EJERCICIOS DE ESTADÍSTICA

Asignatura: Bioestadística

Curso: 1º de Licenciatura en Medicina

Santiago Angulo Díaz-Parreño (sangulo@ceu.es)
José Miguel Cárdenas Rebollo (cardenas@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

Curso 2010-2011

CEU Universidad San Pablo

Índice

1. 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.

   Solución

   d) $\bar{x} = 2{,}225$ medicamentos, $Med = 2$ medicamentos y $Mod = 2$ y 3 medicamentos.
   e) $s^2 = 1{,}974$ medicamentos^2, $s = 1{,}405$ medicamentos.
   f) $cv = 0{,}632$, lo que indica que la hay una dispersión moderada-alta.

2. En un hospital se realizó un estudio sobre el número de personas que ingresaron en urgencias en el mes de noviembre. Los datos observados fueron:

$$15-23-12-10-28-7-12-17-20-21-18-13-11-12-26$$
$$30-6-16-19-22-14-17-21-28-9-16-13-11-16-20$$

   Se pide:

   a) Obtener una distribución de frecuencias agrupadas.
   b) Dibujar el histograma y el polígono de frecuencias correspondientes.
   c) Calcular la media aritmética, la mediana y la moda.
   d) Calcular la varianza y la desviación típica.
   e) Calcular el coeficiente de variación de Pearson.
   f) Calcular el tercer decil.
   g) Calcular el percentil 62.

   Solución

   Agrupando en las clases $(5, 10]$, $(10, 15]$, $(15, 20]$, $(20, 25]$, $(25, 30]$.

   c) $\bar{x} = 16{,}67$ ingresos, $Med = 16{,}11$ ingresos y $Mod = (10, 15]$ y $(15, 20]$ ingrsos.
   d) $s^2 = 36{,}75$ ingresos^2, $s = 6{,}06$ ingresos.
   e) $cv = 0{,}365$, lo que indica que hay poca dispersión.
   f) $D_3 = 12{,}78$ ingresos.
   g) $P_{62} = 18{,}1$ ingresos.

★ 3. 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.

Solución

c) $C_1 = 1$ lesión, $C_2 = 2$ lesiones y $C_3 = 3$ lesiones. $RI = 2$ lesiones, lo que indica que hay bastante dispersión central.

d) $\bar{x} = 1{,}125$ lesiones, $s^2 = 0{,}776$ lesiones^2, $s = 0{,}88$ lesiones y $g_1 = 0{,}49$ lo que indica que la distribución es un poco asimétrica a la derecha.

★ 4. El tiempo de recuperación, medido en días, de 10 pacientes con una determinada lesión de rodilla ha sido

46, 54, 48, 62, 51, 50, 54, 50, 49, 52.

Calcular el coeficiente de asimetría y de apuntamiento e interpretarlos.

Solución

$g_1 = 1{,}22$ (bastante asimétrica hacia la derecha) y $g_2 = 1{,}17$ (bastante leptocúrtica).

5. En un cuestionario se ha preguntado a un grupo de individuos por su nivel de estudios (SE: sin estudios, EB: estudios básicos, ES: estudios secundarios, EU: estudios universitarios, ED: estudios de doctorado) obteniendo los siguientes resultados:

EU, ES, ES, EU, EB, EB, ED, SE, EB, SE, EU, ES, ES, ES, EB,
EB, SE, ED, EB, ES, ES, SE, EU, EU, EB, ES, EU, ES, EB, ES

Se pide:

a) Construir la tabla de frecuencias y los diagramas asociados.

b) Dar una medida de representatividad.

c) Calcular los cuartiles y el percentil 90.

Solución

b) Me =ES y Mo =ES.

c) C_1 =EB, C_2 =ES, C_3 =EU y P_{90} =EU.

★ 6. 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:

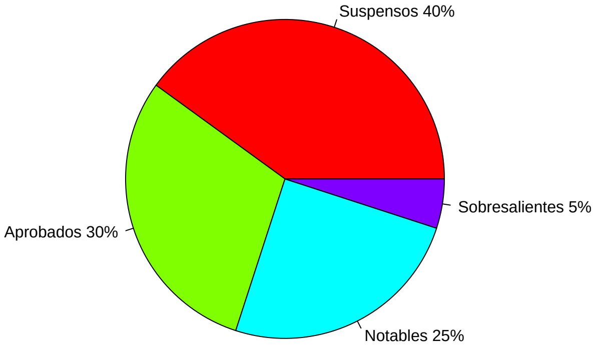
Recién nacidos: 51-50-51-53-49-50-53-50-47-50.
Niños de un año: 62-65-69-71-65-66-68-69.

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

Solución

Llamando X a las estaturas de los niños recién nacidos e Y a las estaturas de los niños de 1 año: $\bar{x} = 50{,}4$ cm, $s_x = 1{,}685$ cm, $cv_x = 0{,}034$, $\bar{y} = 66{,}875$ cm, $s_y = 2{,}713$ cm, $cv_y = 0{,}041$, lo que indica que ambas medias son muy representativas pero un poco más la de los niños recién nacidos.

★ 7. 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.

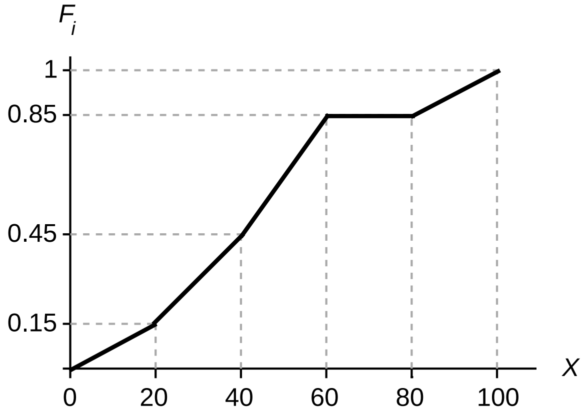
Solución

c) $Me =$ Aprobado y $Mo =$ Suspenso.

d) $\bar{x} = 5{,}275$ puntos, $s = 2{,}447$ puntos y $cv = 0{,}464$, de manera que la media es moderadamente representativa.

8. 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.

Solución

c) $Med = 42{,}5$ y $Mod = (40, 60)$.

d) =44, $s^2 = 564$, $s = 23{,}75$.

★ 9. 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.

Solución

$\bar{x} = 11{,}5$, $s^2 = 24$, $s = 4{,}9$, $cv = 0{,}426$, lo que indica que hay una dispersión moderada. $C_1 = 7{,}5$, $C_3 = 15{,}83$, $RI = 8{,}33$, lo que indica que la dispersión central también es moderada.

---

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		0,20	
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.

Solución

b) $\bar{x} = 1{,}62$ radiografías, $s^2 = 1{,}875$ radiografías^2, $s = 1{,}37$ radiografías y $cv = 0{,}845$, lo que indica que hay bastante dispersión y la media no es muy representativa.

★ 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?

Solución

b) $\bar{x} = 47{,}75$ minutos, $s^2 = 66{,}188$ minutos^2, $s = 8{,}136$ minutos. $g_1 = 0{,}268$ lo que indica que la distribución es ligeramente asimétrica hacia la derecha.

c) $P_{95} = 62{,}5$ minutos.

★ 12. Como parte de un proyecto de investigación, los investigadores obtuvieron los siguientes datos respecto a los niveles de peróxido lípido (en nmol/ml) en el suero de 30 individuos adultos bajo tratamiento de Diabetes Mellitus:

3,09	6,06	7,34	5,32	4,29	5,36	6,01	7,84	3,87	5,23
4,67	7,89	5,16	6,32	6,45	3,21	5,98	6,45	7,12	4,13
5,16	3,04	4,56	5,67	5,98	6,23	7,34	5,32	4,21	7,13

Agrupar los datos en 5 clases de amplitud unidad, comenzando en 3, y sobre la distribución obtenida:

a) Calcular media, desviación típica y coeficiente de variación de los niveles de peróxido lípido. Interpretar el coeficiente de variación.

b) Calcular cuartiles de la distribución e interpretarlos.

c) Dibujar el diagrama de caja y bigotes y comprobar si hay o no datos atípicos.

Solución

a) $\bar{x} = 5{,}667$ nmol/ml, $s^2 = 1{,}668$ (nmol/ml)^2, $s = 1{,}292$ nmol/ml y $cv = 0{,}228$, lo que indica que hay poca dispersión y la media es bastante representativa.

b) $C_1 = 4{,}7$ nmol/ml, $C_2 = 5{,}667$ nmol/ml y $C_3 = 6{,}75$ nmol/ml.

c) Las vallas son $v_1 = 1{,}625$ y $v_2 = 9{,}825$. Todos los datos están entre las vallas y no hay datos atípicos. Los bigotes son $b_1 = 3{,}04$ nmol/ml y $b_2 = 7{,}89$ nmol/ml.

★ 13. 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

a) Sabemos que una distribución normal es simétrica y mesocúrtica. Por tanto, una primera idea de si los datos muestrales provienen de una distribución normal nos la puede dar ver si estos valores se encuentran en el intervalo [-2, 2]. ¿Podríamos suponer según esto que los datos provienen de una distribución normal?

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

Solución

a) $\bar{x} = 65{,}769$ años, $s^2 = 114{,}823$ años, $s = 10{,}716$ años. $g_1 = -0{,}228$ y $g_2 = -0{,}55$, como tanto el coefiente de asimetría como el de apuntamiento están entre -2 y 2, podemos suponer que los datos provienen de una población normal.

b) $P_{85} = 77{,}5$ años.

