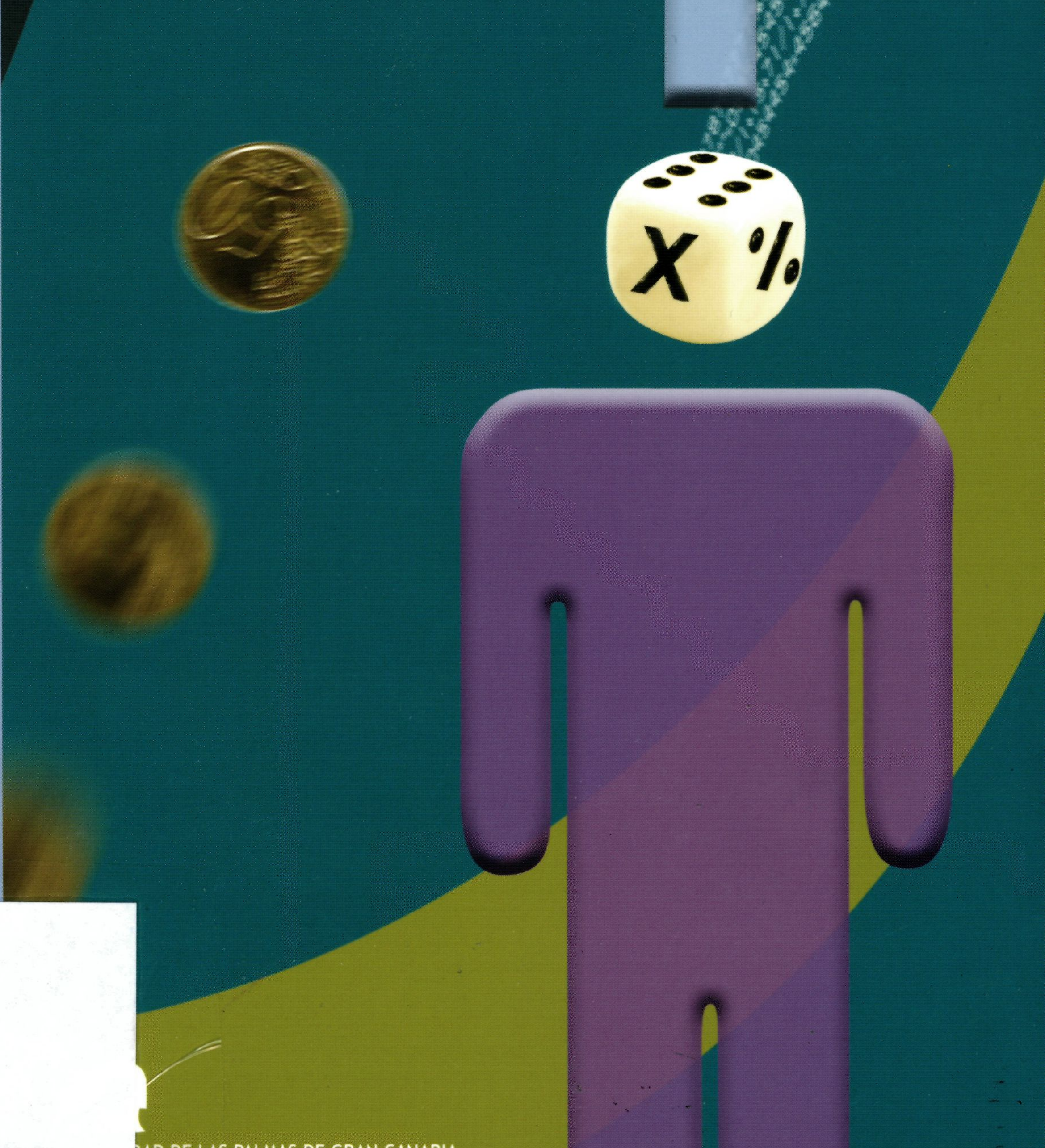
Problemas resueltos de cálculo de probabilidades y estadística

Inmaculada Luengo Merino



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Servicio de Publicaciones

PROBLEMAS RESUELTOS DE CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA

INMACULADA LUENGO MERINO

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Servicio de Publicaciones

La Caja de CANARIAS

2006

Maquetación y diseño: Servicio de Publicaciones de la ULPGC
serpubli@ulpgc.es

ISBN: 84-96502-40-6
Depósito Legal: GC 6-2006

Impresión: COMETA

A Carlos siempre,
y a David y Bárbara.
Y a mi madre.

Índice

Presentación 9

Relación de contenidos que abarcan los ejercicios del presente libro 13

Enunciados 1. Álgebra de sucesos, probabilidad y variables aleatorias 19

Enunciados 2. Inferencia estadística 49

Soluciones 1. Álgebra de sucesos, probabilidad y variables aleatorias 81

Soluciones 2. Inferencia estadística 183

Presentación

El presente libro es una recopilación de ejercicios que han sido presentados en algún examen de la asignatura Cálculo de Probabilidades y Estadística de segundo curso de la Escuela Universitaria de Informática de la Universidad de Las Palmas de Gran Canaria y por tanto, abarcan, o lo pretenden, precisamente el programa de esa asignatura, que presento explícitamente en las páginas siguientes.

Muchos libros que presentan ejercicios de Probabilidad y/o Estadística y en general de casi todas las materias, vienen ordenados por temas u objetivos de conocimiento. Eso que es imprescindible en un libro de teoría, no es siempre tan bueno ni recomendable cuando lo que se presenta son problemas, porque cuando el alumno conoce de antemano a qué tema se refiere un determinado problema limita enormemente su campo de decisión sobre el mismo, ya sabe que sólo entre los contenidos de ese tema debe buscar las herramientas de resolución del problema que tiene delante y si le presentas un problema similar en otro contexto en que tiene esas y otras herramientas mezcladas ya no se siente tan capaz de elegir la manera correcta de resolución.

Por este motivo, los problemas aquí presentados están perfectamente desordenados, los he ido copiando tal cual me surgía y sólo han sido divididos en dos grandes grupos tal y como está dividida la asignatura por su contenido: Problemas de Cálculo de Probabilidades y Problemas de Estadística.

En la resolución de cada uno de ellos he pretendido ser lo más clara posible y explicar en cada caso el método de resolución elegido y las razones que llevaron a elegirlo. El contenido teórico necesario, así como la nomenclatura se corresponde con los libros:

1. *Probabiliad y Estadística*, de Walpole y Myers, editado por McGraw-Hill.
2. *Probabiliad y Estadística,* de G.C. Canavos, editado por McGraw-Hill.

Muchos de estos problemas no son de aplicación inmediata y sin más de cierto conocimiento puntual, sino que hace falta primero la comprensión clara de la situación descrita en el enunciado, cosa que puede parecer una perogrullada, pero tras muchos años corrigiendo exámenes sé que la mala comprensión del enunciado suele ser uno de los problemas que frecuentemente se presenta: la mayoría de los errores son de planteamiento. Con frecuencia el alumno se lanza a resolver un problema sin haber comprendido la situación descrita en el enunciado. No se trata de si tiene conocimientos probabilísticos, o estadísticos en su caso, para lograr hacerlo correctamente: es un problema semántico.

Si deseamos hacer una silla no debemos coger madera, martillo y clavos y lanzarnos a martillar. Para hacer una silla no se empieza directamente martillando, tenemos que conocer qué tipo de silla queremos y planear detenidamente donde hay que clavar los clavos y para qué queremos cierto clavo en un sitio y no en otro y **sólo entonces** empezaremos a martillar, luego además hay que ser diestro con el martillo, pero solo después de ser diestro con el cerebro.

Mi recomendación personal al posible alumno que pretenda usar este libro quizás para comprobar su preparación básica en esta materia es la misma que repito y repetiré a mis alumnos: nunca, insisto nunca, debes lanzarte a resolver el problema inmediatamente después de haberlo leído; suelta el bolígrafo, levanta la cabeza y si el problema habla de un experimento con ratas, ponte la bata blanca, mira las jaulas encima de la mesa, observa las demás cosas que necesitas en el laboratorio, si se trata de inyectar un fármaco mírate con la jeringuilla en la mano, puedes escribir algo, poner nombre a las cosas, etc. Sólo después de que te hayas visto allí empieza a mirar la situación con ojos estadísticos o probabilísticos. Si el enunciado habla de tornillos, ponte inmediatamente un mono (el color queda a tu elección), mira la máquina que los fabrica, como entra el metal por un lado, como aquel brazo articulado les da forma, como pasan un control de calidad, el proceso de empaquetado... todo lo *in situ* que tu imaginación sea capaz de ponerte, y de nuevo, sólo entonces empieza a mirarlo con ojos de estadístico.

No pretendo ser graciosa ni original cuando digo y repito incansablemente lo anterior, ni siquiera pretendo ser pesada. Sólo pretendo ser útil.

Los enunciados de todos o casi todos los ejercicios aquí reunidos han sido sacados de diversos libros de problemas sobre estas materias. De este modo, los enunciados no son originales. La resolución es mía. No sé si algunos de ellos venían ya resueltos en el libro de donde los leí, pero si fue así he perdido el tiempo con ellos porque las soluciones aquí presentadas han sido elaboradas enteramente por mí. Dado que son simples ejercicios, espero que nadie sentirá usurpados sus derechos sobre la propiedad intelectual, o cosa similar.

Es altamente probable que haya errores de cálculo en algunos de los ejercicios. Agradecería a quien los detecte que me lo haga saber a mluengo@dis.ulpgc.es.

