararon los niveles de ácido ascórbico en plasma de mujeres embarazadas fumadoras y no fumadoras, obteniéndose los siguientes resultados expresados en gramos de ácido ascórbico por mililitro de plasma:

Mujeres no fumadoras: 0.97-0.72-1.00-0.81-0.62-1.32-1.24.

Mujeres fumadoras: 0.48-0.71-0.98-0.68-1.18.

Suponiendo que las varianzas poblacionales en fumadoras y no fumadoras son iguales, ¿existe suficiente evidencia para concluir que el nivel de ácido ascórbico en la sangre de mujeres fumadoras es mayor que el de mujeres no fumadoras?

147. Verificar la hipótesis de que el contenido medio de unos recipientes de ácido sulfúrico es de 10 litros, si los contenidos de una muestra aleatoria de 10 recipientes son 10.2, 9.7, 10.1, 10.3, 10.1, 10.1, 9.8, 9.9, 10.4, 10.3 y 9.8 litros. Utilizar un nivel de significación de 0.01 y suponer que la distribución de los contenidos es normal.

148. Un fabricante de equipos de medida afirma que sus equipos pueden realizar al menos 12 mediciones más que los de la competencia sin necesidad de un nuevo ajuste. Para probar esta afirmación se realizan mediciones con 50 equipos de este fabricante y 50 de la competencia. En los suyos el número de mediciones hasta necesitar un nuevo ajuste tuvo de media 86.7 y cuasidesviación típica 6.28, mientras que en los de la competencia estos valores fueron 77.8 y 5.61 respectivamente. Verificar la afirmación del fabricante con $\alpha = 0{,}05$.

149. Para determinar si un nuevo suero detiene la leucemia, se seleccionan 9 ratones con leucemia en una fase avanzada. Cinco reciben el tratamiento y cuatro no. Los tiempos de supervivencia, en años, desde el momento que comenzó el experimento son los siguientes:

Con tratamiento: 2.1 – 5.3 – 1.4 – 4.6 – 0.9. Sin tratamiento: 1.9 – 0.5 – 2.8 – 3.1.

¿Puede afirmarse con un $\alpha = 0{,}05$ que el suero es eficaz? Suponer que ambas distribuciones son normales con varianzas iguales.

150. Un estudio afirma que el 70 % de los habitantes de la capital lee diariamente algún periódico. ¿Estaríamos de acuerdo con las conclusiones de dicho estudio si al preguntar a 15 personas elegidas aleatoriamente, 8 leen diariamente algún periódico?

151. Un distribuidor de tabaco asegura que el 20 % de los fumadores de su ciudad prefiere los cigarrillos de marca A. Se selecciona al azar una muestra de 20 fumadores, y 6 de ellos prefieren la marca A. ¿Qué conclusión se obtiene con $\alpha = 0{,}05$?

152. en un estudio sobre el consumo de alcohol entre los jóvenes durante los fines de semana, se preguntó a 100 chicos y a 125 chicas, de los que 63 chicos y 59 chicas contestaron que consumían. En vista de estos datos, ¿existe alguna diferencia significativa entre las respuestas de chicos y chicas? Utilizar $\alpha = 0{,}10$.

153. Un fabricante de baterías para automóvil asegura que la duración de sus baterías tiene una distribución aproximadamente normal con desviación típica no superior a 0.9 años. Si una muestra aleatoria de 10 de estas baterías tiene una cuasidesviación típica de 1.2 años, ¿qué se puede concluir sobre la afirmación del fabricantes?

154. En un estudio sobre el contenido de ortofósforo de las aguas de un río, se realizaron medidas en dos estaciones distintas. Se sacaron 15 muestras de la estación 1 y 12 de la estación 2. Las muestras de la estación 1 presentaron un contenido medio de ortofósforo de 3.84 mg/l y una cuasidesviación típica de 3.07 mg/l, mientras que las de la estación 2 tuvieron media 1.49 mg/l y una cuasidesviación típica 0.8 mg/l. Se pide:

   a) Calcular el intervalo de confianza para el cociente de varianzas.
   b) Realizar el contraste de hipótesis de igualdad de varianzas.

   Utilizar un $\alpha = 0{,}05$.

Notas:

- Los problemas marcados con un asterisco (*) son problemas de exámenes de cursos anteriores en otras titulaciones (Enfermería, Fisioterapia, Nutrición Humana y Dietética...).
- Los problemas marcados con dos asterisco (**) son problemas de exámenes de cursos anteriores en la titulación de Medicina.