★ 14. 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.

Solución

b) $\bar{x} = 66{,}3125\,\%$, $s^2 = 9{,}9023\,\%^2$, $s = 3{,}1468\,\%$. $C_1 = 64\,\%$, $C_2 = 66\,\%$, $C_3 = 69\,\%$.

c) $cv = 0{,}05$ lo que indica que hay muy poca dispersión y la media es muy representativa.

15. Para determinar la eficacia de un nuevo método para la medición del hematocrito en sangre, se repitió la medida 8 veces sobre una misma muestra de sangre, obteniéndose los siguientes resultados (en porcentaje de hematocrito sobre volumen de plasma sanguíneo):

42,2 42,1 41,9 41,8 42 42,1 41,9 42.

¿Se puede afirmar que se trata de un buen método de medición?

Solución

$\bar{x} = 42\,\%$, $s^2 = 0{,}015\,\%^2$, $s = 0{,}1225\,\%$ y $cv = 0{,}003$ lo que indica que la variabilidad entre las mediciones es ínfima y por tanto se trata de un buen método de medición.

16. En cuestionario sobre la dependencia de las personas mayores de 75 años se preguntaba sobre la necesidad de ayuda en el desarrollo normal de su vida. Las posibles respuestas eran:

    a) Ninguna ayuda.

    b) Ayuda al subir las escaleras.

    c) Ayuda al subir las escaleras y al incorporarse de una posición sentada o tumbada.

    d) Ayuda al subir las escaleras, al incorporarse, y al vestirse.

    e) Ayuda para prácticamente todo.

    El cuestionario lo respondieron 20 personas, y los resultado obtenidos fueron

    b – d – a – b – c – c – b – c – d – e – a – b – c – e – a – b – c – d – b – b

    Se pide:

    a) Representar gráficamente la distribución de frecuencias.

    b) Calcular las medidas de tendencia central.

    c) Calcular los cuartiles y el decil 8.

    d) ¿Qué se puede decir sobre la dispersión?

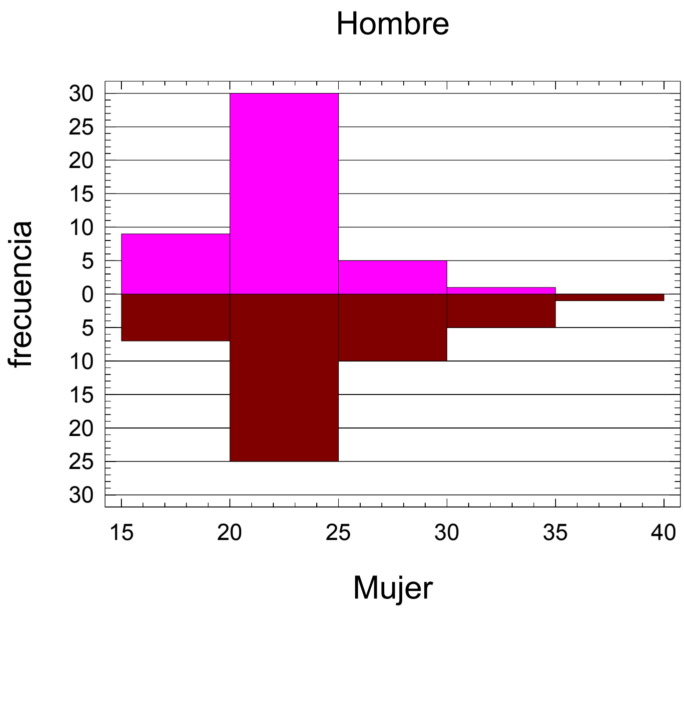
    Solución

    b) Me ente b y c y $Mo = b$.

    c) C_1 =b, C_2 entre b y c, C_3 =d.

    d) Suponiendo que hay la misma distancia entre categorías y asignando rangos a cada valor ($a = 1, b = 2, c = 3, d = 4, e = 5$), tenemos $\bar{x} = 2{,}7$, $s^2 = 1{,}41$, $s = 1{,}187$ y $cv = 0{,}44$, lo que indica una dispersión moderada.

★ 17. El siguiente histograma refleja la distribución del índice de masa corporal en una muestra de hombres y mujeres.



Se pide:

a) Construir la tabla de frecuencias para hombres y mujeres por separado.

b) Dibujar el diagrama de sectores para el sexo.

c) ¿En qué grupo es más representativa la media? Justificar la respuesta.

d) ¿Cómo calcularías la media de toda la muestra a partir de las medias de hombres y mujeres? ¿Cuánto vale?

e) Calcular el rango intercuartílico del índice de masa corporal en los hombres.

Solución

c) $\bar{m} = 24{,}17$ Kg/m^2, $s_m^2 = 21{,}1806$ (Kg/m^2)^2, $s_m = 4{,}6$ Kg/m^2 y $cv_m = 0{,}19$.
$\bar{h} = 22{,}18$ Kg/m^2, $s_m^2 = 9{,}9506$ (Kg/m^2)^2, $s_m = 3{,}15$ Kg/m^2 y $cv_h = 0{,}14$, luego es más representativa la media en los hombres.

d) $\bar{x} = 23{,}25$ Kg/m^2.

e) $RI = 3{,}75$ Kg/m^2.

★ 18. Se ha medido el tiempo necesario, en días, para curar un tipo concreto de infección (tiempo transcurrido desde el comienzo del tratamiento hasta el alta) en dos grupos de pacientes a los que se han aplicado antibióticos diferentes. En un grupo de 80 pacientes tratados con el antibiótico A, los resultados obtenidos fueron:

Tiempo (días)	n_i
$[0, 30)$	50
$[30, 60)$	20
$[60, 90)$	10

Mientras que en un grupo de 100 pacientes tratados con el antibiótico B se obtuvo:

Tiempo (días)	n_i
$[0, 30)$	40
$[30, 60)$	50
$[60, 90)$	10

a) Calcular la media y la desviación típica del número de días hasta el alta, para cada uno de los antibióticos.

b) ¿En qué antibiótico es más representativa la media?. Justificar la respuesta numéricamente.

c) ¿Cuánto vale el percentil 90 del tiempo de curación para los pacientes tratados con B?

d) ¿Qué tiempo de recuperación ha sido relativamente más alto, uno de 40 días de un paciente tratado con A, o uno de 45 días de otro tratado con B? Justificar la respuesta numéricamente.

2. Regresión y Correlación

★ 19. La artrosis reumatoide es una enfermedad reumática que aparece con frecuencia en las personas mayores. Uno de los índices más utilizados para ver el grado de actividad de la enfermedad es el RADAI (Rheumatoid Arthritis Disease Activity Index), que mide el grado de actividad en una escala de 0 (mínima actividad) a 3 (máxima actividad). Para ver de qué manera influye la edad en el grado de actividad de la enfermedad se ha seleccionado un grupo de personas mayores y se ha medido el índice RADAI en ellos, obteniendo la siguiente tabla de frecuencias:

RADAI\Edad	40-50	50-60	60-70	70-80
0-1	8	6	2	1
1-2	4	7	5	2
2-3	0	2	6	7

Se pide:

a) Estudiar si existe relación lineal entre la edad y el RADAI.

b) Calcular la recta de regresión del RADAI sobre la edad. Según la recta, ¿cuánto aumentaría el grado de actividad de la enfermedad por cada año que pasa?

c) Si se considera que los pacientes don un RADAI de 2 o superior necesitan ayuda en sus actividades diarias, ¿a qué edad se empezaría a necesitar esta ayuda?

Solución

Llamando X a la variable que mide la edad e Y a la que mide el RADAI.

a) $r = 0{,}59$ que indica una relación lineal moderada.

b) Recta de regresión de Y sobre X: $y = 0{,}0442x - 1{,}1575$. Cada año que pase la actividad de la enfermedad aumentará 0,0442 puntos en el RADAI.

c) A los 63,4 años.

★ 20. 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.

Solución

Llamando X a la dosis e Y al tiempo de curación:

a) $\bar{x} = 3{,}05$ mg, $\bar{y} = 4{,}55$ días, $s_x^2 = 0{,}648$ mg^2, $s_y^2 = 1{,}448$ días^2, $s_x = 0{,}805$ mg, $s_y = 1{,}203$ días y $s_{xy} = -0{,}678$ mg·días.
$r = -0{,}7$, que quiere decir que hay buena relación lineal entre la dosis y el tiempo de curación, y además es decreciente (a mayor dosis, menor tiempo de curación).

b) Recta de regresión del tiempo de curación sobre la dosis: $y = -1{,}046x + 7{,}741$.
$y(5) = 2{,}511$ días.

---

21. Se determina la pérdida de actividad que experimenta un medicamento desde el momento de su fabricación a lo largo del tiempo, obteniéndose el siguiente resultado:

Tiempo (en años)	1	2	3	4	5
Actividad restante (%)	96	84	70	58	52

Se desea calcular:

a) La relación fundamental (recta de regresión) entre actividad y tiempo transcurrido.

b) El tiempo en meses que corresponde al 80 % de actividad.

c) ¿Cuándo será nula la actividad?