Relación de contenidos que abarcan los ejercicios del presente libro

Lección 1. Introducción a la estadística descriptiva

1.1. Introducción.

1.2. Descripción gráfica de los datos.

1.3. Medidas numéricas descriptivas.

Lección 2. Conceptos de probabilidad

2.1. Introducción.

2.2. La definición clásica de probabilidad.

2.3. Definición de probabilidad como frecuencia relativa.

2.4. Interpretación subjetiva de la probabilidad.

2.5. Desarrollo axiomático de la probabilidad.

2.6. Probabilidades conjunta, marginal, y condicional.

2.7. Sucesos estadísticamente independientes.

2.8. Teorema de Bayes.

2.9. Repaso de combinatoria.

Lección 3. Variables aleatorias y distribuciones de probabilidad

3.1. El concepto de variable aleatoria.

3.2. Distribuciones de probabilidad de variables aleatorias discretas.

3.3. Distribuciones de probabilidad de variables aleatorias continuas.

3.4. Valor esperado de una variable aleatoria.

3.5. Momentos de una variable aleatoria.

3.6. Otras medidas de tendencia central y dispersión.

3.7. Funciones generadoras de momentos.

Lección 4. Algunas distribuciones discretas de probabilidad

4.1. Introducción.

4.2. La distribución Binomial.

4.3. La distribución de Poisson.

4.4. La distribución Geométrica.

4.5. La distribución Binomial negativa.

4.6. La distribución Multinomial.

4.7. La distribución Hipergeométrica Multivariada.

Lección 5. Algunas distribuciones continuas de probabilidad

5.1. Introducción.

5.2. La distribución Uniforme.

5.3. La distribución Exponencial.

5.4. La distribución Gamma.

5.5. La distribución Beta.

5.6. La distribución Normal.

5.7. Algunas distribuciones relacionadas con la Normal.

5.8. La distribución de una función de una variable aleatoria.

Lección 6. Distribuciones conjuntas de probabilidad

6.1. Introducción.

6.2. Variables aleatorias bidimensionales.

6.3. Distribuciones marginales.

6.4. Momentos de una variable aleatoria bidimensional.

6.5. Variables estadísticamente independientes.

6.6. Distribuciones de probabilidad condicional.

6.7. La distribución Normal bivariada.

Lección 7. Muestras aleatorias y distribuciones de muestreo

7.1. Introducción.

7.2. Muestras aleatorias.

7.3. Estadísticos: Estimadores.

7.4. La distribución de la media muestral.

7.5. La distribución de la cuasivarianza muestral.

7.6. La distribución de la diferencia de dos medias muestrales.

Lección 8. Estimación puntual y por intervalo

8.1. Introducción.

8.2. Propiedades deseables de los estimadores puntuales.

8.2.1. Estimadores insesgados.

8.2.2. Estimadores consistentes.

8.2.3. Estimadores de mínima varianza.

8.3. Métodos de estimación puntual.

8.3.1. Método de la máxima verosimilitud.

8.3.2. Método de los momentos.

8.4. Estimación por intervalo.

8.4.1. Intervalos de confianza para la media de una población normal.

8.4.2. Intervalos de confianza para el parámetro de una binomial.

8.4.3. Intervalos de confianza para la diferencia de medias de dos poblaciones independientes.

8.4.4. Intervalos de confianza para la varianza de una población normal.

8.4.5. Intervalos de confianza para el cociente de varianzas de dos poblaciones normales independientes.

Lección 9. Pruebas de hipótesis estadísticas

9.1. Introducción.

9.2. Conceptos básicos.

9.3. Tipos de regiones críticas. Función de potencia.

9.4. Principios generales para probar una hipótesis nula simple contra una alternativa uni- o bilateral.

9.5. Pruebas de hipótesis respecto a la media en poblaciones normales.

9.5.1. Pruebas para una muestra.

9.5.2. Pruebas para dos muestras independientes.

9.5.3. Pruebas para dos muestras: datos emparejados.

9.6. Pruebas de hipótesis respecto a la varianza en poblaciones normales.

9.6.1. Pruebas para una muestra.

9.6.2. Pruebas para dos muestras.

9.7. Pruebas de hipótesis respecto a las proporciones de dos poblaciones binomiales independientes.

Lección 10. Pruebas de bondad de ajuste y análisis de tablas de contingencia

10.1. Introducción.

10.2. La prueba de bondad de ajuste de la Chi-cuadrada.

10.3. La prueba de bondad de ajuste de Kolmogorov-Smirnov.

10.4. La prueba de la Chi-cuadrada para el análisis de tablas de contingencia.

Lección 11. Diseño y análisis de experimentos estadísticos

11.1. Introducción.

11.2. Experimentos estadísticos.

11.3. Diseños estadísticos.

11.4. Análisis de experimentos unifactoriales en un diseño completamente aleatorio.

11.5. Análisis de experimentos unifactoriales en un diseño en bloque completamente aleatorizado.

11.6. Experimentos factoriales.

Lección 12. Análisis de regresión: el modelo lineal simple

12.1. Introducción.

12.2. Significado de la regresión y suposiciones básicas.

12.3. Estimación por mínimos cuadrados para el modelo lineal simple.

12.4. Propiedades de los estimadores de mínimos cuadrados.

12.5. Inferencia estadística para el modelo lineal simple.

12.6. El uso del análisis de la varianza.

12.7. Correlación lineal.

Lección 13. Métodos no parámetros

13.1. Introducción.

13.2. Pruebas para una muestra.

13.2.1. Test de los signos.

13.2.2. Test de rangos con signo.

13.3. Pruebas para dos muestras.

13.3.1. Test de la suma de rangos para dos muestras independientes.

13.3.2. Test para dos muestras con datos emparejados.

13.4. Pruebas para k muestras.

13.4.1. Test de Kruskal-Wallis para k muestras independientes.

13.4.2. Test de Friedman para k muestras con datos emparejados.

Lección 14. Programación y análisis estadísticos básicos con SPSS

14.1. Introducción.

14.2. Programación con SPSS.

14.2.1. Normas generales de sintaxis.

14.2.2. Comandos de ayuda e información.

14.2.3. Comandos de configuración del sistema.

14.2.4. Comandos iniciales de definición de datos.

14.2.5. Recodificación y generación de variables.

14.2.6. Modificación de los ficheros de sistema.

14.3. Análisis estadísticos básicos.

14.3.1. Estadística descriptiva.

14.3.2. Correlación lineal.

14.3.3. Gráficas planas.

14.3.4. Inferencia estadística.

14.3.5. Análisis de la varianza.

14.3.6. Tests no paramétricos.

Enunciados 1.

Álgebra de sucesos, probabilidad y variables aleatorias

1. Diez amigos quieren sortear quién paga la cena: cada uno coge una moneda, que suponemos equilibrada, de su bolsillo y las lanzan todos a la vez. Si todos menos uno obtienen el mismo resultado, éste paga la cena; en caso contrario se repite el lanzamiento.

a) Probabilidad de que en una determinada jugada haya perdedor.

b) Distribución de probabilidad de las tiradas hasta para obtener un perdedor y el número medio de tiradas necesarias.

2. Dos de los 8 camiones de ocasión que va a comprar una compañía no tienen el motor en perfecto estado, hecho fácil de apreciar con una simple prueba. ¿Cuál de los tres casos siguientes es más probable?

a) Encontrar los dos camiones defectuosos eligiendo al azar, sin remplazamiento, cuatro de ellos.

b) Encontrar un camión defectuoso entre tres elegidos al azar, sin remplazamiento.

c) No encontrar ningún camión defectuoso entre dos elegidos al azar, y con remplazamiento.

3. Un estudio calcula en 0.34 la probabilidad de que un camión pase la ITV a la primera; si el camión tiene menos de dos años esta probabilidad aumenta a 0.82. Sabiendo que el 60% del parque de camiones tiene más de dos años, calcular la probabilidad de que un camión de más de dos años pase la ITV a la primera.

4. Supongamos que hay una prueba para diagnosticar el cáncer que da positivo el 95% de las veces cuando una persona tiene cáncer y da negativo el 95% de las veces cuando se aplica en personas que no tienen la enfermedad. Si la probabilidad de que una persona tenga cáncer es 0.005, ¿cuál es la probabilidad de que una persona tenga realmente cáncer cuando la prueba ha dado positiva?

5. Sea un circuito con n componentes conectadas en paralelo que funcionan de manera independiente; es decir, el sistema funciona si y solo si funciona al menos uno de los componentes. Si el tiempo de duración de cada componente se distribuye normalmente con media de 50 horas y desviación típica de 5 horas, ¿cuál debe ser el número de componentes para que la probabilidad de que el sistema falle durante las primeras 55 horas sea aproximadamente 0.01?

6. En el jardinero del Sr. Rodríguez no se puede confiar: la probabilidad de que olvide regar el rosal durante la ausencia del Sr. Rodríguez es 2/3.

El rosal es muy delicado: si se lo riega tiene igual probabilidad de progresar que de secarse, pero si no es regado tiene solo un 0.25 de probabilidad de progresar.

A su regreso, el Sr. Rodríguez se encuentra con que el rosal está seco. ¿Cuál es la probabilidad de que el jardinero se haya olvidado de regarlo?

7. Sea un dado tal que la probabilidad de las distintas caras es proporcional al número de puntos inscritos en ellas. Hallar la probabilidad de obtener con este dado un número par.

8. Cierto profesor lleva siempre en el bolsillo dos cajas iguales de N fósforos. Cada vez que va a encender un cigarrillo toma al azar una de las cajas; al cabo de cierto tiempo una de las cajas está vacía. Calcular la probabilidad de que en ese momento la otra caja contenga r fósforos, supuesto que se da cuenta de que la caja se queda vacía al coger el último fósforo de la misma.

9. Las probabilidades de que tres hombres hagan diana en una competición de tiro son, respectivamente, 1/5, 1/4 y 1/3. Cada hombre dispara una vez. Calcular:

a) Probabilidad de que solo uno de ellos haga diana.

b) Si sabemos que solo uno de ellos ha hecho diana, probabilidad de que haya sido el tercero.

10. Un señor piensa comprar acciones de cierta compañía. La cotización de las acciones en bolsa durante los seis meses inmediatamente anteriores es un dato de interés y se observa que dicha cotización se relaciona con el Producto Interior Bruto (PIB). Si el PIB aumenta, la probabilidad de que las acciones aumenten de valor es de 0.8; si el PIB se mantiene esa prob. es de 0.2; si el PIB disminuye la prob. es de 0.1. Si para los próximos seis meses las probs. de que el PIB suba, se mantenga, o baje, son 0.4, 0.3 y 0.3, respectivamente, calcular la probabilidad de que esas acciones aumenten de valor en los próximos seis meses.

11. Se sabe que el 40% de personas de cierto grupo grande estarían dispuestas a formar parte del comité ecologista. Para constituir el comité, formado por 5 personas, seleccionamos al azar una persona del grupo y le preguntamos si quiere formar parte mismo, hasta encontrar las cinco personas buscadas. Calcular:

a) Probabilidad de que tras quince consultas aún no esté formado el comité.

b) Probabilidad de que sean necesarias doce consultas para formar el comité.

12. El gerente de un restaurante que solo atiende mediante reserva, sabe por experiencia que el 15% de las personas que reservan una mesa, no asisten. El restaurante solo tiene 20 mesas, pero acepta 25 reservas. ¿Cuál es la probabilidad de que determinado día, a todas las personas que acuden al restaurante se les asigne mesa?

13. A un sospechoso se le aplica un suero de la verdad del que se sabe que es fiable en un 90% de los casos cuando la persona es culpable y en un 99% cuando la persona es inocente. En otras palabras, el 10% de los culpables son considerados inocentes cuando se usa el suero y el 1% de los inocentes son considerados culpables. Si el sospechoso se escogió de un grupo del cual solo el 5% ha cometido alguna vez algún delito y el suero indica que la persona es culpable, ¿cuál es la probabilidad de que sea inocente?

14. Un fabricante de medicamentos afirma que cierto fármaco cura una enfermedad de la sangre el 80% de las veces, en promedio. Para comprobar la afirmación una oficina gubernamental utilizó el producto medicinal en una muestra de 100 individuos y decidió aceptar tal afirmación si 75 o más eran curados.

a) ¿Cuál es la probabilidad de que la afirmación sea rechazada cuando la probabilidad de cura sea, de hecho, 0.8?

b) ¿Cuál es la probabilidad de que la afirmación sea aceptada cuando la probabilidad de cura sea solo de 0.7?

15. En la asignatura de Bioestadística de cada 3 mujeres que se presentan al examen aprueban 2, y de cada 5 hombres aprueban 3. En la clase hay 4 veces más mujeres que hombres. En estas condiciones:

a) Si elegimos al azar un alumno y resulta que está aprobado, ¿cuál es la probabilidad de que sea mujer?

b) Si elegimos al azar a un alumno y resulta que ha suspendido, ¿cuál es la probabilidad de que sea un hombre?

16. Un club de estudiantes extranjeros tiene entre sus miembros a dos canadienses, tres japoneses, cinco italianos y dos alemanes. Si se selecciona al azar un grupo de cuatro personas, calcular:

a) Probabilidad de que todas las nacionalidades estén representadas.

b) Probabilidad de que todas las nacionalidades excepto la italiana estén representadas.

17. En el lejano reino de Falandia a los condenados a muerte se les concedía la gracia de que su vida dependiera de sacar bola blanca en el siguiente sorteo: se colocaban 50 bolas blancas en una urna y 50 negras en otra, entonces el condenado sacaba una bola, eligiendo la urna al azar con los ojos vendados.

En cierta ocasión, un reo pidió la gracia de que se le dejara distribuir las bolas de otro modo antes de hacer el sorteo. Tras alguna discusión de los expertos se le concedió la gracia, y colocó una bola blanca en una urna y en la otra 49 blancas y 50 negras. ¿Cuál resultó de este modo la probabilidad de salvar la vida? ¿Aumentó respecto a la que hubiera tenido sin la modificación o, por el contrario, se redujo?

18. Se lanza un dado equilibrado y el número que resulta, es el número de bolas blancas que ponemos en una urna; se vuelve a lanzar el dado, poniendo esta vez tantas bolas negras en la misma urna anterior como indique el resultado del lanzamiento. De la urna así formada, se extrae una bola al azar. ¿Cuál es la probabilidad de que sea blanca?

19. Hallar la probabilidad de obtener al menos un 4 en dos lanzamientos de un dado honrado.

20. Cada vez que un cliente compra un dentífrico elige la marca A o la marca B. Supóngase que la probabilidad de que elija la misma marca que en la compra anterior es $1/3$. Si la primera marca que adquiere es A ¿cuál es la probabilidad de que la quinta compra sea de la marca B?

21. En un sistema de alarma, la probabilidad de que ocurra una situación de peligro es 0,1; si éste se produce, la probabilidad de que la alarma funcione es 0,95; la probabilidad de que la alarme se dispare sin haber habido peligro es 0,03. Hallar:

a) Probabilidad de que habiendo funcionado la alarma, no haya habido peligro.

b) Probabilidad de que haya peligro y la alarma no funcione.

c) Probabilidad de que no habiendo funcionado la alarma, haya habido peligro.

22. Supongamos que a una piscina se ha arrojado cierta cantidad de peces de dos tipos A y B, y que tenemos un método de extracción al azar, y con remplazamiento. Sabemos que la proporción de peces tipo A es p, con $0<p<1$. Sea X = número de la extracción en que aparece por primera vez un pez tipo A. Calcular la función de probabilidad y la de distribución acumulada de X, su media y su varianza. Ídem con Y = nº de peces tipo B antes de la n-síma extracción tipo A.

23. Cuatro aspirinas y dos tabletas para purificar el agua son colocadas accidentalmente dentro del mismo bote. Una profesora de Estadística tiene una palpitante jaqueca, probablemente producida por sus alumnos. Selecciona apresuradamente dos pastillas del bote y las traga empleando un vaso de agua.

a) ¿Cuál es la probabilidad de que simplemente purifique el agua que bebe?

b) ¿Cuál es la probabilidad de que seleccione una aspirina y una tableta para purificar el agua?

24. Un jurado de tres miembros decide por mayoría. Dos de los miembros dan su veredicto según su criterio y tienen una probabilidad p de decidir correctamente. El tercer miembro se limita a lanzar una moneda y si sale cara votará culpable. Cada uno de los tres da su voto en un papel sin consultar con los demás. Otro tribunal está compuesto por un sólo juez que tiene probabilidad p de dar un veredicto correcto. ¿Cuál de los dos tribunales juzga mejor?

25. La probabilidad de un tipo de accidente industrial es 0,001. Una compañía de seguros propone a una empresa un seguro de accidentes cuyo coste anual es de 10 000 pts., comprometiéndose en caso de accidente a pagar 5 millones de pts. en concepto de indemnización. Calcular el beneficio esperado para la compañía de seguros.

26. Una compañía aérea ha observado que el 12% de las plazas reservadas no se cubren y decide aceptar reservar por un 10% más de las plazas disponibles en aviones de 450 plazas. Calcular la proporción de vuelos en que algún pasajero con reserva, no tiene plaza. (Indicar las hipótesis que hay que hacer para resolver el problema).

27. Una empresa recibe piezas de un proveedor en lotes de 2000 que se someten al siguiente control de calidad: se toman 20 al azar y si hay más de una defectuosa se rechaza el lote; en otro caso se acepta. La calidad garantizada por el proveedor es un 8 por mil de defectuosas. Calcular la probabilidad de:

a) Aceptar un lote que contenga un 2% de defectuosas.

b) Rechazar un lote que debería ser aceptado al tener sólo el 8 por mil de defectuosas.

28. Se desea capturar 14 gacelas. Para la expedición se dispone de tres vehículos que pueden transportar 500, 750 y 1000 kg de carga respectivamente. Decidir cuál sería el vehículo más idóneo para la expedición, si se desea usar el más ligero capaz de transportar las 14 gacelas. Se sabe que el peso de una gacela sigue una distribución N(50,6).