Solución

Llamando T al tiempo y A a la actividad del fármaco:

a) $\bar{t} = 3$ años, $\bar{a} = 72\,\%$, $s_t^2 = 2$ años^2, $s_a^2 = 264\,\%^2$, $s_{ta} = -22{,}8$ años$\cdot$ %.
Recta de regresión de actividad sobre tiempo: $a = -11{,}4t + 106{,}2$.

b) Recta de regresión de tiempo sobre actividad: $t = -0{,}086a + 9{,}2182$.
$t(80) = 2{,}3091$ años.

c) $t(0) = 9{,}2182$ años.

---

22. Para comprobar el efecto de la herencia genética sobre la inteligencia se desarrolló un estudio en el que se midió el coeficiente intelectual de varias parejas de gemelos, obteniendo los siguientes resultados:

$$(128, 132)\ (116, 112)\ (86, 98)\ (65, 81)\ (104, 96)\ (111, 111)\ (101, 105)\ (72, 75)$$

Calcular el coeficiente de determinación lineal e interpretarlo. ¿Tiene sentido calcular el coeficiente de correlación?

Solución

Llamando X al coeficiente intelectual del primer hermano e Y al del segundo: $\bar{x} = 97{,}875$, $\bar{y} = 101{,}25$, $s_x^2 = 418{,}3594$, $s_y^2 = 288{,}4375$, $s_{xy} = 326{,}5313$ y $r^2 = 0{,}8836$, lo que indica que existe bastante relación lineal entre el coeficiente intelectual de los gemelos. No tiene sentido el coeficiente de correlación lineal porque es indiferente el orden en que tomemos a los gemelos.

★ 23. Después de tomar un litro de vino se ha medido la concentración de alcohol en la sangre en distintos instantes, obteniendo:

Tiempo después (minutos)	30	60	90	120	150	180
Concentración (gramos/litro)	1.6	1.7	1.5	1.1	0.7	0.2

Se pide:

a) Calcular la recta de regresión de la concentración en función del tiempo.

b) ¿Qué concentración de alcohol habrá a los 100 minutos?

c) Si la concentración máxima de alcohol en la sangre que permite la ley para poder conducir es 0.8 g/l, ¿cuánto tiempo habrá que esperar después de tomarse un litro de vino para poder conducir sin infringir la ley?

Solución

---

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.

Solución

a) $\bar{x} = 8$ y $\bar{y} = 10$.

b) $r = 0{,}5$.

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.

Solución

 a) $y(10) = 1$.

 b) $x(5) = 2$.

★ 26. 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) Dibujar el diagrama de dispersión tomando una escala logarítmica para los días de dieta.

 c) Calcular el modelo lineal y el logarítmico de los kilos perdidos con respecto a los días de dieta. Nota: Utilizar los datos muestrales sin agrupar.

 d) Utilizar el mejor de los modelos anteriores para predecir en número de kilos perdidos tras 40 días de dieta y tras 100 días. ¿Son fiables estas predicciones?

Solución

Llamando X a los días de dieta, Y a los Kg perdidos y $Z = \log X$.

 c) $\bar{x} = 47$ días, $\bar{y} = 5{,}3$ Kg, $s_x^2 = 143{,}833$ días^2, $s_y^2 = 0{,}885$ Kg^2, $s_{xy} = 9{,}942$ días·Kg. Modelo lineal: $y = 0{,}069x + 2{,}051$.
 $\bar{z} = 3{,}82$ logdías, $s_z^2 = 0{,}07$ $\log^2$días, $s_{yz} = 0{,}22$ logdías·Kg.
 Modelo logarítmico: $y = 3{,}4 \log y - 7{,}67$.

 d) Modelo lineal: $r^2 = 0{,}78$, modelo logarítmico: $r^2 = 0{,}86$.
 Predicciones con el modelo logarítmico: $y(40) = 4{,}86$ Kg y $y(100) = 7{,}98$ Kg. Las predicciones son fiables ya que el coeficiente de determinación es alto.

★ 27. 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 $t = 2$?

Solución

a) $\bar{s} = 0$, $\bar{t} = 1{,}5$ y $s_{st} = 0$.

b) No podemos afirmar que S y T son independientes, sólo se puede afirmar que no hay relación lineal.

c) $s(2) = 0$.

---

28. En un estudio sobre la influencia del tabaco en los embarazos se ha medido en una muestra de 20 madres el número medio de cigarrillos diarios que fumaban las madres y el peso del recién nacido, obteniendo los siguientes resultados

Cigarrillos	2	3	10	8	12	6	6	5	4	9	14	3	7	8	2
Peso (kg)	3.1	3.3	2.5	3.3	2.6	3.1	3.0	3.4	3.4	2.7	2.5	3.7	3.1	3	3.6

Se pide:

a) Construir el modelo de regresión logarítmico del peso sobre el número de cigarrillos.

b) Según este modelo, ¿cuanto pesará el recién nacido si la madre fumaba 15 cigarrillos diários? Es fiable esta predicción.

c) ¿Es mejor el modelo lineal a la hora de hacer predicciones?

Solución

★ 29. 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 se explica, 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.

Solución

a) $\bar{x} = 40$ años, $\bar{y} = 130{,}867$ mmHg, $s_x^2 = 200$ años^2, $s_y^2 = 26{,}295$ mmHg^2 y $s_{xy} = 58{,}667$ años·mmHg.
$r^2 = 0{,}654$, luego el modelo lineal explica el 65,4 % de la varianza de la presión sistólica.

b) Recta de regresión de la edad sobre la presión sistólica: $x = 2{,}231y - 251{,}978$.
$x(133) = 44{,}745$ años. La predicción es bastante fiable pues el coeficiente de determinación es alto.

★ 30. En un banco de sangre se mantiene el plasma a 0ºF. Cuando se necesita para una transfusión se calienta en un horno a una temperatura constante de 120ºF. En un experimento se ha medido la temperatura del plasma a distintos instantes desde el comienzo del calentamiento. Los resultados son:

Tiempo (min)	5	8	15	25	30	37	45	60
Temperatura (ºF)	25	50	86	102	110	114	118	120

Se pide:

a) Dibujar el diagrama de dispersión. ¿Qué modelo expliaría la relación entre la temperatura y el tiempo?

b) ¿Qué transformación de escala tendríamos que realizar en las variables para tener una nube de puntos con una tendencia lineal? Hacer la representación gráfica.

c) Construir el modelo de regresión logarítmico de la temperatura sobre el tiempo.

d) Según el modelo, ¿qué temperatura habrá a los 15 minutos? ¿Es fiable la predicción? Justificar la respuesta.

Solución

★ 31. 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/dl? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.

Solución

Llamando T al tiempo, C a la concentración y Z al logaritmo de la concentración:

a) $\bar{t} = 5$ horas, $\bar{z} = 4{,}1639$ log(mg/dl), $s_t^2 = 4$ horas^2, $s_z^2 = 0{,}3785$ log^2(mg/dl), $s_{tz} = 1{,}2291$ horas· log(mg/dl).
Modelo exponencial de C sobre T: $c = e^{0{,}3073x+2{,}6275}$.
$c(4{,}8) = 60{,}498$ mg/dl y es bastante fiable ya que $r^2 = 0{,}999$.

b) $\bar{c} = 77{,}2857$ mg/dl, $s_c^2 = 2160{,}7755$ (mg/dl)^2, $s_{tc} = 89$ horas(mg/dl).
Modelo lineal de T sobre C: $t = 0{,}0412c + 1{,}8167$.
$t(100) = 5{,}9356$ y también es fiable ya que $r^2 = 0{,}9165$.

32. 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 e investigadora? Justificar la respuesta.

Solución

$r_s = -0{,}83$, lo que indica una fuerte relación inversa entre la excelencia docente y la excelencia investigadora.

★ 33. En un grupo de pacientes se analiza el efecto de una sustancia dopante en el tiempo de respuesta a un determinado estímulo. Para ello, se suministra en sucesivas dosis, de 0 a 70 mg, la misma cantidad de dopante a todos los miembros del grupo, y se anota el tiempo medio de respuesta al estímulo, expresado en centésimas de segundo.

X (mg)	0	10	20	30	40	50	60	70
Y (10^{-2}s)	28	46	62	81	100	132	195	302

a) Representar la nube de puntos. A la vista de la representación, ¿crees que el modelo lineal es el que mejor explica el tipo de relación entre las variables?

b) Calcular la recta de regresión del tiempo en función de la cantidad de dopante, y utilizarla para predecir el tiempo de reacción medio para una cantidad de dopante de 25 mg.

c) Hacer la misma predicción del apartado anterior con el modelo exponencial. ¿Qué predicción es mejor?

d) Si para el estímulo estudiado los tiempos de reacción superiores a un segundo se consideran peligrosos para la salud, ¿a partir de qué nivel debería regularse, e incluso prohibirse, la administración de la sustancia dopante?

Solución

b) Recta de regresión de Y sobre X: $y = 3{,}44x - 2{,}25$. $y(25) = 83{,}82$ centésimas de segundo.

c) Recta de regresión de X sobre Y: $x = 0{,}25y + 5{,}57$. $x(100) = 30{,}46$ mg.

★ 34. Los investigadores están estudiando la correlación entre obesidad y la respuesta individual al dolor. La obesidad se mide como porcentaje sobre el peso ideal (x). La respuesta al dolor se mide utilizando el umbral de reflejo de flexión nociceptiva (y), que es una medida de sensación de punzada. Se obtuvieron los siguientes resultados:

x	89	90	75	30	51	75	62	45	90	20
y	10	12	4	4,5	5,5	7	9	8	15	3

a) Dibujar el diagrama de dispersión. Según la nube de puntos, ¿qué modelo explicaría mejor la relación entre el umbral de reflejo y el porcentaje sobre el peso ideal?

b) Obtener la recta de regresión que mejor exprese la dependencia del umbral de reflejo en función del porcentaje sobre el peso ideal.

c) Obtener el modelo exponencial que mejor exprese la dependencia del umbral de reflejo en función del porcentaje sobre el peso ideal.

d) Predecir el porcentaje de sobrepeso que se espera obtener para un umbral de reflejo de 10, ¿Es fiable esta predicción?.