29. La probabilidad de que un niño sea varón es de 0.51. Sin considerar partos múltiples, calcular:

a) Probabilidad de que una pareja tenga tres niñas antes de tener un varón.

b) Ídem. antes de tener dos varones.

c) Número medio de hijos de una pareja para tener dos varones.

30. Por un punto de una autopista pasan coches aleatoriamente, siguiendo una distribución de Poisson, con una media de 6 coches por minuto.

a) Probabilidad de que transcurran más de 20 s desde que pasa un coche hasta que pasan 5 más.

b) Si un perro tarda 10 s en cruzar la autopista y empieza a cruzar inmediatamente después de que pase un coche, probabilidad de que lo atropellen.

31. La vida de las plantas de determinada especie sigue una distribución exponencial de parámetro $r=1/120$.

a) Calcular la función de distribución de la variable.

b) ¿Qué proporción de plantas mueren en los cien primeros días?

c) Si una cierta planta ha vivido ya cien días, ¿qué probabilidad hay de que viva al menos otros 50 días más?

32. La vida media de cierto tipo de motor pequeño es de 10 años, con una desviación estándar de 2 años. El fabricante repone gratis todos los motores que se estropeen siempre que estén en periodo de garantía. Si piensa reponer solo el 3% de los motores que fallen, ¿cuánto tiempo debe durar la garantía? (Suponemos que la vida útil de los motores sigue una distribución normal).

33. Al inspeccionar 1450 juntas de soldadura realizadas por una máquina de soldar, se encontraron 148 uniones defectuosas. Al soldar 5 uniones, ¿cuál es la probabilidad de obtener X piezas defectuosas? Calcular la media.

34. Un fabricante vende un artículo a un precio fijo de 100 pts. Si el peso del artículo es inferior a 8 g no lo puede vender y representa una pérdida total. La distribución de los pesos es una normal $N(\mu, 1)$ y el coste de producción de cada artículo viene dado por $c=5x + 30$. Determinar el peso medio μ que haga máximo el beneficio esperado.

35. El peso de los botes de conserva fabricados en cierta industria se distribuye normalmente. Se han fabricado 4000 botes en un mes, de los cuales 800 pesaron menos de 1 kg y 1000 más de 2 kg. Hallar la media y la desviación típica de la distribución.

36. En una situación de selección de personal un psicólogo utiliza un test y pone como norma aceptar a los sujetos que se aparten más de 4 pero menos de 6 puntos de la media de dicho test. Sabiendo que la distribución de las puntuaciones es normal, que la varianza es 16 y que el número de sujetos elegidos resulta ser 58. Se pide:

a) Hallar el número de sujetos que realizó el test.

b) ¿Será seleccionado el sujeto que obtuvo el percentil 67? Justificar la respuesta.

37. En una centralita se reciben un promedio de cinco llamadas entre las 9:00 y las 10:00 horas, recibiéndolas al azar. Calcular:

a) Probabilidad de que se reciban una o más llamadas entre las 9:00 y las 10:00 horas de un día determinado.

b) Probabilidad de recibir exactamente 2 llamadas entre las 9:00 h. y las 9:12 h.

c) Probabilidad de que durante una semana de 5 días, haya exactamente dos días en los que entre las 9:00 h. y las 9:12 h. no se reciba ninguna llamada.

38. En un laboratorio se dispone de una población de ratas. Se definen dos variables aleatorias X e Y de la siguiente forma:

$$X = \begin{cases} 0 & \text{si el animal no ha recibido entrenamiento.} \\ 1 & \text{si el animal ha recibido entrenamiento.} \end{cases}$$

Y = número de éxitos cuando se le hace realizar cierta tarea dos veces consecutivas.

La función de distribución conjunta viene dada por la tabla siguiente:

X\Y	0	1	2
0	0.2	0.4	0.5
1	0.25	0.75	1

a) Calcular la probabilidad de que una rata elegida al azar haya recibido entrenamiento.
b) Calcular la prob. de que una rata elegida al azar haya obtenido algún acierto al realizar la tarea dos veces consecutivas.
c) Calcular la Cov(X,Y) y estudiar la independencia de las variables.

39. Consideremos dos distribuciones normales con desviaciones típicas de 5 unidades. Una de ellas tiene media -8. Sabiendo que existe un punto A que deja a su derecha un 10% del área bajo la densidad que tiene por media -8 y a su izquierda un 15% del área bajo la densidad de la otra. Calcular la media desconocida.

40. Una licorería da servicio tanto a clientes que llegan en automóvil como a los que llegan caminando. En un día elegido al azar sean X e Y, respectivamente, las proporciones de tiempo en que se utilizan las instalaciones para atender a los automovilistas y a los que llegan caminando, y su función de densidad conjunta la siguiente:

$$f(x,y)=\begin{cases}\dfrac{2(x+2y)}{3} & \text{si} \quad 0\le x\le 1,\ 0\le y\le 1\\ 0 & \text{en otro caso}\end{cases}$$

a) Calcular la densidad marginal de X.
b) Calcular la densidad marginal de Y.
c) Calcular la probabilidad de que las instalaciones para atender a los automovilistas estén ocupadas menos de la mitad del tiempo.

41. El peso neto de una lata de conservas es aleatorio, con una distribución uniforme entre 485 y 535 gramos. Si el producto sale al mercado garantizando un peso neto de por lo menos medio kilo, ¿cuál es la probabilidad de que al inspeccionar seis latas, se detecte que más de una no cumple la garantía?

42. La variable aleatoria bidimensional (X,Y) es tal que:
a) X tiene distribución exponencial de parámetro δ.
b) Supuesto que X = x, Y sigue una distribución exponencial de parámetro x.
Calcular la distribución de Y.

43. Como parte de un experimento se decide inyectar a cada ratón de una muestra aleatoria de 25, con un fármaco cuya dosificación es de 0.004 mg/g de peso. Se sabe que el peso de esta cepa de ratones se distribuye normalmente con una media de 19 g y una desviación típica de 4 g.

a) Si el investigador consigue para sus pruebas un total de 2 mg del fármaco, ¿qué probabilidad hay de quedarse sin él?

b) ¿De qué cantidad de fármaco debería disponer para que el riesgo de quedarse sin él sea inferior al 1%?

44. Se lanzan un dado y una moneda al aire, y se consideran las variables aleatorias:

$$X = \begin{cases} 1 & \text{si sale cara} \\ -1 & \text{si sale cruz} \end{cases} \qquad Y = \begin{cases} 1 & \text{si sale impar} \\ 2 & \text{si sale par} \end{cases}$$

Hallar la función de probabilidad, media, varianza y desviación típica de las variables:

a) X+Y.

b) XY.

c) Y/X (Y condicionada a X).

45. Los accidentes de trabajo que se producen en una fábrica por semana, siguen una ley de Poisson tal que la probabilidad de que haya 5 accidentes es 16/15 de la probabilidad de que haya 2 accidentes. Calcular:

a) El parámetro de la ley de Poisson.

b) Número máximo de accidentes semanales, con seguridad del 90%.

c) Probabilidad de que no haya ningún accidente en 4 semanas.

46. Una moneda es lanzada hasta que aparece dos veces seguidas el mismo resultado. Describir el espacio muestral y hallar la probabilidad de que:

a) El experimento termine antes del sexto lanzamiento.

b) El experimento termine en un lanzamiento par.

47. Para obtener el permiso de conducir se realiza un test con 20 ítems. Se sabe que una determinada persona tiene una probabilidad 0.8 de contestar bien cada ítem. Calcular:

a) Probabilidad de que la primera pregunta que contesta bien sea la tercera que hace.

b) Probabilidad de que apruebe al contestar el doceavo ítem, sabiendo que para aprobar el test es necesario contestar 10 ítems bien.

48. Un abogado se traslada diariamente de su casa en las afueras a su oficina en el centro de la ciudad. En promedio el viaje dura 24 minutos con una desviación estándar de 3.8 minutos y suponemos que los tiempos de traslado siguen una distribución normal.

a) Probabilidad de que cierto día tarde al menos media hora en llegar a la oficina.

b) Si la oficina abre a las 9:00 a.m. y él sale de su casa a las 8:45 a.m. diariamente, ¿qué porcentaje de días llega tarde?

c) Si sale de su casa a las 8:35 a.m. y en la oficina se sirve un café entre las 8:45 y las 9:00, ¿cuál es la probabilidad de que se pierda el café?

d) Encontrar el tiempo por encima del cual duran el 15% de los traslados más lentos.

e) Encontrar la probabilidad de que 2 de los 3 siguientes traslados duren menos de media hora.

49. Una moneda se lanza dos veces. Sea Z el número de caras en el primer lanzamiento y W el número total de caras en los dos lanzamientos. La moneda no está equilibrada, siendo la probabilidad de cara 0.4, encontrar:

a) La distribución de probabilidad conjunta de W y Z.

b) La distribución marginal de W.

c) La distribución marginal de Z.

d) La probabilidad de que ocurra al menos una cara.

50. La longitud de una cierta pieza se distribuye en el intervalo [1,3] con función de densidad $f(x) = k\,(x-1)\,(3-x)$ con k constante y $f(x) = 0$ en el resto. Sólo es válida una pieza si su longitud está comprendida entre 1.7 y 2.4.

a) ¿Cuál es la probabilidad de que una determinada pieza sea útil?

b) Si las piezas se empaquetan en lotes de 5 unidades, y se acepta el lote si contiene menos de dos piezas defectuosas, ¿cuál es la probabilidad de que un cierto lote sea rechazado?

51. Se informó a dos estudiantes que habían recibido puntuaciones estándar de 0.8 y -0.4, respectivamente, en una prueba de inglés. Si sus puntuaciones fueron 88 y 64, respectivamente, hallar la media y la desviación típica de las puntuaciones de esa prueba, sabiendo que dichas puntuaciones siguen una distribución normal.

52. Cinco personas lanzan simultáneamente monedas, para determinar quién ha de comprar los refrescos para todos. El sistema es el siguiente: el primero que obtenga un resultado (ya sea cara o cruz) distinto de cada uno de los resultados obtenidos por los demás, debe pagar los refrescos de todos. Sea X el número de ensayos requeridos para concluir el juego. Se pide:

a) Determinar su función de probabilidad.

b) Determinar su función de distribución.

c) Calcular $P(X>2)$, $P(X<2)$ y $P(3\leq X\leq 6)$.

d) Calcular $P(X>5 / X>2)$.

e) Calcular $E[X]$ y $V(X)$.

53. En un cierto hospital se comprobó que el peso en kilos de los niños al nacer era una variable aleatoria cuya función de densidad es:

$$f(x)=\begin{cases} kx & si\ 2\leq x\leq 4 \\ 0 & en\ otro\ caso \end{cases}$$

Se pide:

a) Hallar k para que f(x) sea función de densidad.

b) Hallar la función de distribución, media y varianza.

c) Probabilidad de que un niño elegido al azar pese más de 3 kilos.

d) Probabilidad de que pese entre 2 y 3.5 kilos.

e) ¿Cuánto debe pesar un niño para tener inferior o igual a su peso al 90% de los niños?

54. Los empresarios de una fábrica de automóviles deciden otorgar un premio entre los distribuidores si venden 320 automóviles o más por día. El número de automóviles vendidos al día por los distribuidores A y B está normalmente distribuido de la forma siguiente:

Distribuidor	Media	Desviación Típica
A	290 automóviles	20 automóviles
B	300 automóviles	10 automóviles

Se pide:

a) ¿Qué porcentaje de los días obtendrá premio el distribuidor A?

b) ¿Qué porcentaje de los días obtendrá premio el distribuidor B?

c) ¿A quién de los dos distribuidores beneficia la decisión de la empresa?

d) Si se asociaran los dos distribuidores A y B, ¿qué porcentaje de los días obtendrían premio?

55. En el niquelado de ciertas láminas metálicas se producen desperfectos que se distribuyen aleatoriamente sobre toda la superficie niquelada. El número de defectos por lámina es una v. a. de Poisson de media 2.

a) Si se eligen al azar 10 láminas de la cadena de montaje, ¿cuál es la probabilidad de que exactamente 3 tengan defectos?

b) Se toman láminas de la cadena de montaje hasta obtener 7 sin defectos, ¿cuál es la probabilidad de que haya habido que elegir 10 láminas?

c) ¿Qué porcentaje de láminas tendrá 4 defectos o más?

56. El tiempo, en horas, empleado diariamente en transporte por los trabajadores de cierta ciudad, sigue una distribución Gamma con parámetros a=2, p=2.

a) Calcular la probabilidad de que un trabajador emplee más de media hora en transporte.

b) Calcular el tiempo mínimo que emplea el 50% de los trabajadores que más tiempo pierden en transporte.

57. La empresa Olis, encargada de la fabricación y el mantenimiento de ascensores, desea mejorar la calidad de sus productos y el servicio al cliente. A tal fin, se lleva a cabo un análisis de la situación actual del que se concluye que la distribución conjunta de la variable aleatoria X, tiempo transcurrido, en días, desde la instalación de un ascensor hasta la primera avería, e Y, tiempo en días que tarda la empresa en acudir al lugar del problema una vez se ha dado el aviso, viene dada por:

$$f(x,y)=\begin{cases}\dfrac{1}{100}e^{-\frac{x+900y}{300}} & si\ x>0,\ y>0\\ 0 & en\ otro\ caso\end{cases}$$

a) Calcular la función de densidad del tiempo de funcionamiento hasta la primera avería.

b) El jefe de escalera de un vecindario está indignado porque hace 3 días que dio parte a la empresa de una avería y todavía no han acudido a arreglarla. ¿Cuál es la probabilidad que se dé una situación como ésta?

c) ¿Son X e Y independientes?