Solución

★ 35. 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	9,0	9,0	11	13	14	14,5	16
Esteroide	30,0	25,0	31,5	27,5	39,5	38,0	43,0	49,0	55,0	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.

Solución

★ 36. 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) Calcular el coeficiente de correlación de Spearman e interpretarlo. ¿Está en consonancia con el coeficiente de correlación lineal?

Solución

---

3. Probabilidad

37. Se dispone de dos urnas, la primera con 10 bolas blancas y 6 bolas negras, y la segunda con 5 bolas rojas, 8 bolas azules y 3 bolas verdes. Construir el espacio muestral del experimento que consiste en sacar una bola de cada urna, y del experimento que consistiría en sacar dos bolas de cada urna.

38. Un experimento consiste en seleccionar a una pareja de personas y medir su grupo sanguineo y si son fumadores o no. Expresar el espacio muestral en forma de árbol.

39. En una estantería en la que hay 3 cajas de un medicamento A y 2 de un medicamento B, se eligen 3 al azar. ¿Cuál es la probabilidad de que se hayan elegido 2 cajas del medicamento A y 1 del B?

Solución

36/60.

---

40. En un laboratorio hay 4 frascos de ácido sulfúrico y 2 de ácido nítrico, y en otro hay 1 frascos de ácido sulfúrico y 3 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.

d) Calcular la probabilidad de estos mismos sucesos si el frasco elegido en el primer laboratorio se introduce en el segundo antes de sacar el frasco de este.

Solución

a) 4/24.
b) 6/24.
c) 14/24.
d) 8/30, 8/30 y 14/30 respectivamente.

---

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

a) $P(A \cup B)$.
b) $P(\overline{A})$ y $P(\overline{B})$.
c) $P(\overline{A} \cap \overline{B})$.
d) $P(A \cap \overline{B})$.
e) $P(A/B)$.
f) $P(A/\overline{B})$.

Solución

a) $P(A \cup B) = 5/8$.
b) $P(\overline{A}) = 5/8$ y $P(\overline{B}) = 1/2$.
c) $P(\overline{A} \cap \overline{B}) = 3/8$.
d) $P(A \cap \overline{B}) = 1/8$.
e) $P(A/B) = 1/2$.
f) $P(A/\overline{B}) = 1/4$.

---

42. 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?

Solución

0,0199.

---

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

a) A y B son independientes.
b) A y B son incompatibles.

Solución

a) $P(B) = 0{,}75$.
b) $P(B) = 0{,}3$.

---

44. El tétanos es mortal en el 70 % de los casos. Si tres personas contraen el tétanos, ¿Cuál es la probabilidad de que mueran al menos dos de los tres?

Solución

0,784.

★ 45. 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.

Solución

Llamando HT a tener hipertensión y HC a tener hipercolesterolemia:

a) $P(HT \cap HC) = 0{,}12$.
b) $P(HT/\overline{HC}) = 0{,}0698$.

★ 46. En un estudio sobre el tabaco, se informa que el 40 % de los fumadores tiene un padre fumador, el 25 % tiene una 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?

Solución

Llamando PF al suceso que consiste en tener un padre fumador y MF a tener una madre fumadora:

a) $P(MF/PF) = 0{,}33$.
b) $P(MF/\overline{PF}) = 0{,}2$.
c) No son independientes.

47. 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?

★ 48. 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?

★ 49. 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?

Solución

Llamado S a sanar y E a tener efectos secundarios:

a) $P(S) = 0{,}868$ y $P(E) = 0{,}114$.

b) $P(A/S) = 0{,}495$, $P(B/S) = 0{,}318$ y $P(C/S) = 0{,}187$.
$P(A/\overline{E}) = 0{,}497$, $P(B/\overline{E}) = 0{,}291$ y $P(C/\overline{E}) = 0{,}212$.
En ambos casos el tratamiento más probable es el A.

c) $P(S \cap \overline{E}) = 0{,}806$.

★ 50. 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 grado3?

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?

★ 51. En una población se ha vacunado a la tercera parte de los individuos contra la gripe. Trascurrido el invierno, se comprueba que la probabilidad de estar vacunado si se tiene la gripe es 0,2, y que el 10 % de los vacunados tuvieron gripe.

a) ¿Cuál fue la incidencia de la epidemia de gripe?
(Nota: La incidencia de una epidemia es la probabilidad de personas infectadas).

b) ¿Cuál es la probabilidad de que una persona no vacunada contraiga la gripe?

c) ¿Se puede afirmar que la vacuna tiene alguna eficacia?

SOLUCIÓN

Llamando G al suceso consistente en tener la gripe y V a estar vacunado:

a) $P(G) = 1/6$.

b) $P(G/\overline{V}) = 0{,}2$.

c) Si resulta eficaz, aunque poco.

★ 52. Una enfermedad se trata con 3 medicamentos diferentes: A en un 50 % de los casos, B en un 30 % y C en un 20 %, todo ello independientemente de si se es hombre o mujer. Si sabemos que el medicamento A produce efectos secundarios en un 5 % de los hombres y en un 10 % de las mujeres, el B en un 15 % de los hombres y en un 5 % de las mujeres, y el C en un 8 % de los hombres y en un 13 % de las mujeres, se pide:

a) ¿En qué colectivo resulta más probable que haya efectos secundarios, en hombres o en mujeres? Justificar adecuadamente la respuesta.

b) Calcular la probabilidad de que un hombre que presenta efectos secundarios haya sido tratado con C, y la de que una mujer que no los presenta haya sido tratada con A.

c) Si en total de enfermos hay un 65 % de hombres y un 35 % de mujeres, ¿qué probabilidad hay de que un enfermo que no presenta efectos secundarios sea mujer?

SOLUCIÓN

Llamando EH a que un hombre tenga efectos secundarios y EM a que los tenga una mujer:

a) $P(EH) = 0{,}086$ y $P(EM) = 0{,}091$.

b) $P(C/EH) = 0{,}186$, y $P(A/\overline{EM}) = 0{,}495$.

c) $P(M/\overline{E}) = 0{,}349$.

---

53. Se sabe que los grupos sanguíneos en una determinada población se distribuyen con las siguientes frecuencias:

$$0 : 30\,\% \quad A : 45\,\% \quad B : 18\,\% \quad AB : 7\,\%$$

Por otro lado, también se sabe que la octava parte de los individuos del grupo 0 tienen RH negativo, así como la cuarta parte del grupo A, la mitad del grupo B, y la tercera parte del grupo AB.

Se pide:

a) ¿Cuál es la probabilidad de que un indiviudo elegido al azar sea del tipo A y tenga RH positivo?

b) ¿Cuál es la probabilidad de que un individuo elegido al azar tenga RH negativo o sea del grupo universal?

c) Si un individuo elegido al azar tiene RH positivo, ¿cuál es la probabilidad de que pertenezca al grupo B?

SOLUCIÓN

Llamando 0 a tener grupo 0, A a tener grupo A, B a tener grupo B, AB a tener grupo AB, $+$ a tener RH positivo y $-$ a tener RH negativo:

a) $P(A \cap +) = 0{,}34$.

b) $P(- \cup 0) = 0{,}53$.

c) $P(B/+) = 0{,}12$.

---

54. Sabemos que el test de Bender para detectar alteraciones cerebrales tiene una sensibilidad del 88 % y una especificidad del 96 %. Por otro lado, sabemos que en una población la probabilidad de que un individuo elegido al azar tenga alteraciones cerebrales y además de positivo en el test es 0,08. Se pide:

    a) Calcular el porcentaje de personas que tienen alteraciones cerebrales en la población.

    b) ¿Es efectivo el test en esta población para detectar la ausencia de alteraciones cerebrales?

Solución

Llamando E a presentar alteraciones cerebrales y $+$ y $-$ a que el test de positivo y negativo respectivamente:

a) $P(E) = 0{,}0901$, es decir, un 9,09 %.

b) $P(\overline{E}/-) = 0{,}99$, lo cual indica que es muy efectivo para detectar la ausencia de alteraciones cerebrales.

---

55. Los estudios epidemiológicos indican que el 20 % de las personas mayores sufren un deterioro neuropsicológico. También se sabe que la tomografía axial computerizada (TAC) puede detectar ese trastorno en el 90 % de los que lo padecen, pero que también puede diagnosticarlo en el 5 % de las personas que no lo tienen. Si se toma una persona al azar y el TAC da positivo, ¿cuál es la probabilidad de que realmente esté enfermo?

Solución

Llamando E a sufrir el deterioro neuropsicológico y $+$ a que el TAC de positivo: $P(E/+) = 0,82$.

★ 56. 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.

Solución

★ 57. Para diagnosticar una misma enfermedad se utilizan dos test A y B completamente independientes. 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.

Solución

4. Variables Aleatorias

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

X	4	5	6	7	8
$f(x)$	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) $P(5 < X < 6)$.