58. Se estudia un proceso de fabricación que tiene unos límites de tolerancia superior e inferior, y se sabe que la probabilidad de rechazo correspondiente a estos límites es, respectivamente:

$$p_1 = 0{,}06 \qquad p_2 = 0{,}02$$

En una muestra de 100 elementos cierto número de ellos serán rechazados por exceder las especificaciones y otros por no alcanzarlas. La función de probabilidad conjunta del número de unidades rechazadas es:

$$P(x_1, x_2, x_3) = \frac{100}{x_1!\,x_2!\,x_3!}(0{,}06)^{x_1}(0{,}02)^{x_2}(0{,}92)^{x_3}$$

a) ¿Cual es la probabilidad de que sean rechazadas 2 piezas por defecto y 5 por exceso de tolerancia?

b) ¿Cuál es la probabilidad de que sean rechazadas cuatro piezas?

c) ¿Cual es la probabilidad de que sean rechazadas dos piezas?

59. El tiempo que se tarda en reparar una máquina tiene la función de distribución siguiente:

$$F(x) = \begin{cases} 0 & si\, x \leq 0 \\ x/2 & si\; 0 < x \leq 1 \\ 1/2 & si\; 1 < x \leq 2 \\ x/4 & si\; 2 < x \leq 4 \\ 1 & si\, x > 4 \end{cases}$$

a) Dibujar la función de distribución.

b) Obtener la función de densidad e interpretarla.

c) Si el tiempo de reparación de cierta máquina es superior a una hora, ¿cuál es la probabilidad de que sea superior a tres horas y media?

60. Una máquina de empaquetado automático deposita por término medio en cada paquete 81,5 g de producto, con desviación típica de 8 g. El peso medio del paquete vacío es de 14,5 g, con desviación típica de 6 g. Ambas variables son normales e independientes. Se pide:

a) Calcular la distribución del peso de los paquetes llenos.

b) Escribir la función de densidad conjunta de la variable:

(X , Y) = (peso producto, peso envase)

61. En dos grupos de 2º año de carrera de una Facultad se ha medido el cociente de inteligencia de los alumnos. En el grupo A la media fue 100 y la desviación típica 10, y en el grupo B fueron 105 y 12, respectivamente. Ambos grupos tienen el mismo número de alumnos, y escogido un alumno al azar se comprueba que su cociente de inteligencia es superior a 120. Suponiendo normalidad, ¿cuál es la probabilidad de que el citado alumno provenga del grupo B?

62. En un laboratorio el 40% de los ratones sometidos a un cierto estímulo son machos. El tiempo, en minutos, que un ratón tarda en responder al estímulo es una variable aleatoria exponencial, siendo el tiempo medio de respuesta el doble en los machos que en las hembras.

Se toma un ratón al azar y se le somete al experimento. Si la probabilidad de que tarde más de 4 minutos en responder es 0,8, obténgase el tiempo medio de respuesta para los machos y para las hembras.

63. Supongamos que la probabilidad de seleccionar una palabra de la oración LA NIÑA SE PUSO SU PRECIOSO SOMBRERO ROJO es proporcional al número de letras que forman dicha palabra. Si X es el número de letras de la palabra seleccionada, decir cuáles son: el experimento; el espacio muestral; la función de probabilidad asociada; la variable aleatoria, diciendo de qué tipo es; el recorrido de la variable; la función de probabilidad de la variable; y cuál es el valor de E[X] y V[X].

64. El número de coches que diariamente pasan por cierta aldea se distribuye según una ley de Poisson de intensidad $\lambda=20$. La probabilidad de que uno de estos coches se detenga en la taberna de la aldea es $p=0,2$. Estudiar la distribución del número de coches que paran en la taberna en un día determinado y calcular la probabilidad de que en un día determinando se detengan 5 coches.

65. Al efectuar la suma de 300 números reales se aproxima cada uno de los sumandos por el entero más próximo. Así si uno de los sumandos fuera el número 13,845..., para efectuar la suma este sumando se tomaría como 14.

Suponiendo que la distribución de decimales es uniforme en el intervalo [-0,5 , 0,5] se pide calcular la probabilidad de que la diferencia entre la suma de los números reales S y la suma S* que resulta de los redondeos no sea superior a cuatro unidades.

66. Se está estudiando si interesa fabricar una máquina automática nueva para soldar. Se considerará exitosa si tiene una efectividad del 99% en sus soldaduras, y si no, no se fabricará. Se lleva a cabo una prueba con un prototipo y se realizan 100 soldaduras. La máquina se aceptará para su fabricación si no son defectuosas más de tres soldaduras.

a) ¿Cuál es la probabilidad de que una máquina eficiente sea rechazada?
b) ¿Cuál es la probabilidad de que una máquina ineficiente sea rechazada? (Supongamos que el verdadero porcentaje de soldaduras buenas fuera el 95%).

67. En un campeonato de ajedrez entre dos jugadores, las tablas no se tienen en cuenta y el juego continúa hasta que uno de los contrincantes alcanza seis puntos (ganar = 1 punto, perder = 0 puntos). Considerando a los participantes igualmente expertos, hallar la probabilidad de que, con tales reglas, en el momento de finalizar la confrontación, el perdedor tenga k puntos, k = 0, 1, 2, 3, 4, 5.

68. Un bolso contiene tres monedas, dos de las cuales son normales y equilibradas y la otra tiene dos caras. Se escoge una moneda al azar y se tira cuatro veces. Si las cuatro veces sale cara, ¿qué probabilidad hay de que se haya escogido la moneda con dos caras?

69. La duración de cierto tubo de radio es una variable aleatoria continua con función de densidad:

$$f(x)=\begin{cases}\dfrac{100}{x^2} & \text{si } x>100\\ 0 & \text{en el resto}\end{cases}$$

a) ¿Cuál es la probabilidad de que un tubo dure menos de 200 horas si se sabe que todavía funciona después de 150 horas?
b) ¿Cuál es la probabilidad de que si se instalan 3 de tales tubos, exactamente uno deba ser sustituido después de 150 horas?
c) ¿Cuál es el número máximo de tubos que se debe poner en un sistema para que haya una probabilidad de al menos 0,5 de que tras 150 horas de servicio todos funcionen?

70. Tenemos dos tipos de urnas: U y V. Cada una contiene bolas blancas y negras. En cada una de las urnas tipo U hay 20 bolas blancas y 80 negras. En cada una de las urnas tipo V hay 40 bolas blancas y 60 negras. Calcular en qué proporción se encuentran ambos tipos de urnas, sabiendo que la probabilidad de que una bola blanca proceda de una urna tipo U es 0,2.

71. La probabilidad de que la media aritmética de 64 variables aleatorias independientes, igualmente distribuidas, con media α y varianza 16α, esté comprendida entre $-\alpha$ y 3α es 0,6. Calcular α.

72. Se observa una fuente radioactiva durante 5 intervalos de tiempo no solapados de 8 segundos de duración cada uno. Las partículas son emitidas según una ley de Poisson con una media de 0,5 partículas por segundo y registradas en un contador. Hallar la probabilidad de que:

a) En cada uno de los cinco intervalos se cuenten 4 o más partículas.

b) Se cuenten 4 o más partículas al menos en uno de los intervalos.

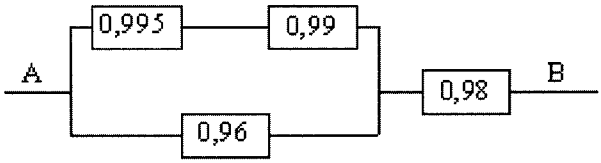
73. Un empleado de un banco sustituye cautelosamente un billete bueno por uno falso en cada fajo de cien billetes. Si el interventor del banco toma un billete al azar de cada uno de 50 fajos, ¿cuál es la probabilidad de que tope con algún billete falso?

74. En Alemania, durante el siglo XVI, la unidad de longitud se determinaba así: Un domingo, los primeros dieciséis hombres que salían de la iglesia eran puestos en fila. Se tomaban a continuación las medidas de sus pies izquierdos, se sumaban y se dividía entre dieciséis. El resultado era el *pie correcto y legal* a utilizar en las transacciones.

Se sabe que la longitud (en mm) del pie de un hombre adulto de la época era aleatoria con distribución N(262,5 , 12).

Encontrar la probabilidad de que dos valores del *pie correcto y legal*, determinados en dos diferentes iglesias, difieran entre sí en más de 5 mm.

75. Consideremos una red de comunicaciones como la de la figura. Dos puntos están comunicados si hay alguna línea que los una, todos cuyos repetidores (representados en el diagrama mediante rectángulos) funcionan. Sobre cada repetidor se indica la probabilidad de que esté en servicio. ¿Cuál es la probabilidad de que A y B queden incomunicados?



76. Cinco personas juegan a los chinos. ¿Cuál es la probabilidad de que el total de monedas escondidas sea 7? (El juego de los chinos consiste en adivinar el número total de monedas que un grupo de jugadores tienen escondidas, sabiendo que cada jugador puede tener 0, 1, 2 ó 3 monedas, a voluntad).

77. El número de hamburguesas que se venden semanalmente en un *burguer*, en cientos de unidades, es una v. a. con función de densidad:

$$f(x)=\begin{cases}\dfrac{x^2}{9} & \text{si } 0<x<3\\ 0 & \text{en el resto}\end{cases}$$

Si en una semana se venden x hamburguesas, la cantidad de ellas que llevan queso, Y, se distribuye según la siguiente función de densidad:

$$f(y/X=x)=\begin{cases}\dfrac{1}{x} & \text{si } 0<y<x<3\\ 0 & \text{en el resto}\end{cases}$$

a) Sabiendo que cierta semana se han vendido cien hamburguesas, ¿cuál es la probabilidad de que por lo menos la mitad de ellas tuvieran queso?
b) Calcular la probabilidad de que en una semana el número de hamburguesas sin queso sea superior a cien?

78. La demanda diaria, en unidades de cierto producto durante 30 días laborables fue:

38	35	76	58	48	59
67	63	33	69	53	51
30	21	32	32	59	62

49	78	48	42	72	52
47	66	52	44	44	56

a) Agrupar en cinco intervalos y construir las tablas de frecuencias agrupadas. Dibujar el histograma de frecuencias y el polígono de frecuencias acumuladas.

b) Calcular la mediana y la desviación mediana para los datos sin agrupar y agrupados.

79. La tasa de suicidios de un país es de uno al mes por cada millón de habitantes.

Se pide:

a) Calcular el número medio de meses sin suicidios al año.

b) Calcular, en una ciudad de dos millones y medio de habitantes, la probabilidad de que en un mes haya más de cuatro suicidios.

80. Supongamos que los diámetros de los pernos de una caja grande siguen una distribución normal con una media de 2 cm y desviación típica de 0,03 cm. Además supongamos que los diámetros de los agujeros de las tuercas de otra caja grande siguen una distribución normal de media 2,02 cm. Y desviación típica de 0,04 cm. Un perno y una tuerca ajustarán si el diámetro del agujero de la tuerca es mayor que el diámetro del perno y la diferencia entre ambos diámetros no es mayor que 0,05 cm. Si se seleccionan al azar un perno y una tuerca, ¿cuál es la probabilidad de que ajusten?

81. La llegada del correo a una empresa se produce con igual probabilidad en cualquier instante entre las 10 y las 11 horas. El conserje lo entrega a la secretaria del director, también con igual probabilidad en cualquier instante desde el momento que llega pero nunca más tarde de las 11. ¿A qué hora, en promedio, recibirá la secretaria el correo? ¿Cuál es la probabilidad de recibirlo antes de las 10 y media?

82. Una persona tira dos dados, uno de ellos es un cubo y el otro un tetraedro regular, tomando el número de la cara inferior cuando se trata del tetraedro, ¿cuál es la probabilidad de que la suma de los números obtenidos no sea menor que 5?

83. En función de su prioridad un programa para ordenador espera en la fila de entrada cierto tiempo, después del cual lo ejecuta el procesador central en un determinado tiempo. La función de densidad conjunta para los tiempos de espera y de ejecución viene dada por:

$$f(t_1,t_2)=\begin{cases} 2\ e^{-(\frac{t_1}{5}+10t_2)} & \text{si } t_1,\ t_2>0 \\ 0 & \text{en el resto} \end{cases}$$

Se pide:

a) Obtener la probabilidad conjunta de que el tiempo de espera no sea mayor de 8 minutos y el de ejecución no sea mayor de 12 segundos.

b) Obtener las funciones de densidad marginales y analizar la independencia de ambas variables.

c) Determinar el tiempo esperado de ejecución de un programa, si esperó en la fila de entrada un tiempo t_1^*.

84. Supongamos que cuando cierta máquina está correctamente ajustada, el 50% de los artículos que produce son de alta calidad y el otro 50% son de calidad media. Supongamos, sin embargo, que la máquina está mal ajustada durante el 10% del tiempo y que, en estas condiciones, el 25% de los artículos producidos por ella son de alta calidad y el 75 % de los artículos son de calidad media.

a) Supongamos que 5 artículos producidos por la máquina en un tiempo determinado son seleccionados al azar e inspeccionados. Si cuatro de ellos son de alta calidad y uno de calidad media, ¿cuál es la probabilidad de que la máquina estuviera correctamente ajustada durante ese tiempo?

b) Supongamos que se selecciona un artículo adicional , que fue producido por la máquina al mismo tiempo que los otros cinco, y resulta ser de calidad media. ¿Cuál es la nueva probabilidad final de que la máquina estuviera correctamente ajustada?

85. Una persona apuesta a que tirando tres veces dos dados, obtendrá en una de ellas un total de 5 puntos y en otra un total de 7 puntos. ¿Qué probabilidad tiene de ganar la apuesta?

86. Supongamos que la distribución del número de defectos en un determinado rollo de tela sigue una distribución de Poisson con media 5 y que para una muestra aleatoria de 125 rollos se cuenta el número de defectos de cada rollo. Calcular la probabilidad de que el número promedio de defectos por rollo en la muestra sea menor que 5,5.

87. Sea X una variable aleatoria con media μ y desviación típica σ.

Calcular $E\left[(X-c)^2\right]$ en función de μ y σ, siendo c una constante. ¿Qué valor de c hace *mínima* esa expresión?

88. En un campeonato de tenis usted tiene la opción de escoger la secuencia de partidos A-B-A o la B-A-B donde A y B indican sus oponentes. Para clasificarse debe usted ganar dos partidos consecutivos. El jugador A es mejor que el B. ¿Qué secuencia elegiría usted?

89. Se admite que las retribuciones percibidas en una empresa siguen una distribución normal. Por las relaciones de los seguros sociales, se sabe que el 1% son superiores a 5 800 000 pts. y el 10% son inferiores a 1 200 000 pts. Calcular qué proporción de las retribuciones son superiores a 3 000 000 pts.

90. En un aparato de control actúan dos variables, X e Y independientes, ambas con una distribución uniforme, la primera entre 1 y 9, la segunda entre 1 y *a*. El aparato funciona bien cuando $x<4y^2$. Calcular el valor de *a* para que:

$$P(X > 4Y^2) \leq 0{,}01.$$

91. Una moneda cuyas caras están marcadas con un 2 y un 3, respectivamente, se tira 5 veces. La probabilidad de salir la cara que tiene el 2 es 0,6. ¿Cuál es la probabilidad de que la suma de los puntos obtenidos en las 5 tiradas sea 12?

92. Un portamonedas contiene cuatro monedas que sólo pueden ser pesetas o duros. Se sacan dos monedas y resultan ser dos pesetas. Si las monedas se vuelven a guardar en el portamonedas, y a continuación se extrae de nuevo una moneda, ¿cuál es la probabilidad de que sea un duro?

93. Una empresa fabrica ascensores con capacidad para 16 personas. Cuando el peso de las personas que pretenden utilizar el ascensor excede del que puede elevar el motor, suena una alarma ordenando el desalojo. Si el peso de una persona (en kg) se distribuye como una normal de media 74 y varianza de 144, ¿cuál debe ser la potencia del motor para que la alarma suene no más del 1% de las veces en que 16 personas ocupan el ascensor?

94. Sea X una variable exponencial.

a) ¿Cuál es la probabilidad de que X tome un valor mayor que su media?

b) ¿Cuáles son las probabilidades de que X tome un valor que se encuentre como máximo a una distancia de una desviación estándar de la media, primero, y a dos desviaciones de la media, después?

95. Sean X e Y dos variables aleatorias continuas con función de densidad conjunta:

$$f(x\ ,\ y)=\begin{cases} x\ \ e^{-x(y+1)} & \text{si}\ \ x,\ \ y>0 \\ \quad 0 & \text{en otro}\ \ \text{caso} \end{cases}$$

a) Demostrar que f(x , y) es función de densidad.

b) Calcular P(X<2,Y<1).

c) Obtener las funciones de densidad marginales de X y de Y.

d) ¿Son X e Y estadísticamente independientes?

96. Un fabricante de escapes para automóviles desea garantizar su producto durante un tiempo igual a la vida del vehículo. El fabricante supone que el tiempo de duración de su producto es una variable aleatoria con una distribución normal, con una vida promedio de tres años y una desviación estándar de seis meses. Si el costo de reemplazo de una unidad es de 10 $, ¿cuál puede ser el costo de reemplazo para dos años si se instalan 1 000 000 de unidades?

97. Cierta enfermedad puede ser producida por tres tipos de virus, A, B y C. En un laboratorio hay tres tubos con el virus A, dos con el virus B y cinco con el virus C. La probabilidad de que el virus A produzca la enfermedad citada es

$P\{|X| < 4\}$ siendo $X \in N(3{,}5)$. La probabilidad de que el virus B produzca la enfermedad es $P\{Y \leq 3\}$, siendo $Y \in B(5\ ,\ 0{,}7)$. Por último la probabilidad de que el virus C produzca la enfermedad es $P\{W \leq 5\}$ siendo $W \in P(4)$ (variable de Poisson).

Se elige un tubo al azar y se inocula el virus a un animal, éste contrae la enfermedad; probabilidad de que el virus inoculado sea del tipo C.

98. La calificación promedio en una prueba de Estadística fue de 62,5 con una desviación estándar de 10. El profesor sospecha que el examen fue difícil. De acuerdo con lo anterior desea ajustar las calificaciones de manera que el promedio sea de 70 y la desviación estándar de 8. ¿Qué ajuste de tipo $aX+b$ debe hacer?

99. Se lanza una moneda, si sale cara se echan *a* bolas blancas en una urna y si sale cruz son *2a* bolas blancas las que se ponen en la urna. El mismo procedimiento se sigue con una segunda tirada, poniendo *b* bolas negras si sale cara y *2b* bolas negras si sale cruz. De la urna así compuesta se extrae una bola al azar. ¿Cuál es la probabilidad de que sea negra?

100. El peso de cereal que contiene una caja se aproxima a una distribución normal con una media de 600 g. El proceso de llenado de las cajas está diseñado para que, de entre 100 cajas, el peso de una se encuentre fuera del intervalo 590-610 gramos. ¿Cuál es el valor máximo de la desviación estándar para alcanzar este requerimiento?

101. Los jurados para una corte federal de distrito se seleccionan de manera aleatoria entre la lista de votantes del distrito. En un determinado mes se selecciona una lista de 25 candidatos. Ésta contiene los nombres de 20 hombres y cinco mujeres.

a) Si la lista de votantes se encuentra igualmente dividida por sexo, ¿cuál es la probabilidad de obtener una lista de 20 hombres y cinco mujeres?

b) ¿Cuál es la probabilidad de elegir de entre esa lista un jurado de 12 personas, de las cuales sólo una sea mujer?

c) Si el alumno fuera el abogado de la defensa, ¿qué podría argumentar mediante el empleo de las respuestas de las partes a) y b)?

102. Se tiene una moneda con una probabilidad de 2/3 de que al tirarla el resultado sea cara. Se lanza la moneda: si el resultado es cara, se saca una bola aleatoriamente de una urna que contiene dos bolas rojas y tres verdes; si el resultado es cruz se saca una bola de otra urna que contiene dos bolas rojas y dos verdes. ¿Cuál es la probabilidad de sacar una pelota roja?

103. ¿Coinciden la media y la mediana en una variable que presenta una distribución exponencial?

104. Considérense las dos apuestas siguientes:

a) Se tira 50 veces una moneda equilibrada, usted gana si el número de caras oscila entre 20 y 30; pierde en caso contrario.

b) Se tira 5000 veces la misma moneda, usted gana si el número de caras oscila entre 2000 y 3000; pierde en caso contrario.

Respóndase razonada y detalladamente qué apuesta prefiere y por qué.

105. En una oficina de empleo se aplica un test a los aspirantes a cierto puesto de trabajo. De los 250 presentados, 210 obtienen calificación inferior a 5,5. Se sabe que la distribución de las calificaciones es normal con desviación típica 1. Si hay 150 plazas, ¿cuál es la mínima puntuación que tiene que alcanzar un candidato para obtener una plaza?

106. Consideremos una situación ideal de la Bolsa. Al empezar un año parte de 100 enteros y la probabilidad de que suba un entero en un día es de un tercio y de que baje un entero de dos tercios. Supongamos que las alzas y bajas son independientes de un día a otro. Sea X la variable que da la posición de la Bolsa después de cuatro sesiones consecutivas. Se pide:

a) Hallar la distribución de X.

b) Calcular su esperanza.

107. Sea una urna con 1 bola blanca, 2 rojas y 3 negras. Se extraen dos bolas sin remplazamiento. Denotamos X_1, X_2, X_3, las variables que representan el número de bolas blancas, rojas y negras, respectivamente:

a) Calcular las distribuciones marginales de las tres variables.

b) Calcular la esperanza y la desviación típica de cada una de las variables.

108. En una mesa de juego, en 1654, Meré propuso a Pascal la siguiente afirmación: "Es más probable obtener al menos un as en una sola tirada con cuatro dados, que al menos un doble as en veinticuatro tiradas con dos dados". Demostrar que Meré tenía razón.

109. Supongamos que la distribución de estaturas, en cm, de los jóvenes en edad militar en cierto país es una v. a. con distribución $N(175,100)$ y que son rechazados aquellos cuya estatura es inferior a 155 cm o superior a 195 cm. ¿Cuál es la probabilidad de que un soldado en la mili mida entre 172 y 181 cm?

110. Un individuo tiene un aparato que necesita para funcionar una válvula C cuya duración en horas viene dada por la función de densidad:

$$f(x) = \frac{1}{10} e^{-\frac{x}{10}}, \quad x \geq 0$$

Tiene dos válvulas de ese tipo y quiere saber cuál es la probabilidad de que el aparato funcione más de 21 horas, suponiendo que los demás componentes del sistema tienen garantizada una duración de al menos 100 horas y que si se estropea una válvula empieza a funcionar la otra automáticamente.

111. Un profesor realiza un test con diez ítems a una clase, teniendo cada ítem cuatro posibles respuestas A, B, C y D, de las cuales sólo una es correcta. Suponiendo que un alumno no se ha preparado el test y tiene la misma probabilidad de responder A, B, C o D, hallar la probabilidad de los siguientes sucesos:

a) Que conteste todos los ítems mal.

b) Que conteste al menos 5 ítems bien.

c) Que el séptimo ítem sea el tercero que contesta bien.

d) ¿Cuántos ítems se espera que conteste bien?

112. Un país está habitado por dos grupos étnicos: A (alpinos) y N (nórdicos), que se encuentran en las proporciones 0,75 y 0,25. Se sabe que la talla de los individuos adultos varones es N (170,7), en cm para el grupo A y N (176,7) para el grupo N. Se dice que un individuo es "alto" si su talla es superior a 180 cm. Calcular:

a) La proporción de altos de los grupos A y N.

b) Probabilidad de que un individuo sea nórdico, si se sabe que es alto.

c) Probabilidad de que de seis individuos elegidos al azar, 4 de ellos sean "altos".

113. Sea (X,Y) una v. a. bidimensional con función de densidad conjunta:

$$f(x,\ y)=\begin{cases} e^{-y} & \text{si} \quad 0<x<y<\infty \\ 0 & \text{en otro caso} \end{cases}$$

Describir el recinto donde está definida la variable bidimensional. Calcular la covarianza de las dos variables y decir si son independientes.

114. Supóngase que el 75% de las personas de cierta área metropolitana viven en la ciudad y el 25% restante en las afueras. Si las 1200 personas que asisten a un concierto son consideradas una muestra aleatoria, ¿cuál es la probabilidad de que el número de personas que asisten al concierto y que viven en las afueras sea menor de 270?

115. Una urna contiene 5 bolas cuyo color se desconoce. Se extrae una bola y se devuelve a la urna. La bola extraída es de color rojo. A continuación se sacan simultáneamente dos bolas. ¿Cuál es la probabilidad de que las dos bolas sean también rojas?

116. Uno de cada cuatro habitantes de una región ha sido vacunado contra cierta enfermedad. En el curso de una epidemia se observa que entre los enfermos hay uno vacunado por cuatro no vacunados.

a) ¿Es la vacuna de alguna eficacia?

b) Se sabe, además, que hay un enfermo de entre doce vacunados. ¿Cuál es la probabilidad de caer enfermo de un individuo no vacunado?

117. Un test detecta la presencia de un cierto tipo T de bacterias en el agua con probabilidad 0,9 en el caso de haberlas. Si no las hay, detecta la ausencia con probabilidad 0,8. Sabiendo que la probabilidad de que una muestra de agua contenga bacterias tipo T es 0,2, calcular la probabilidad:

a) De que haya realmente bacterias T, cuando el test ha dado positivo.

b) Ídem cuando el resultado del test es negativo.

c) De que haya bacterias y el test sea positivo.

d) De que no haya bacterias y el test dé negativo.

118. Si se seleccionan 3 dígitos de una tabla de dígitos aleatorios. ¿Cuál es la probabilidad de que su suma sea 6?

Enunciados 2.

Inferencia estadística

1. Un metalúrgico ha realizado cuatro determinaciones del punto de fusión del manganeso: 1269°, 1265°, 1271° y 1263°. Según un manual el punto de fusión es 1260°. ¿Están de acuerdo los datos con el manual o debe aceptar que el punto de fusión es mayor?

2. Las compañías de auditorías generalmente seleccionan una muestra aleatoria de los clientes de un banco y verifican los balances contables reportados por el banco. Si una compañía de este tipo se encuentra interesada en estimar la proporción de cuentas para las cuales existe una discrepancia entre el cliente y el banco, ¿cuántas cuentas deberán seleccionarse de manera tal que con una confiabilidad del 0,995, la proporción muestral se encuentre a no más de 0.02 unidades de la proporción real?

3. Al terminar la Segunda Guerra Mundial se intentó dar razones científicas para decidir si el bombardeo de Londres por los proyectiles V2 había sido al azar o se había hecho contra lugares de interés militar. Con tal fin se cuadriculó el plano de Londres en 576 casillas y se contó el número de casillas en que habían caído 0, 1, 2, etc., bombas V2, obteniéndose:

n° bombas V2	0	1	2	3	4
n° de casillas	229	211	93	35	8

A la vista de estos datos, decidir si las bombas obedecían a un plan de ataque a objetivos de interés militar o no.