Solución

a)

$$F(x) = \begin{cases} 0 & \text{si } x < 4, \\ 0{,}15 & \text{si } 4 \leq x < 5, \\ 0{,}5 & \text{si } 5 \leq x < 6, \\ 0{,}6 & \text{si } 6 \leq x < 7, \\ 0{,}85 & \text{si } 7 \leq x < 8, \\ 1 & \text{si } 8 \leq x. \end{cases}$$

b) $P(X < 7{,}5) = 0{,}85$, $P(X > 8) = 0$, $P(4 \leq x \leq 6{,}5) = 0{,}6$ y $P(5 < X < 6) = 0$.

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

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

Se pide:

a) Distribución de probabilidad.

b) Obtener:

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

Solución

a)

X	1	4	6
$f(x)$	1/5	11/20	1/4

b) $P(X=6)=1/4$, $P(X=5)=0$, $P(2<X<5{,}5)=11/20$ y $P(0\le X<4)=1/5$.

★ 60. Se realiza un experimento aleatorio consistente en inyectar un virus a tres tipos de ratas y observar si sobreviven o no. Se sabe que la probabilidad de que viva el primer tipo de rata es 0,5, la de que viva el segundo es 0,4 y la de que viva el tercero 0,3. 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 $P(X\le 1)$, $P(X\ge 2)$ y $P(X=1{,}5)$.

d) Calcular la media y la desviación típica.

Solución

a)

X	0	1	2	3
$f(x)$	0,21	0,44	0,29	0,06

b)

$$F(x)=\begin{cases} 0 & \text{si } x<0,\\ 0{,}21 & \text{si } 0\le x<1,\\ 0{,}65 & \text{si } 1\le x<2,\\ 0{,}94 & \text{si } 2\le x<3,\\ 1 & \text{si } 3\le x. \end{cases}$$

c) $P(X\le 1)=0{,}65$, $P(X\ge 2)=0{,}35$ y $P(X=1{,}5)=0$.

d) $\mu=1{,}2$ ratas, $\sigma^2=0{,}7$ ratas^2 y $\sigma=0{,}84$ ratas.

★ 61. En las siguientes tablas, indicar razonadamente, en los caso que sea posible, los valores de h que deben ponerse en cada tabla para que se tenga una distribución de probabilidad:

x	$f(x)$
−2	0,3
5	h
8	0,1

x	$f(x)$
1	−0,2
3	0,7
4	h

x	$f(x)$
2	h
3	0,5
4	0,6

En las tablas que constituyan una distribución de probabilidad:

a) Representar gráficamente la función de distribución.

b) Calcular media y desviación típica.

c) Calcular la mediana.

d) Si a los valores de x se multiplican por una constante $k<0$, ¿cómo se ve afectada la media? ¿Y la desviación típica?

Solución

La única tabla que puede ser una distribución de probabilidad es la primera para $h=0{,}6$.

b) $\mu = 3{,}2$ y $\sigma = 3{,}516$.

c) $Me = 5$.

d) $\mu_y = k\mu_x$ y $\sigma_y = |k|\sigma_x$.

---

62. 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.

Solución

Llamando X al número de pacientes curados de los 6 sometidos al tratamiento, se tiene que $X \sim B(6,\, 0{,}85)$.

a) $P(X = 3) = 0{,}041$.

b) $P(X \geq 4) = 0{,}9526$.

---

63. Se sabe que la probabilidad de que aparezca una bacteria en un mm^3 de cierta disolución es de 0,002. Si en cada mm^3 a los sumo puede aparecer una bacteria, determinar la probabilidad de que en un cm^3 haya como máximo 5 bacterias.

Solución

Llamando X al número de bacterias en 1 cm^3 de disolución, se tiene $X \sim B(1000,\, 0{,}002) \approx P(2)$. $P(X \leq 5) = 0{,}9834$.

---

64. 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?

65. 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?

★ 66. 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$\}$.

Solución

a) Si X es el número de llamadas en 2 segundos, entonces $X \sim P(4)$ y $P(X < 4) = 0{,}4335$.

b) Si Y es el número de llamadas en 3 segundos, entonces $Y \sim P(6)$ y $P(Y \geq 3) = 0{,}938$.

---

67. 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?

Solución

Llamando X al número de preguntas acertadas de las 7 contestadas al azar, se tiene $X \sim B(7, 1/3)$ y $P(X \geq 4) = 0{,}1733$.

★ 68. En un estudio sobre un determinado tipo de parásito que ataca el riñón de las ratas, se sabe que el número medio de parásitos en cada riñón es 3. Se pide:

a) Calcular la probabilidad de que una rata tenga más de 8 parásitos.
(Nota: se supone que una rata normal tiene dos riñones).

b) Si se tienen 10 ratas, ¿cuál es la probabilidad de que haya al menos 9 con parásitos?

Solución

a) Si X es el número de parásitos en una rata, $X \sim P(6)$ y $P(X > 8) = 0{,}1528$.

b) Si Y es el número de ratas con parásitos en un grupo de 10 ratas, entonces $Y \sim B(10, 0{,}9975)$ y $P(Y \geq 9) = 0{,}9997$.

★ 69. 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?

★ 70. 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?

Solución

a) Llamando X al número de pacientes en un turno, se tiene que $X \sim P(16)$ y $P(X > 5) = 0{,}9986$.

b) Llamando Y al número de horas en el que el servicio se vea desbordado porque lleguen más de 4 pacientes, se tiene que $Y \sim B(8, 0{,}0527)$ y $P(Y \geq 1) = 0{,}3515$.

★ 71. El síndrome de Turner es una anomalía genética que se caracteriza porque las mujeres tienen sólo un cromosoma X. Afecta aproximadamente a 1 de cada 2000 mujeres. Además, aproximadamente 1 de cada 10 mujeres con síndrome de Turner, como consecuencia, también sufren un estrechamiento anormal de la aorta. Se pide:

a) En un grupo de 4000 mujeres, ¿cuál es la probabilidad de que haya más de 3 afectadas por el síndrome de Turner? ¿Y de que haya alguna con estrechamiento de aorta como consecuencia de padecer el síndrome de Turner?

b) En un grupo de 20 chicas afectadas por el síndrome de Turner, ¿cuál es la probabilidad de que menos de 3 sufran un estrechamiento anormal de la aorta?

Solución

a) Si X es el número de mujeres afectadas por el síndrome de Turner en el grupo de 4000 mujeres, entonces $X \sim B(4000,\, 1/2000) \approx P(2)$ y $P(X > 3) = 0{,}1429$.
Si Y es el número de mujeres con estrechamiento de la aorta en el grupo de 4000 mujeres, entonces $Y \sim B(4000,\, 1/20000) \approx P(0{,}2)$ y $P(Y > 0) = 0{,}1813$.

b) Si Z es el número de mujeres con estrechamiento de la aorta en el grupo de 20 mujeres con el síndrome de Turner, entonces $Z \sim B(20,\, 1/10)$ y $P(Z < 3) = 0{,}6769$.

---

72. Se sabe que 2 de cada 1000 pacientes son alérgicos a un fármaco A, y que 6 de cada 1000 lo son a un fármaco B. Además, el 30 % de los alérgicos a B, también lo son a A. Si se aplican los dos fármacos a 500 personas,

a) ¿Cuál es la probabilidad de que no haya ninguna con alergia a A?

b) ¿Cuál es la probabilidad de que haya al menos 2 con alergia a B?

c) ¿Cuál es la probabilidad de que haya menos de 2 con las dos alergias?

d) ¿Cuál es la probabilidad de que haya alguna con alergia?

Solución

a) Llamando X_A al número de personas alérgicas al fármaco A en el grupo de 500 personas, se tiene que $X_A \sim B(500,\, 0{,}002) \approx P(1)$ y $P(X_A = 0) = 0{,}3678$.

b) Llamando X_B al número de personas alérgicas al fármaco B en el grupo de 500 personas, se tiene que $X_B \sim B(500,\, 0{,}006) \approx P(3)$ y $P(X_B \geq 2) = 0{,}8009$.

c) Llamando $X_{A\cap B}$ al número de personas alérgicas a ambos fármacos $A \cap B$ en el grupo de 500 personas, se tiene que $X_A \sim B(500,\, 0{,}0018) \approx P(0{,}9)$ y $P(X_{A\cap B} < 2) = 0{,}7725$.

d) Llamando $X_{A\cup B}$ al número de personas alérgicas a alguno de los fármacos $A \cup B$ en el grupo de 500 personas, se tiene que $X_A \sim B(500,\, 0{,}0062) \approx P(3{,}1)$ y $P(X_{A\cup B} \geq 1) = 0{,}9550$.

---

73. En una clase hay 40 alumnos de los cuales el 35 % son fumadores. Si se toma una muestra aleatoria con reemplazamiento de 4 alumnos, ¿cuál es la probabilidad de que haya al menos 1 fumador? ¿Cuál sería dicha probabilidad si la muestra se hubiese tomado sin reemplazamiento?

Solución

Si X es el número de fumadores en una muestra aleatoria con reemplazamiento de tamaño 4, entonces $X \sim B(4,\, 0{,}35)$ y $P(X \geq 1) = 0{,}8215$.
Si la muestra es sin reemplazamiento $P(X \geq 1) = 0{,}8364$.

★ 74. Suponiendo una facultad en la que hay un 60 % de chicas y un 40 % de chicos:

a) Si un año van 6 alumnos a hacer prácticas en un hospital, ¿qué probabilidad hay de que vayan más chicos que chicas?

b) En un período de 5 años, ¿cuál es la probabilidad de que más de 1 año no haya ido ningún chico?