4. Un agricultor quiere saber si hay diferencia entre las producciones de dos variedades de trigo A y B. La tabla indica la producción por unidad de área de ambas variedades. ¿Puede concluirse que existe diferencia a los niveles de significación a) 0.01; b) 0.05?

trigo A	15.9	15.3	16.4	14.9	15.3	16.0	14.6	15.3	14.5	16.6	16.0
trigo B	16.4	16.8	17.1	16.9	18.0	15.6	18.1	17.2	15.4		

5. La tabla adjunta muestra datos referentes a horas trabajadas en un taller y número de unidades producidas:

horas	80	79	83	84	78	60	82	85	79	84	80	62
unidades	300	302	315	330	300	250	300	340	315	330	310	240

a) Estudiar la recta de regresión y decir si el modelo lineal es suficientemente bueno para relacionar las variables.

b) ¿Cuántas unidades se espera producir con 75 horas de trabajo?

c) ¿Se puede admitir, a nivel de significación 0.05, que la pendiente de la recta de regresión es mayor que 2.5?

6. Un técnico opina que introduciendo un nuevo tipo de maquinaria en cierto proceso de producción se disminuye sustancialmente el tiempo requerido en la misma. A causa del alto costo de mantenimiento, el empresario dice que salvo que se reduzca el tiempo de producción en más de 8 minutos, no vale la pena la inversión. Seis experiencias arrojan una disminución media de dicho tiempo de 8.4 minutos con desviación típica de 0.32. Contrastar la hipótesis de que el proceso merece ser renovado, con niveles de significación a) 0.01 y b) 0.05.

7. La tabla adjunta recoge una muestra de 40 notas de un examen de ámbito nacional. Contrastar, a un nivel de significación de 0.05, la hipótesis de que la nota mediana de todos los participantes es a) 66, b) 75.

71	67	55	64	82	66	74	58	79	61
78	46	84	93	72	54	78	86	48	52
67	95	70	43	70	73	57	64	60	83
73	40	78	70	64	86	76	62	95	66

8. En una investigación se midió el largo de las alas y de la lengua de 16 abejas, obteniéndose una correlación muestral r = 0.415. ¿Es significativa esta correlación, a nivel α = 0.01?

9. En el proceso de fabricación de un tipo de piezas se ha llegado a la conclusión de que el porcentaje de defectuosas es 2 ó 4. Se quiere contrastar la hipótesis de que el porcentaje es 2 tomando una muestra de tamaño 100, se rechaza la hipótesis si en la muestra hay más de tres defectuosas. ¿Qué valores tienen las probabilidades de error tipo I y tipo II?

10. Se espera que dos trabajadores produzcan en promedio, el mismo número de unidades terminadas en el mismo tiempo. El número de unidades terminadas por cada operario en cierta semana fue:

operario 1	12	11	18	16	13
operario 2	14	18	18	17	16

Si se supone que el número de unidades terminadas diariamente por cada trabajador es una variable aleatoria normal y que ambas tienen la misma varianza, ¿hay alguna diferencia entre los trabajadores a nivel de significación α =0.1?

11. Para estudiar las diferencias sobre la producción de remolacha de cuatro fertilizantes, se dispuso de cinco fincas, cada una de las cuales se dividió en cuatro parcelas del mismo tamaño y tipo. Los fertilizantes fueron asignados al azar en las parcelas de cada finca y el rendimiento en toneladas fue:

		FERTILIZANTE			
		A	B	C	D
FINCA	1	2.1	2.2	1.8	2.1
	2	2.2	2.6	1.9	2.0
	3	1.8	2.7	2.6	2.2
	4	2.0	2.5	2.0	2.4
	5	1.9	2.8	1.9	2.1

Se desea saber si hay diferencias entre los fertilizantes y entre las fincas prescindiendo del tipo de terreno.

12. Se ha dividido un terreno en 100 parcelas de 1 m^2 y se ha contado el número de semillas de cierta especie que han germinado espontáneamente en cada parcela, con los resultados siguientes:

nº de semillas germinadas	0	1	2	3	4	5
frecuencia	9	31	30	16	10	4

¿Puede admitirse que el número de semillas germinadas por m^2 sigue una ley de Poisson? (Con nivel de significación α =0.05).

13. Una compañía de helados desea saber si debe enfocar la publicidad de un nuevo sabor sólo a un grupo determinado de edades o si la debe dirigir a todas las edades. Para hacer un estudio se establecen tres grupos de edades y se seleccionan aleatoriamente 20 personas de cada grupo. A cada uno se le da a probar el nuevo helado y se le pide que lo clasifique como 1) excelente, 2) normal o 3) malo.

a) ¿Qué procedimiento se debe usar? Describirlo con detalle.

b) ¿Cuál es la región crítica para a = 0.05?

c) Si el estadístico del contraste toma el valor 0.5, ¿cuál es la conclusión y por qué?

d) ¿Tenemos alguna información sobre si el nuevo helado se venderá bien?

14. En una fábrica de componentes electrónicos se sabe que la proporción usual de piezas defectuosas es del 10% y la máxima proporción aceptable es del 20%. Se establece la siguiente regla de decisión para control de calidad: se toma una muestra de 10 unidades y se cuenta el número de piezas defectuosas; si el número de defectuosas es 3 o más se decide que la proporción real de defectuosas es mayor que el 10%, en otro caso se admite que la proporción real de tales piezas no sobrepasa el 10%.

a) Decir explícitamente cuáles son las hipótesis del contraste.

b) Calcular el error tipo I.

c) Calcular el error tipo II, cuando la verdadera proporción de defectuosas fuera del 20%.

d) Comentar si es buena esta regla de decisión.

15. Una perfumería lleva a cabo la siguiente prueba: Se le pregunta a 15 clientes, elegidos aleatoriamente, cuál de dos fragancias que se le presentan prefiere. Sin saberlo los entrevistados se trata del mismo perfume, la única diferencia es que uno de los envases tiene una apariencia mucho más cara. El dueño piensa que los compradores tenderán a elegir el perfume del envase con mejor apariencia. Resulta que 10 de los entrevistados eligen ese perfume. ¿Tiene razón el dueño de la perfumería?

16. Los siguientes datos representan los tiempos de funcionamiento en horas de tres tipos de calculadoras de bolsillo, antes de que necesiten ser recargadas:

	CALCULADORA	
A	B	C
4.9	5.5	6.4
6.1	5.4	6.8
4.3	6.2	5.6
4.6	5.8	6.5
5.3	5.5	6.3
5.2	6.6	4.8

Estudiar la hipótesis de que los tiempos de funcionamiento de las tres calculadoras son iguales, con nivel de significación 0.01.

17. Se tiene una muestra aleatoria de una población normal cuya varianza es 36. A partir de la media muestral, al 95% de confianza, se obtiene que los límites del intervalo de confianza para la media poblacional son 21.08 y 28.92.

a) Calcular la media muestral y el tamaño de la muestra.

b) Calcular el tamaño que debería tener la muestra para que el intervalo de confianza reduzca su amplitud a la mitad sin variar el nivel de confianza.

18. La tabla adjunta da las edades X y las presiones sanguíneas en sístole Y de 12 mujeres.

Edad	56	42	72	36	63	47	55	49	38	42	68	60
Presión	147	125	160	118	149	128	150	145	115	140	152	155

Calcular:

a) El coeficiente de correlación y la recta de regresión de Y sobre X.

b) Decir si es válido el modelo lineal.

c) Intervalo de confianza al 95% de la presión sistólica de las mujeres de 45 años.

19. Supongamos que se desea contrastar la hipótesis H_0: a = 5, contra la alternativa H_1: a = 8, por medio de un solo valor observado de una variable aleatoria con función de densidad:

$$f(x,\ a) = (1/a)\, e^{-x/a}$$

Si el tamaño máximo del error tipo I es 0.15, ¿cuál de las siguientes pruebas es la mejor para contrastar nuestras hipótesis?

a) Rechazar H_0 si $x \geq 9$.

b) Rechazar H_0 si $x \geq 10$.

c) Rechazar H_0 si $x \geq 11$.

20. El número de horas de sueño ganadas por 7 personas después de la administración de un somnífero fue:

1 , 0,5 , 1,1 , -0,7 , 3 , 1,2 , 1,4

Testar si el número de horas ganadas se desvía significativamente de cero, usando el test T de Student, el test de rangos con signo de Wilcoxon y el test de los signos. Observar que cada test es más sensible que el siguiente. Decir cuáles son las hipótesis en cada caso y sacar conclusiones.

21. En 22 parcelas de la misma extensión sembradas con remolacha se han medido las variables X = "número de plantas, en millares" e Y = "producción de azúcar", en Tm/Ha". Los datos obtenidos fueron:

$$\overline{X} = 31,2 \qquad \overline{Y} = 59,3 \qquad S_X = 7,4 \qquad S_Y = 12,6 \qquad r = 0,918$$

a) Respecto a la plantación ¿se puede aceptar que aumentando 1000 plantas por parcela, el aumento esperado de azúcar será 2 Tm/ha? (α = 0.01).

b) Si en una parcela se han plantado 28 millares de plantas:

i) ¿Entre qué valores se encuentra la producción esperada de azúcar, con probabilidad 0.95?

ii) Hallar el intervalo de confianza para la predicción de la producción de azúcar.

c) ¿Qué diferencia hay entre los resultados i) y ii)?

22. Una v.a. discreta toma los valores 0, 1 y 2 con función de probabilidad:

$$P(X=0) = p^2 \qquad P(X=1) = 2\,p\,(1\text{-}p) \qquad P(X=2) = (1\text{-}p)^2 \quad \text{con} \quad 0 < p < 1.$$

Estimar p, por el método de máxima verosimilitud, a partir de una muestra de tamaño 100 en la que se ha presentado 22 veces el 0, 53 veces el 1 y 25 veces el 2.

23. Por fistulización se obtuvo el pH de seis muestras de bilis hepática, con los siguientes resultados:

7.83 , 8.52, 7.32, 7.79, 7.57, 6.98.

Se desea saber, al nivel de significación 0.05, si la bilis hepática puede considerarse neutra (se supone normalidad).

Si se conociera que la desviación típica σ =0.5, ¿qué decisión tomaríamos? (El pH neutro es 7).

24. Se ensayaron dos tratamientos antirreumáticos administrados al azar sobre 10 y 22 pacientes; los resultados, medidos en una escala convencional (a mayor puntuación más eficacia) fueron:

tratamiento 1	12 15 21 17 38 42 10 23 35 28
tratamiento 2	21 18 42 25 14 52 65 40 43 35 18 56 29 32 44 15 15 68 41 37 43 58

Decidir si existe diferencia entre los tratamientos.

25. Al medir el tiempo de reacción ante cierto estímulo, un psicólogo estima que la desviación típica del mismo es de 0.05 s. ¿Cuál será el número de medidas que deberá realizar para tener a) el 99%, b) el 95% de confianza de que el error en su estimación del tiempo medio de reacción no excederá de 0.01 s?

26. En un experimento sobre percepción extrasensorial (P.E.) un sujeto fue preguntado sobre el color (rojo o azul) de una carta elegida por una persona que se hallaba en otra habitación de entre 50 cartas bien barajadas. Ambos individuos desconocen el número de cartas de cada color que hay en el mazo de cartas. Si el sujeto identifica correctamente 32 cartas, determinar si este resultado es significativo.

27. Los resultados de una encuesta, hecha para determinar si la edad de los conductores a partir de 21 años tiene influencia en el número de accidentes de tráfico, se indican en la tabla adjunta. Estudiar la hipótesis de que el número de accidentes es independiente de la edad del conductor.

	Edad del conductor					
nº accid.	21 - 30	31 - 40	51 - 50	51 - 60	61 - 70	totales
0	748	821	786	720	672	3747
1	74	60	51	66	50	301
2	31	25	22	16	15	109
3 ó más	9	10	6	5	7	37
totales	862	916	865	807	744	4194

28. El test de personalidad de Eysenk tiene dos formas: A y B. Se supone que ambas miden igualmente la extroversión. Probar esta afirmación a partir de los siguientes datos datos:

	Sujeto											
	1	2	3	4	5	6	7	8	9	10	11	12
forma A	12	18	21	10	15	27	31	6	15	13	8	10
forma B	10	17	20	5	21	24	29	7	11	13	8	11

29. Un contratista encarga un pedido de vigas cuya longitud promedio es de 5 m. Se sabe que la longitud de una viga se distribuye normalmente con desviación estandar de 0.02 m. Después de recibir el pedido se seleccionan 16 vigas al azar y se miden: si la media muestral es menor que la esperada, se devuelve el pedido.

a) Si la probabilidad de rechazar el pedido es de 0.04, ¿cuál debe ser el valor de la media muestral para que el pedido sea rechazado?

b) Si la longitud promedio real es de 4.98 m, ¿cuál es la potencia de la prueba del apartado a)?

30. Se pide a un laboratorio que compare la durabilidad de pelotas de golf de cuatro marcas. Para ello se seleccionan de forma aleatoria 8 pelotas de cada marca y una máquina las golpea repetidamente, con fuerza constante. Se mide el número de golpes que cada pelota resiste sin deteriorarse.

MARCA			
A	B	C	D
205	242	237	212
229	253	259	244
238	226	265	229
214	219	229	272
242	251	218	255
225	212	262	233
209	224	242	224
204	247	234	245

a) ¿Existe alguna razón para creer que la durabilidad promedio es diferente para alguna de las cuatro marcas? ($\alpha = 0.05$)

b) ¿Existe alguna razón para dudar de la suposición de que las varianzas de los errores son iguales?