Solución

a) Si X es el número de chicos, $X \sim B(6,\, 0{,}4)$ y $P(X \geq 4) = 0{,}1792$.

b) Si Y es el número de años que no ha ido ningún chico, $Y \sim B(5,\, 0{,}0467)$ y $P(Y > 1) = 0{,}0199$.

★ 75. 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?

★ 76. ¿Cuánto habría que restar a cada pregunta errada en un examen de tipo test de 5 preguntas con cuatro opciones y sólo una correcta, para que un individuo que responda al azar tenga una puntuación esperada de 0?

Solución

1/3.

★ 77. A un hospital llegan pacientes por la mañana a efectuarse extracciones de sangre. Se ha medido la frecuencia de llegada de los mismos en intervalos de 15 minutos. La distribución de probabilidad (medida de forma frecuentista) se muestra en la siguiente tabla:

X	0	1	2	3	4	5	6
$P(x)$	0,1	0,2	0,25	0,15	0,15	0,1	0,05

Se pide:

a) Calcular la probabilidad de que en un intervalo de 15 minutos lleguen 2 o más personas, y probabilidad de que lleguen menos de 8 personas.

b) ¿Cuál es el número medio esperado de personas que llegarán a sacarse sangre cada 15 minutos?

c) Suponiendo que el número de personas que llegan a sacarse sangre en 15 minutos sigue una distribución de Poisson de media la calculada en el apartado anterior, ¿cuál es la probabilidad de que llegue alguna en 15 minutos? ¿Y de que llegue alguna en 5 minutos?

Solución

Llamemos J al suceso consitente en que una persona con la lesión sea joven, C al suceso consistente en curarse, y A y B a los sucesos consistentes en aplicar las respectivas técnicas de rehabilitación:

a) Llamando X al número de personas que llegan en un intervalo de 15 minutos: $P(X \geq 2) = 0{,}7$ y $P(X < 8) = 1$.

b) $\mu = 2{,}55$ personas.

c) Suponiendo $X \sim P(2{,}55)$, $P(X \geq 1) = 0{,}9219$.
Llamando Y al número de personas que llegan en un intervalo de 5 minutos, $P(Y \geq 1) = 0{,}5726$.

★ 78. 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?

Solución

★ 79. 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?

Solución

80. Un empleado suele acudir al trabajo en cualquier instante entre las 6 y las 7 con igual probabilidad. Se pide:

a) Calcular la función de densidad de la variable que mide el instante en que acude a trabajar y dibujarla.

b) Calcular la función de distribución y dibujarla.

c) Calcular la probabilidad de que llegue entre las 6 y cuarto y las 6 y media.

d) Calcular la hora media a la que se espera que llegue.

Solución

a)

$$f(x) = \begin{cases} 0 & \text{si } x < 6, \\ 1 & \text{si } 6 \leq x \leq 7, \\ 0 & \text{si } 7 < x. \end{cases}$$

b)

$$F(x) = \begin{cases} 0 & \text{si } x < 6, \\ x - 6 & \text{si } 6 \leq x \leq 7, \\ 1 & \text{si } 7 < x. \end{cases}$$

c) $P(6{,}25 < X < 6{,}5) = 0{,}25$.

d) $\mu = 6{,}5$, es decir, a las 6 horas y media.

81. 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 $P(X \leq 120 \text{ mg/100 ml})$.

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.

Solución

a) $P(X \leq 120) = 0{,}9599$.

b) $P(90 < X < 120) = 0{,}9371$, es decir, un 93,71 %.

c) 100,64 mg/100 ml.

---

82. 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.

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

Solución

a) $\mu = 0{,}7$ mg y $\sigma = 0{,}25$ mg.

b) $P(X > 1) = 0{,}1151$ y el número de personas con más de 1 mg de bilirrubina es $0{,}1151 \cdot 40000 = 4604$.

★ 83. 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?

Solución

a) $P(X > 16) = 0{,}1587$ y el número de personas con tensión por encima de 16 mmHg es $0{,}1587 \cdot 20000 = 3174$.

b) Debe disminuir 3,75 mmHg.

★ 84. El peso de los recién nacidos no prematuros en una ciudad sigue una distribución normal de media y desviación típica desconocidas. Teniendo en cuenta que, de un total de 200 recién nacidos no prematuros, 15 han pesado más de 4 kg y 25 menos de $2,5$ kg:

 a) ¿Cuáles son la media y la desviación típica del peso?

 b) ¿Cuántos niños no prematuros habrán nacido con un peso entre 3 y 3,5 kg?

 c) Si los médicos consideran peligrosos los pesos por debajo del percentil 10, ¿cuál será dicho peso?, ¿cuántos niños habrán nacido con un peso por debajo de dicho percentil?

Solución

Llamando X al peso de los recién nacidos no prematuros, se tiene que $X \sim N(\mu, \sigma)$.

a) $\mu = 3{,}17$ kg y $\sigma = 0{,}58$ kg.

b) 66 niños.

c) $P_{10} = 2{,}43$ kg y habrán nacido 20 niños por debajo de este peso.

★ 85. Suponiendo que la duración del embarazo sigue una distribución normal de media y desviación típica desconocidas, y teniendo en cuenta que el 80 % de las mujeres dan a luz antes de 40 semanas, y que el 70 % lo hacen después de 38 semanas y 2 días, se pide:

a) Calcular la media y la desviación típica de la distribución dadas en número de semanas.

b) Suponiendo un hospital en el que se han producido 2000 nacimientos, ¿cuántos lo habrán hecho con más de 282 días de gestación?

c) Suponiendo una mujer que ha tenido 2 embarazos y que la duración del segundo ha sido independiente del primero, ¿cuál es la probabilidad de que alguno de los partos se haya producido antes de los 275 días de embarazo?

★ 86. El gasto mensual en medicamentos de las familias españolas antes de la crisis seguía una distribución normal $X \sim N(160, \sigma)$, mientras que ahora sigue una distribución normal $Y \sim N(\mu, 2\sigma)$. Sabiendo que antes de la crisis el 10 % de las familias gastaba más de 200 euros y que después el 40 % gastaba menos de 100 euros, se pide:

a) ¿Qué porcentaje de familias gastarán ahora entre 110 y 120 euros?

b) ¿Qué percentil de la distribución actual se corresponde con el tercer decil de la distribución de antes de la crisis?

Solución

$X \sim N(160,\ 31{,}25)$ y $Y \sim N(115{,}625,\ 62{,}50)$.

a) $P(110 \leq Y \leq 120) = 0{,}0638$.

b) El decil tercero en X es $D_3 = 143{,}75$ euros y se corresponde aproximadamente con el percentil 67 de Y.

★ 87. Según el teorema central del límite, se sabe que para muestras grandes ($n \geq 30$) la media muestral $\bar{x}$ sigue una distribución normal $N(\mu, \sigma/\sqrt{n})$, donde μ es la media de la población y σ su desviación típica.

Se sabe que en una población la elongación del triceps sural tiene una media de 60 cm y una desviación típica de 15 cm. Si se toma una muestra de tamaño 30, ¿cuál es la probabilidad de obtener una media muestral mayor de 62? Si se considera que una muestra es atípica si su media muestral está por debajo del percentil 5 de su distribución, ¿se puede considerar atípica una muestra de 60 individuos con $\bar{x} = 57$?

Solución

a) Llamando $\bar{X}$ a la variable que mide la elongación media del triceps sural en muestras de tamaño 30, $P(\bar{X} > 62) = 0{,}2327$.

b) Llamando $\bar{Y}$ a la variable que mide la elongación media del triceps sural en muestras de tamaño 60, $P_5 = 56{,}8$ cm, luego la muestra no es atípica.

---

88. Un test diagnóstico para niños de 10 años con problemas de lectura da puntuaciones que se distribuyen normalmente con una media de 80 y una varianza de 100. Se pide:

a) Dar la probabilidad de que un niño seleccionado al azar tenga una puntuación de:
   1) menos de 68;
   2) entre 75 y 90;

b) El test indica que el niño tiene un problema del lenguaje si su puntuación está por debajo del 10 % de la población que realizó el test. ¿Por debajo de qué puntuación el test diagnosticará problemas de lenguaje?

c) Si se selecciona una muestra de 16 niños:
   1) ¿Cuantos se espera que tengan una puntuación por encima de 68?
   2) ¿Cuál es la probabilidad de que su puntuación media supere los 84 puntos?

Solución

Llamando X a la puntuación del test diagnóstico, se tiene que $X \sim N(80, 10)$.

a) $P(X < 68) = 0{,}1151$ y $P(75 < X < 90) = 0{,}5328$.

b) $P_{10} = 67{,}18$ puntos.

c) $P(X > 68) = 0{,}8849$ y el número esperado niños con puntuaciones por encima de 68 es $0{,}8849 \cdot 16 = 14{,}16$.
$\bar{x} \sim N(80,\ 2{,}5)$ y $P(\bar{x} > 84) = 0{,}0548$.

5. Estimación de parámetros

89. Las notas en Estadística de una muestra de 10 alumnos han sido:

$$6{,}3,\ 5{,}4,\ 4{,}1,\ 5{,}0,\ 8{,}2,\ 7{,}6,\ 6{,}4,\ 5{,}6,\ 4{,}3,\ 5{,}2$$

Dar una estimación puntual de la nota media, de la varianza y del porcentaje de aprobados en la clase.