31. En el censo de población de una ciudad se han obtenido los siguientes datos referidos al número de miembros por unidad familiar:

nº de miembros	1 a 2	3 a 4	5 a 6	7 ó más
nº de familias	9200	12400	6300	4100

Escogida una muestra de 600 familias se obtuvo la siguiente distribución:

nº de miembros	1 a 2	3 a 4	5 a 6	7 ó más
nº de familias	190	260	100	50

¿Podemos considerar que esta es una muestra aleatoria a un nivel de significación del 10%?

32. Un investigador desea evaluar el porcentaje de habitantes de una población que están vacunados contra la polio. Para ello planifica un muestreo de la población, deseando obtener resultados correctos dentro del ±3% con una confianza del 99%. De su experiencia previa, el investigador cree que el porcentaje de la población que está vacunada contra la polio está entre el 70% y el 80%. Si puede obtener una muestra aleatoria de individuos, ¿cuántos debe incluir en la muestra?

33. Para probar las diferencias respecto a la conducción eléctrica de tres tipos de alambres A, B y C, se obtuvo muestras y se midió la resistencia en ohmios, con los resultados que se acompañan. Estudiar si existen diferencias significativas al nivel 0.05.

A	0.135	0.138	0.144	0.142	0.139
B	0.138	0.143	0.140	0.137	0.141
C	0.141	0.143	0.145	0.142	0.138

34. La tabla adjunta muestra las respectivas alturas X e Y de una muestra de 12 padres y sus hijos varones mayores.

X = estatura del padre, en pulgadas

Y = estatura del hijo, en pulgadas

X	65	63	67	64	68	62	70	66	68	67	69	71
Y	68	66	68	65	69	66	68	65	71	67	68	70

a) Construir un diagrama de dispersión.

b) Hallar la recta de regresión de Y sobre X.

c) Calcular la variación total y la variación explicada por el modelo lineal.

d) Calcular el coeficiente de regresión y el coeficiente de determinación, y explicar si podemos usar la recta de corrrelación para predicción de las alturas de los hijos a partir de la altura de los padres.

35. Dos personas A y B juegan a cara y cruz con una moneda. A elige cara y B cruz. Al cabo de 100 partidas A ha ganado 62 veces. Tras este resultado, B afirma que la moneda está trucada, y que la probabilidad de salir cara es mayor que 2/3. A mantiene que la moneda es correcta. ¿Quien tiene razón? Elegir un nivel de significación del 5% y calcular la potencia del test.

36. La tabla muestra el número de soldados muertos en el ejército prusiano por coces de caballos, en diferentes unidades militares de caballería y es debida a Bortkiewicz. ¿Puede decirse que la variable sigue un modelo de Poisson?

Nº de muertos	0	1	2	3	4
Unidades con dichos muertos	109	65	22	3	1

37. En un estudio para determinar el efecto de una droga sobre la agresividad, se formaron dos grupos: el A, al que se administró la droga, y el B que recibió un placebo. Después de la administración se realizó una prueba para medir la agresividad. Las puntuaciones obtenidas, a mayor puntuación mayor agresividad, fueron:

grupo A	10 8 12 16 5 9 7 11 6
grupo B	12 15 20 18 13 14 9 16

Expresar el estudio en términos estadísticos formales y establecer la conclusión.

38. La tabla adjunta representa el rendimiento de la cosecha de un cereal en 10 pares de parcelas, con un abono fosfórico y la correspondiente sin él, siendo las demás condiciones las mismas. Admitiendo que la distribución de rendimientos es normal, ¿se puede concluir la eficacia del abono fosforado?

parcela 1	6.5	5.6	6.6	6.1	5.8	6.0	6.1	6.3	6.1	6.6
parcela 2	5.4	5.8	5.4	5.8	5.7	5.4	5.7	6.0	5.3	6.0

39. Contrastar si la muestra siguiente de duración de vida puede suponerse exponencial de parámetro 0.089:

16 8 10 12 6 10 20 7 2 24

40. El producto nacional bruto (PNB) *per capita* es reconocido como un estimador del nivel de vida de una nación. A su vez, se sabe que el consumo de energía eléctrica *per capita* es un buen indicador del PNB *per capita.* A continuación se da una lista de algunas naciones con sus correspondientes datos del PNB, en dólares, y de consumo de energía en kWs/h *per capita.*

PAÍS	PNB	KWS/H
India	110	81
Ecuador	240	399
Colombia	290	503
Perú	330	330
Chile	510	370
Costa Rica	510	754
Méjico	580	461
Panamá	660	1233
España	820	1013

Venezuela	1000	715
Argentina	1060	2487
URSS	1200	2007
Japón	1430	2377
Reino Unido	1890	3796
Alemania Federal	2190	3205
Francia	2460	2287
Estados Unidos	4240	6612

Hacer un estudio de la regresión lineal entre el consumo de energía y el PNB:

a) Dar la recta de regresión.

b) Estudiar la validez del modelo lineal.

c) Sabiendo que el consumo de energía de Canadá es 2310 Kws/h. predecir su PNB y dar un intervalo de confianza al 95% para el verdadero valor del PNB *per cápita* de Canadá.

41. Los ingenieros consideran válido suponer que la dureza de un aluminio tratado térmicamente está normalmente distribuida con una desviación estandar igual a 0.5. Deseamos probar la hipótesis H_0: m = 13.7 contra H_1: m≠ 13.7. La regla de decisión adoptada es: a) tomar una muestra de 30 piezas de aluminio; b) calcular la media muestral; c) si la media muestral es menor que 13.60 o mayor que 13.80 rechazamos H_0. En este caso calcular α y luego calcular β suponiendo que la verdadera dureza media del metal es 13.85. Comentar esta regla de decisión.

42. Para una red de alto voltaje se necesitan cables uniformes de gran resistencia a la tensión. Tenemos 4 tipos de cables y los resultados de medir la resistencia a la tensión de doce cables de cada tip se indican en la tabla. Como la resistencia a la tensión de cada cable es de 340 Kg. aproximadamente, para simplificar los cálculos hemos restado de todas las observaciones 340. Interesa contrastar la uniformidad de los cables.

cable 1	cable 2	cable 3	cable 4
5	-11	0	-12
-13	-13	-10	4
-5	-8	-15	2

-2	8	-12	10
-10	-3	-2	-5
-6	-12	-8	-8
-5	-12	-5	-12
0	-10	0	0
-3	5	-4	-5
2	-6	-1	-3
-7	-12	-5	-3
-5	-10	-11	0

43. La tabla muestra las producciones por acre de cuatro semillas sembradas en campos tratados con tres fertilizantes distintos. Con nivel de significación 0.01 determinar si hay diferencias en producción por acre: a) debida a los fertilizantes y b) debida a las semillas.

	Finca 1	Finca 2	Finca 3
A	12	15	14
B	15	19	18
C	10	12	15
D	14	11	12

44. En un experimento diseñado para saber si hay diferencia en la eficacia de 5 máquinas A, B, C, D y E, cinco operarios **expertos** trabajaron con cada una de las máquinas, todos el mismo tiempo con cada una. Los resultados se recogen en la tabla, en número de unidades producidas. Contrastar la hipótesis de que no hay diferencia entre ellas.

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

45. En dos provincias españolas se realizó una encuesta a 70 y 60 personas respectivamente, sobre la distribución de gastos semanales en alimentación (en miles de pts.), obteniéndose:

	0-5	5-10	10-15	15-20	20-25	total
prov. A	9	25	21	11	4	70
prov. B	5	20	15	11	9	60

a) Estudiar la homogeneidad del gasto en alimentación en las dos provincias.

b) Describir detalladamente todos los pasos del contraste.

46. Con objeto de analizar si existe relación entre el consumo de energía eléctrica Y_i (en kW/h) y el volumen de producción X_i (en millones de pts.) de una empresa, se ha obtenido la siguiente información:

X_i millones de pts.	2.1	3.2	5.7	8.5	8.9	10.2	10.3	12.3	13.2
Y_i kW/h	5.5	7.2	7.5	9.2	9.0	9.3	8.7	10.2	11.1

a) Determinar la ecuación de la recta que mejor explique el consumo de electricidad en función del volumen de producción.

b) Si la empresa tuviese un volumen de producción de 9.2 millones de pts. ¿entre qué valores estaría el gasto en electricidad con confianza del 90%?

47. Supongamos que una caja contiene circuitos de tres colores: rojo, marrón y azul y que se desea contrastar la hipótesis nula de que las proporciones de circuitos de cada uno de los tres colores son iguales, frente a la alternativa de que esas proporciones no son todas iguales. Supongamos que el método de contraste elegido consiste en seleccionar al azar tres circuitos de la caja y rechazar la hipótesis nula si, y solo si, al menos dos de los circuitos de la muestra son del mismo color.

a) Calcular el nivel de significación del contraste.

b) Calcular la potencia del contraste si 1/7 de los circuitos son rojos, 2/7 son marrones y los restantes azules.

48. Se ha aplicado un cierto test de memoria a un gran número de estudiantes de bachillerato, que dio como desviaciones típicas para las alumnas y alumnos 33,5 y 38,5, respectivamente. Aplicando el test a una muestra aleatoria simple de 18 alumnos dio un promedio de 183,7 y a una muestra aleatoria simple de 25 alumnas dio una media de 165,4. Suponiendo que las poblaciones son independientes y normales, ¿son estos resultados significativos de una diferencia de memoria entre ambos sexos?

49. Para contrastar la hipótesis de que una moneda está equilibrada mediante lanzamientos sucesivos, se imponen las siguientes restricciones:

a) La probabilidad de error tipo I debe ser como mucho 0,05.

b) La probabilidad de error tipo II cuando p difiera de 0,5 en 0,1 o más, debe ser a lo sumo 0,05.

Determinar el tamaño mínimo de la muestra necesario y enunciar la regla de decisión.

50. Se desea probar la eficacia de un fármaco sobre la angina de pecho. Se toma una muestra de 10 enfermos y se mide el tiempo que son capaces de realizar un determinado esfuerzo hasta sentir dolor. Los datos en segundos son:

Enfermo	1	2	3	4	5	6	7	8	9	10
Antes	720	600	160	245	125	185	80	136	124	226
Después	780	635	225	310	75	302	80	184	122	288

Se pide:

a) Estudiar si hay diferencias significativas, suponiendo normalidad.

b) Ídem sin suponer normalidad.

51. A los datos obtenidos al observar las roturas de un telar medidas cada 10 000 pasadas, se ajustó una distribución de Poisson. Comprobar la bondad del ajuste.

nº de roturas	0	1	2	3	4	5	6
frecuencias	3	39	48	16	5	1	1

52. Un proceso de producción emplea 5 máquinas en sus 3 procesos de desplazamiento. Se clasificó una muestra aleatoria de 164 fallas de acuerdo con la máquina y el tipo de desplazamiento en que ocurrió la falla obteniéndose los datos de la tabla. De acuerdo con esta información, ¿existe fundamento para dudar de la independencia entre la operación de desplazamiento y la falla de la máquina? (α = 0,01).

	Máquina				
Desplazamiento	A	B	C	D	E
tipo 1	10	12	8	14	8
tipo 2	15	8	13	8	11
tipo 3	12	9	14	12	10

53. Se selecciona una muestra de 200 votantes y se encontró que 114 apoyan determinada propuesta. Obtener un intervalo de confianza del 96% para la proporción de la población votante que está a favor de la propuesta.

¿De qué tamaño debe tomarse la muestra si se quiere una confianza del 96% de que la proporción muestral estará dentro de 0,02 de la proporción verdadera de la población de votantes?

54. Los siguientes datos representan los tiempos de duración de películas producidas por dos compañías. Calcular un intervalo del 95% de confianza para σ_1^2 / σ_2^2.

Compañía A	103	94	110	87	98		
Compañía B	97	82	123	92	175	88	118

55. Una empresa de automóviles quiere averiguar si el sexo de sus clientes tiene o no relación con el modelo elegido. Se toma una muestra de 2000 clientes y se tiene:

		modelo		
		A	B	C
sexo	mujer	340	400	260
	hombre	350	270	380

56. Un investigador afirma que el promedio de vida de los ratones puede ser ampliado hasta en 8 meses cuando se reducen las calorías en su comida aproximadamente un 40% del tiempo en que se les da de mamar. Las dietas restringidas son enriquecidas hasta niveles normales con proteínas y vitaminas. Una muestra aleatoria de 10 ratones es alimentada con una dieta normal y vive un promedio de 32,1 meses con desviación estándar de 3,2 meses, mientras que una muestra aleatoria de 15 ratones es alimentada con la dieta restringida y vive un promedio de 37,6 meses, con una desviación estándar de 2,8 meses. Con un nivel de significación de 0,05, probar la hipótesis de que el promedio de vida de los ratones con la dieta restringida, se incrementa 8 meses, en contra de la alternativa de que el incremento es menor de 8 meses. Suponer que las distribuciones de la duración de vida con las dietas regular y restringida tienen varianzas iguales y son normales.

57. Una zapatería es abastecida por cuatro fabricantes. Cada zapato es inspeccionado antes de ponerlo en venta. Hay tres defectos diferentes que causarían la devolución al fabricante. En cierta muestra se encontraron los siguientes defectos.

		Defecto		
		I	II	III
Fabricante	A	17	10	13
	B	10	10	10
	C	18	15	17
	D	15	5	10

¿Puede decirse que el tipo de defecto es independiente del fabricante?

58. Tenemos el contraste: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$.

Calcular su potencia para dos m. a. s. de tamaños $n_1 = 10$ y $n_2 = 5$, sobre poblaciones normales, con $\alpha = 0,1$ y supuesto que $\sigma_1^2 = 2\sigma_2^2$.

59. De una población se ha extraído una muestra y se midió la variable X, obteniéndose la siguiente tabla de frecuencias:

valores de X	1	2	3	4	5
frecuencia	20	40	85	53	13

Estudiar si se puede aceptar que la distribución de la variable X es uniforme. ¿Qué hipótesis habrá que hacer para llevar a cabo el estudio?