Solución

$\bar{x} = 5{,}81$ puntos, $\hat{s}^2 = 1{,}7721$ puntos^2 y $\hat{p} = 0{,}8$.

90. 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.

Solución

$\mu \in 79 \pm 1{,}742 = (76{,}258,\ 79{,}742)$.

91. 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 $\alpha = 0{,}01$.

Solución

$\mu \in 7{,}874 \pm 0{,}0426 = (7{,}8314,\ 7{,}9166)$.

92. 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.

    Solución

    249 individuos.

93. 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?

    Solución

    $17{,}96 \pm 1{,}278$ mg/mm^3 = (16,682mg/mm^3, 19,238mg/mm^3).

94. Se desea hacer un estudio estadístico sobre el número de hematíes en las mujeres. Se selecciona una muestra de 25 mujeres y se obtiene un número medio de hematíes de 4300000 con una desviación típica de 200000 (en cada milímetro cúbico de sangre). Calcular el intervalo de confianza para la media y la varianza del número de hematíes de las mujeres en la población, con un nivel de significación 0,1.

    Solución

    Intervalo de confianza del 95 % para la media: $\mu \in 4{,}3 \cdot 10^6 \pm 69850$ hematíes = (4230150, 4369850).
    Intervalo de confianza del 95 % para la varianza: $\sigma^2 \in (2747 \cdot 10^10, 7246 \cdot 10^10)$ hematíes^2.

95. 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.

    Solución

    $\sigma^2 \in$ (8,844g^2/l^2, 83,728g^2/l^2).

96. El tiempo que tarda en hacer efecto un analgésico sigue una distribución aproximadamente normal. En una muestra de 20 pacientes se obtuvo una media de 25,4 minutos y una desviación típica de 5,8 minutos. Calcular el intervalo de confianza del 90 % e interpretarlo. ¿Cuántos pacientes habría que tomar para poder estimar la media con una precisión de ±1 minuto?

    Solución

    Intervalo de confianza del 90 % para μ: (23,0992, 27,7008).
    Tamaño muestral para una precisión de ±1 minuto: $n = 96$.

97. Si el porcentaje de indivudos daltónicos de una muestra aleatoria es 18 %, ¿cuál será el mínimo tamaño muestral necesario para conseguir una estimación del porcentaje de daltónicos con una confianza del 95 % y un error menor del 3 %?

Solución

$n = 2520$.

98. 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?

★ 99. 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.

Solución

$p \in (0{,}1580,\ 0{,}4420)$ con un 95 % de confianza.

★100. 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?

Solución

101. 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.

Solución

$\mu_{x_1-x_2} \in 24{,}0909 \pm 19{,}4766$ mg/dl $= (4{,}6143\text{mg},\ 43{,}5675\text{mg/dl})$.

102. 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.

SOLUCIÓN

$\mu_1 - \mu_2 \in 3{,}667 \pm 4{,}712$ días $= (-1{,}045$ días, $8{,}379$ días$)$.

★103. Un equipo de investigación está interesado en ver si una droga reduce el colesterol en la sangre. Con tal fin se 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 la media de dicha variable.

b) A la vista del intervalo anterior, ¿hay pruebas significativas de que la droga disminuye el nivel de colesterol en sangre?

SOLUCIÓN

a) $\mu_{x_1 - x_2} \in 7{,}4 \pm 7{,}572$ mg/100ml $= (-0{,}172$mg/100ml, $14{,}972$mg/100ml$)$.

b) No se puede afirmar que la droga disminuya el colesterol con un 95 % de confianza.

★104. 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?

105. Para estudiar si la estación del año influye en el estado de ánimo de la gente, se ha tomado una muestra 12 personas y se ha medido su nivel de depresión en verano e invierno mediante un cuestionario con puntuaciones de 0 a 100 (a mayor puntuación mayor depresión). Los resultados obtenidos fueron:

Invierno	65	72	84	31	80	61	75	52	73	79	85	71
Verano	60	51	81	45	62	53	70	52	64	51	67	62

¿Se puede afirmar que la estación del año influye en el estado de ánimo de la gente con un 99 % de confianza? ¿Cómo influye?

SOLUCIÓN

No se puede afirmar que la estación influya en el estado de ánimo ya que la media de la diferencia está en $(-0{,}7498, 19{,}0831)$ con un 99 % de confianza.

---

106. Un psicólogo está estudiando la concentración de una encima en la saliba como un posible indicador de la ansiedad crónica. En un experimento se tomó una muestra de 12 neuróticos por ansiedad y otra de 10 personas con bajos niveles de ansiedad. En ambas muestras se midió la concentración de la encima, obteniendo los siguientes resultados:

Con ansiedad:	2,60	2,90	2,60	2,70	3,91	3,15	3,94	2,46	2,91	3,88	3,55	3,96
Sin ansiedad:	2,37	1,10	2,55	2,64	2,20	2,12	2,47	2,90	1,66	2,72		

¿Se puede concluir a partir de estos datos que la población de neuróticos con ansiedad y la población de personas sin ansiedad son diferentes en el nivel medio de concentración de encimas? Justificar la respuesta.

Solución

$\frac{\sigma_1^2}{\sigma_2^2} \in (0{,}3383, 4{,}7486)$ con un nivel de confianza del 95 %, luego se puede suponer que las varianzas son iguales.
$\mu_1 - \mu_2 \in (0{,}4296, 1{,}4510)$ con un 95 % de confianza, luego se puede concluir que hay diferencias entre las medias.

★107. 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.

Solución

- $p_A - p_B \in (-0{,}2126, 0{,}0626)$ con un nivel de confianza del 95 %.
- No puede afirmarse que un tratamiento sea más eficaz que otro pues la diferencia de medias podría ser positiva, negativa o cero.

★108. 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?

Solución

- $p_R \in (0{,}3799,\ 0{,}4548)$ y $p_U \in (0{,}4185,\ 0{,}4856)$ con un 99 % de confianza.
- $p_R - p_U \in (-0{,}0729,\ 0{,}0036)$ con un nivel de confianza del 95 %, luego no se puede afirmar que haya diferencias en las proporciones de niños menores de 5 años con sobrepeso.

109. Se cree que el consumo de tabaco va ligado al consumo de alcohol y para corroborar esta hipótesis se ha realizado un estudio en el que se han obtenido los siguientes datos

	Bebedores	No bebedores
Fumadores	487	137
No fumadores	312	365

¿Se puede afirmar que existe relación entre el consumo de tabaco y el de alcohol? Justificar la respuesta.

★110. 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.

Solución

6. Contraste de hipótesis

111. 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?

Repetir los cálculos si se toma una muestra de 100 personas y se establece que la vacuna es mejor si más de 75 de los vacunados conservan la protección al cabo de 2 años.

Nota: Aproximar la distribución binomial mediante una distribución normal.

Solución

Contraste: $H_0 : p = 0{,}6$, $H_1 : p > 0{,}6$. Muestra de tamaño 10:

- $P(\text{Rechazar } H_0/H_0) = 0{,}0464$.
- $P(\text{Aceptar } H_0/H_1) = 0{,}6242$.

Muestra de tamaño 100:

- $P(\text{Rechazar } H_0/H_0) = 0{,}0011$.
- $P(\text{Aceptar } H_0/H_1) = 0{,}1056$.

---

112. Se sabe que el tiempo de reacción ante un estímulo sigue una distribución normal de media 30 ms y desviación típica 10 ms. Se cree que la alcoholemia aumenta el tiempo de reacción de los sujetos, y para comprobar esta hipótesis se ha tomado una muestra aleatoria de 40 individuos a los que se les ha inducido una alcoholemia de 0,8 g/l y en los que se ha apreciado un tiempo medio de respuesta de 35 ms y una desviación típica de 12 ms. ¿Se puede afirmar que una alcoholemia de 0,8 gm/l influye en el tiempo medio de respuesta con un riesgo $\alpha = 0{,}05$? ¿Y con un riesgo $\alpha = 0{,}01$?

    ¿Cuál será la potencia del contraste para detectar una difererencia en la media del tiempo de reacción de 4 ms? ¿Cuál debería ser el tamaño muestral para aumentar la potencia hasta un 90 %?

    Solución

    Contraste para la media: $H_0 : \mu = 30$, $H_1 : \mu > 30$.
    Región de aceptación para $\alpha = 0{,}01$: $z < 2{,}3263$.
    Estadístico del contraste: $z = 2{,}6020$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula tanto para $\alpha = 0{,}01$ y con mayor motivo para $\alpha = 0{,}05$, de manera que se concluye que la alcoholemia influye en el tiempo de respuesta.
    Potencia del contraste para $\delta = 4$: $1 - \beta = 1 - 0{,}5966 = 0{,}4034$.
    El tamaño muestral para, $\alpha = 0{,}05$, $\delta = 4$ y una potencia del 90 % es $n = 80$.

---

113. 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 13 días y una cuasivarianza de 9. ¿Contradice lo observado en la muestra la afirmación del fisioterapeuta para un $\alpha = 0{,}05$?

    Solución

    Contraste para la media: $H_0 : \mu = 15$, $H_1 : \mu < 15$.
    Región de aceptación para $\alpha = 0{,}05$: $-1{,}6444 < z$.
    Estadístico del contraste: $z = -4$. Como cae fuera de la región de aceptación se rechaza la hipótesis nula y se confirma confirma la afirmación del fisioterapeuta.

---

114. Se cree que el nivel medio de protrombina en plasma de una población normal tiene una media de 19mg/100ml y una desviación típica de 4mg/100ml. Para contrastar estas hipótesis se tomó una muestra de 8 individuos en los que se obtuvieron los siguientes niveles de protrombina en plasma:

    $$16{,}3 - 18{,}4 - 20{,}0 - 17{,}6 - 15{,}4 - 23{,}7 - 17{,}8 - 19{,}5$$

¿Se pueden aceptar ambas hipótesis con un riesgo $\alpha = 0{,}1$?

Solución

Contraste para la media: $H_0 : \mu = 19$, $H_1 : \mu \neq 19$.
Región de aceptación para $\alpha = 0{,}1$: $-1{,}8946 < t < 1{,}8946$.
Estadístico del contraste: $t = -0{,}4552$. Como cae dentro de la región de aceptación, se mantiene la hipótesis de que la media es 19mg/100ml.
Contraste para la varianza: $H_0 : \sigma = 4$, $H_1 : \sigma \neq 4$.
Región de aceptación para $\alpha = 0{,}1$: $2{,}1673 < j < 14{,}0671$.
Estadístico del contraste: $j = 2{,}8743$. Como cae dentro de la región de aceptación, también se mantiene la hipótesis de que la desviación típica es de 4mg/100ml.

115. 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 $\alpha = 0{,}01$).

Solución

Contraste para la proporción: $H_0 : p = 0{,}1$, $H_1 : p > 0{,}1$.
Región de aceptación para $\alpha = 0{,}01$: $z > 2{,}3263$.
Estadístico del contraste: $z = 1{,}6667$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y se concluye que no hay pruebas suficientes para retirar la vacuna.

116. 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.

Solución

Contraste para la comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.
Región de aceptación para $\alpha = 0{,}05$: $-1{,}96 < z < 1{,}96$.
Estadístico del contraste $z = 0{,}6677$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y no se pude concluir que la vitamina C tenga influencia en el tratamiento del cáncer.

117. 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 $\alpha = 0{,}05$.

Solución

Contraste de comparación de medias: $H_0 : \mu_1 < \mu_2 + 12$, $H_1 : \mu_1 \geq \mu_2 12$.
Intervalo de confianza para la diferencia de medias: $\mu_1 - \mu_2 \in (6{,}5659, 11{,}2341)$ con un 95 % de confianza, luego hay diferencias significativas entre el número medio de mediciones, pero no se puede afirmar que sean mayores de 12 mediciones.

118. 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.

119. 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 $\alpha = 0{,}10$.

    Solución

    Contraste de comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.
    Región de aceptación para $\alpha = 0{,}01$: $-1{,}6449 < z < 1{,}6449$.
    Estadístico del contraste: $z = 2{,}4026$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se concluye que hay diferencias significativas entre el consumo de alcohol de chicos y chicas.

120. Un psicólogo está estudiando la concentración de una encima en la saliba como un posible indicador de la ansiedad crónica. En un experimento se tomó una muestra de 12 neuróticos por ansiedad y otra de 10 personas con bajos niveles de ansiedad. En ambas muestras se midió la concentración de la encima, obteniendo los siguientes resultados:

Con ansiedad:	2,60	2,90	2,60	2,70	3,91	3,15	3,94	2,46	2,91	3,88	3,55	3,96
Sin ansiedad:	2,37	1,10	2,55	2,64	2,20	2,12	2,47	2,90	1,66	2,72		

    ¿Se puede concluir a partir de estos datos que la población de neuróticos con ansiedad y la población de personas sin ansiedad son diferentes en el nivel medio de concentración de encimas? Justificar la respuesta.

    Solución

    Contraste de comparación de varianzas: $H_0 : \sigma_1 = \sigma_2$, $H_1 : \sigma_1 \neq \sigma_2$.
    Región de aceptación para $\alpha = 0{,}05$: $0{,}2787 < f < 3{,}9121$.
    Estadístico del contraste: $f = 1{,}2138$. Como cade dentro de la región de aceptación, se acepta la hipótesis nula y se concluye que las varianzas son iguales.
    Contraste de comparación de medias: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.
    Región de aceptación para $\alpha = 0{,}05$: $-2{,}0860 < t < 2{,}0860$.
    Estadístico del contraste: $t = 3{,}8407$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se puede concluir que hay diferencia entre la concentración media de encimas para neuróticos con ansiedad y sin ansiedad.

121. En una investigación para ver la efectividad de una nueva droga antidepresiva, se ha tomado un muestra de 15 pacientes depresivos que han completado un cuestionario para detectar el nivel depresivo, antes y después de recibir la droga. En la puntuación del cuestionario los valores menores indican una mayor depresión. Los resultados obtenidos han sido:

Antes:	18	21	16	19	14	23	16	14	21	18	17	14	16	14	20
Después:	23	20	17	20	16	22	18	18	21	16	19	20	15	15	21

    Realizar un constraste para averiguar si la droga tiene un efecto positivo sobre la depresión. ¿Qué tamaño muestral sería necesario para detectar una diferencia en la puntuación como la que hay entre las medias de las muestras?

Solución

Contraste para la media de la diferencia entre antes y después: $H_0 : \mu = 0$, $H_1 : \mu < 0$.
Región de aceptación par $\alpha = 0{,}05$: $-1{,}7613 < t$.
Estadístico del contraste: $t = -2{,}2563$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se puede afirmar que la droga reduce la depresión.
El tamaño muestral para detectar una diferencia de $\delta = \bar{x}_1 - \bar{x}_2 = 17{,}4 - 18{,}7333 = -1{,}3333$ con una potencia del 90 % es $n = 26$ individuos.

★122. Para ver si la ley antitabaco está influyendo en el número de cigarros que se fuman mientras se está en los bares se seleccionó una muestra en la que se midió el número de cigarros fumados por hora mientras se estaba en un bar antes de la entrada en vigor de la ley y otra muestra distinta en la que también se midió el número de cigarros fumados por hora después de la entrada en vigor de la ley (se entiende que con la ley en vigor los cigarros se fuman en el exterior de los bares). Los resultados aparecen en la siguientes tablas:

Antes	
Cigarros	Personas
0-1	12
1-2	21
2-3	20
3-4	8

Después	
Cigarros	Personas
0-1	22
1-2	18
2-3	8
3-4	4

Se pide:

a) Calcular el intervalo de confianza del 99 % para el número medio de cigarros fumados por hora en los bares antes de la entrada en vigor de la ley. ¿Cuántos individuos serían necesarios para poder estimar dicha media con un margen de error no mayor de $\pm 0{,}1$ cigarros por hora?

b) Contrastar si la nueva ley ha reducido significativamente el consumo medio de tabaco en los bares. ¿Cuánto vale el p-valor del contraste?

Solución

a) El intervalo de confianza del 99 % para el número medio de cigarros fumados por hora en los bares antes de la entrada en vigor de la ley es (1,5789, 2,2079). El tamaño muestral necesario para estimar la media con un margend e error no mayor de $\pm 0{,}1$ cigarros es $n = 603$ individuos.

b) Contraste para de comparación de medias: $H_0 : \mu_x = \mu_y$, $H_1 : \mu_x > \mu_y$.
Región de aceptación para $\alpha = 0{,}05$: $Z \leq z_\alpha = 1{,}6449$.
Estadístico del contraste: $z = 2{,}8447$. Como cae dentro fuera de la región de aceptación, se rechaza la hipótesis nula y se concluye que la ley ha reducido el consumo medio de cigarros por hora.
El p-valor del contraste vale 0,0022.

★123. En un estudio sobre el reparto de género de las tareas domésticas se ha medido el número medio de horas diarias que se destinan a las tareas domésticas en un grupo de personas, obteniendo los siguientes resultados:

Mujeres:	3,2	3,1	2,7	4,4	3,7	3,9	2,4	3,6
Hombres:	3,3	2,1	1,7	2,4	1,6	1,8	2,7	

Contrastar si las mujeres dedican más tiempo que los hombres a las tareas domésticas. ¿Entre qué valores estará la diferencia del tiempo medio destinado a tareas domésticas entre mujeres y hombres para un 95 % de confianza?

Solución

Estadísticos muestrales:
$\bar{x}_M = 3{,}375$, $s^2_M = 0{,}3744$, $\hat{s}^2_M = 0{,}4279$, $\hat{s}_M = 0{,}654$,
$\bar{x}_H = 2{,}2286$, $s^2_H = 0{,}3249$, $\hat{s}^2_H = 0{,}379$, $\hat{s}_H = 0{,}616$.

Contraste de comparación de varianzas: $H_0: \ \sigma^2_M = \sigma^2_H, \quad H_1: \ \sigma^2_M \neq \sigma^2_H,$
Estadístico del contraste: $F = 1{,}129$,
Región de aceptación para $\alpha = 0{,}05$: (0,1954, 5,6955).
Por tanto, se acepta la hipótesis nula y se supone que las varianzas son iguales.

Contraste de comparación de medias: $H_0: \ \mu_M = \mu_H, \quad H_1: \ \mu_M > \mu_H,$
Estadístico del contraste: $T = 3{,}479$.
Región de aceptación para $\alpha = 0{,}05$: $(1{,}7709, \infty)$.
Por tanto, se rechaza la hipótesis nula y se concluye que las mujeres dedican más tiempo que los hombres a tareas domésticas.

Intervalo de confianza del 0,95 % para $\mu_M - \mu_H$: (0,4345, 1,8583).

Nota: Los problemas marcados con una estrella (★) son problemas de exámenes de otros años.