60. Una prueba clínica se basa en la asignación al azar de 40 pacientes a grupos de tratamiento y control de forma que se obtienen 20 individuos en cada grupo. Entre los 20 pacientes control se obtienen 12 aciertos. ¿Cuántos aciertos deben obtenerse entre los 20 pacientes sometidos a tratamiento para poder concluir, con un nivel del 5%, que el índice de aciertos es significativamente mayor en el grupo de tratamiento que en el de control?

61. Se ha observado el número de hijos varones en 1000 familias que tuvieran 5 hijos en total, obteniéndose los siguientes resultados:

nº de hijos varones	0	1	2	3	4	5
nº de familias	31	168	319	308	150	24

Estudiar la bondad de ajuste a una distribución binomial, a un nivel de significación de 0.05.

62. En un estudio acerca de la precipitación pluvial y la cantidad de contaminación de aire eliminada, se obtuvieron los siguientes datos:

Calcular:

a) Recta de regresión de Y sobre X y el coeficiente de correlación.

b) ¿Es válido el modelo lineal? Justificar la respuesta.

c) Intervalo de predicción, al 95% de las partículas eliminadas cuando la cantidad de lluvia diaria asciende a 4.8. cm^3/cm^2.

Lluvia diaria, X (cm^3/cm^2)	4,3	4,5	5,9	5,6	6,1	5,2	3,8	2,1	7,5
Partículas eliminadas, Y (mg/m)	126	121	116	118	114	118	132	141	108

63. Se ensaya un tipo de fertilizante sobre la producción de trigo de dos maneras: con riego y sin riego. Para ello se eligen 20 parcelas y se agrupan en diez pares, siendo cada par de la misma extensión, pero pares distintos no tienen por qué tener la misma extensión. Las producciones obtenidas (en Qm) después de la cosecha fueron:

Parcelas	1	2	3	4	5	6	7	8	9	10
con riego	46	110	70	54	60	120	82	76	37	28
sin riego	42	87	75	50	48	108	80	67	40	25

Con nivel de significación $\alpha=0,06$, ¿puede aceptarse que el fertilizante aplicado con riego, supera al aplicado sin riego?

64. En lanzamientos sucesivos de un dado, se han dado los siguientes resultados:

puntuación	1	2	3	4	5	6
frecuencia	95	92	80	123	140	70

¿Se puede decir que el dado es regular? (nivel de confianza 95%).

65. Se realizó un estudio para saber si el incremento en la concentración de sustrato tiene un efecto apreciable en la velocidad de una reacción química. La reacción se realizó 15 veces con una concentración de sustrato de 1.5 moles por litro, con una velocidad promedio de 7.5 micromoles por 30 minutos y una desviación estándar de 1.5. Con una concentración de sustrato de 2.0 moles por litro, se corrieron 12 pruebas, resultando una velocidad promedio de 8.8 micromoles por 30 minutos con una desviación estándar de 1.2. ¿Hay suficiente razón para creer que este incremento en la concentración de sustrato ocasiona un aumento en la velocidad promedio por más de 0.5 micromoles por 30 minutos? Utilice un nivel de significación de 0.02 y suponga que las poblaciones tienen una distribución aproximadamente normal.

66. En un experimento con guisantes, Gregor Mendel observó que 315 eran redondos y amarillos, 108 redondos y verdes, 101 rugosos y amarillos y 32 rugosos y verdes. De acuerdo con su teoría de la herencia, esos números debían estar en la proporción 9:3:3:1. ¿Hay alguna evidencia para dudar de su teoría al nivel de significación 0.05?

67. Tenemos una muestra de tamaño ***n*** de una población con distribución de Poisson, cuya función de densidad viene dada (como saben) por :

$$f(x) = (e^{-\lambda} \cdot \lambda^{x}) / x! \qquad x = 0, 1, 2, \ldots\ldots$$

Calcular la expresión del estimador de máxima verosimilitud del parámetro λ de dicha distribución.

68. Existe una rara especie de planta, cuya característica principal es su altura. Se sabe que la altura de este tipo de planta se distribuye normalmente con una desviación típica de 0.2 m. Un jardinero encarga para un parque un gran pedido de plantas de esta especie de 5 m de alto. Después de recibir el pedido el jardinero selecciona 16 plantas al azar y mide su altura: si la media muestral es menor que la esperada, se devuelve el pedido al vivero hasta que crezcan más.

a) ¿Cuál debe ser la altura promedio de las plantas seleccionadas para que se devuelva el pedido al vivero, si sabemos que la probabilidad de rechazar el pedido es de 0.04?

b) ¿Cuál es la potencia de la prueba descrita en el apartado anterior, si la altura promedio real de las plantas es de 4.9 m?

69. Si el coeficiente de correlación entre X e Y es 0.50. ¿Qué porcentaje de la variación total de Y queda inexplicado por la ecuación de regresión?

70. Durante un curso, un estudiante obtuvo las calificaciones que figuran en la tabla en algunos de sus exámenes. Determinar si hay diferencia significativa entre las calificaciones de las asignaturas que cursaba, a un nivel de significación de 0.05.

Matemáticas	72 80 83 75
Ciencias	81 74 77
Inglés	88 82 90 87 80
Economía	74 71 77 70

71. Las calificaciones de un grupo de estudiantes en su reporte de medio año (X) y en los exámenes finales (Y) fueron las siguientes:

X	77	50	71	72	81	94	96	99	67
Y	82	66	78	34	47	85	99	99	68

a) Estime la recta de regresión lineal de Y sobre X.

b) Intervalo de confianza al 95% para la calificación del examen final de un estudiante que obtuvo una calificación de 85 en el reporte de medio año.

72. La altura promedio de las mujeres en el grupo de primer año de una institución de enseñanza superior, se considera que se distribuye normalmente con una media de 162.5 cm y con una desviación típica de 6.9 cm. ¿Hay alguna razón, al nivel de significación de 0.05, para creer que existe un cambio en la altura promedio, si una muestra aleatoria de 50 mujeres del grupo actual tiene una altura promedio de 165.2 cm?

73. Supongamos dos muestras aleatorias de tamaños $n_1 = 10$ y $n_2 = 5$ de dos poblaciones normales de igual varianza de valor 9 y medias μ_1 y μ_2. Si la diferencia entre las medias poblacionales fuera 1) 0,1, 2) 3 y 3) 128, y el nivel de significación $\alpha_1 = 0,05$ y $\alpha_2 = 0,01$. Calcular la potencia del siguiente contraste:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

74. Se determina la pérdida de actividad de un preparado hormonal en el curso del tiempo y se obtiene el siguiente resultado.

Y = tiempo (meses)	1	2	3	4	5
X = % de actividad restante	90	75	42	30	21

Se desea calcular:

a) La expresión de la relación lineal entre ambas variables.

b) ¿Cuándo será del 25% la actividad del preparado hormonal?

c) A un nivel de significación de 0,05, ¿es aceptable el modelo lineal considerado?

4) Calcular el intervalo de confianza al 90% para el tiempo promedio transcurrido para que un preparado conserve el 85% de su actividad.

75. Los educadores creen que la habilidad de los alumnos de 2° curso puede ser mejorada por medio de un nuevo método con ayuda de computadoras, en lugar del método convencional. Los educadores saben que la medida de habilidad de los estudiantes sigue una distribución aproximadamente normal y que el promedio de habilidad para los alumnos que han aprendido con el método convencional se ha establecido en 78, siendo la desviación típica para ambos métodos de 7. Para probar esta hipótesis se calcula el promedio de habilidad de una muestra de 200 estudiantes. Si el promedio de la habilidad de esta muestra resulta ser mayor o igual que 79 se aceptará que el método nuevo mejora la habilidad, de otro modo se sigue aceptando como válido el método convencional.

a) Calcular el error tipo I.

b) Calcular el error tipo II, suponiendo que el verdadero promedio de habilidad con el nuevo método sea 78,5.

76. El rendimiento de la cosecha de un cereal se considera muy bueno si la producción es superior a 25 kg por área de cultivo, buena si es superior a 15 kg. y mala si no llega a 15. Se hacen 30 determinaciones del rendimiento en otras tantas parcelas donde se ha sembrado cereal de tipo A y 30 determinaciones en parcelas donde se sembró un cereal de tipo B. Los resultados fueron los siguientes:

	Tipo de cereal	
Rendimiento	A	B
Muy Bueno	10	12
Bueno	14	10
Malo	6	8

¿Son igualmente efectivos para el cultivo los dos tipos de cereales A y B? Tómese $\alpha = 0{,}05$.

77. Un fabricante de unidades de pantallas de vídeo prueba dos diseños de microcircuitos para determinar, con un nivel de significación de 0,05, si producen flujos de corriente equivalentes. La ingeniería de desarrollo ha obtenido los siguientes datos:

Diseño 1	$n_1 = 15$	$\bar{X}_1 = 24{,}2$	$S_1^2 = 10$
Diseño 2	$n_2 = 10$	$\bar{X}_2 = 23{,}9$	$S_2^2 = 20$

Suponiendo que ambas poblaciones son normales, aunque sin estar dispuestos a considerar que las varianzas sean iguales, ¿podemos afirmar que tiene razón el fabricante?

78. Un científico de computadoras ha desarrollado un algoritmo para generar enteros aleatorios en el intervalo 0 – 9. Codifica el algoritmo y genera 1000 dígitos aleatorios. Los datos obtenidos fueron:

0	1	2	3	4	5	6	7	8	9
94	93	112	101	104	95	100	99	108	94

¿Funciona bien el programa del científico? ($\alpha = 0{,}05$)

79. Sobre una muestra de tamaño 5 de una población se midieron dos variables X e Y y se obtuvo que el coeficiente de correlación de Pearson para esa muestra concreta resultó ser –0,082. Estamos interesados en probar:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

a) ¿Conseguiremos probar con este contraste si las dos variables son independientes?, ¿por qué?

b) ¿Cuál es el estadístico adecuado para realizar esta prueba? ¿Qué valor toma el estadístico en este caso?

c) Suponiendo que el valor del estadístico del contraste fuera 1,2 y tomando $\alpha = 0{,}1$, ¿cuál de entre las siguientes sería la conclusión?

i) Rechazar H_0 y decir que la correlación es significativa.

ii) Rechazar H_0 y decir que la correlación no es significativa.

iii) Aceptar H_0 y decir que la correlación es significativa.

iv) Aceptar H_0 y decir que la correlación no es significativa.

v) Suponer que se ha cometido un error de cálculo, ya que el valor del estadístico tiene que estar entre 0 y 1.

(Señalar la respuesta adecuada y explicar por qué es esa la elegida).

80. Se clasificó la progenie de cierta pareja en tres grupos, atendiendo a cierto atributo, obteniéndose las frecuencias 10, 53 y 46. Según un modelo genético las frecuencias deberían estar en la proporción:

$$p^2 : p(1 - p) : (1 - p)^2$$

¿Concuerdan los datos con el modelo?

81. El gerente de una empresa de neumáticos tiene la sospecha de que en una de las dos fábricas de la compañía existe un fallo en el control de calidad. Para contrastar esta sospecha se eligió 8 neumáticos de la primera fábrica y tras someterlos a un proceso acelerado de desgaste se obtuvo una duración media de 14 300 km, con una desviación típica de 4350 km. En la segunda fábrica se eligió una muestra de 21 neumáticos, que sometidos al mismo proceso de desgaste dieron una media de 13 800 km, con desviación típica de 3830 km. Estudiar si los desgastes medios son iguales.

82. De una población normal $N(\mu, 1)$ se observa una muestra de tamaño 5. Se considera el contraste de hipótesis simples:

$$\begin{cases} H_0: & \mu = 1 \\ H_1: & \mu = 3 \end{cases}$$

y la región crítica dada por $RC = \{(X_1, X_2, X_3, X_4, X_5)/\overline{X} > 2{,}1\}$. Calcular la probabilidad de los dos tipos de error.

83. Realizar un estudio lo más completo posible de la dependencia lineal de la concentración de glucosa en sangre y la de adrenalina en función de los siguientes datos, obtenidos para 8 pacientes.

Glucosa mg/c.c.	1.2	1.8	3.1	4.9	3.7	7.1	8.6	9.8
Adrenalina, mg/c.c.	4.5	5.9	7.0	7.8	7.2	6.8	4.5	2.7

84. En una clínica se quiere estudiar si existe diferencia significativa entre tres tipos de calmantes. Con este fin se suministran los calmantes a pacientes comparables. La variable de respuesta fue el tiempo en segundos desde la inyección hasta la desaparición del dolor. Los resultados obtenidos fueron:

Droga A	35	30	60	61	42	
Droga B	41	32	70	56	80	
Droga C	62	57	35	42	78	43

¿Producen las tres drogas el mismo efecto, en promedio? (Suponer normalidad).

85. Las edades al morir de una muestra de 19 individuos fallecidos de tuberculosis dan una media de 50 años con una desviación típica muestral de 6 años.

a) Hallar un intervalo de confianza al 95% para la media, suponiendo normalidad.

b) Qué tamaño muestral debe tomarse para tener una confianza del 95% de que la edad media estimada y la verdadera edad media no difieren en más de dos años.

86. Tenemos un dado del que nos han dicho que cada una de las caras pares tiene el doble de probabilidad de ocurrir que cada una de las caras impares. Realizamos 80 lanzamientos con este dado y obtenemos:

cara	1	2	3	4	5	6
frecuencia	6	22	14	15	5	18

¿Podemos afirmar que el dado no es del modelo descrito?

87. Una máquina embotelladora de gaseosas se considera fuera de control si la varianza del contenido excede el valor de 1,1 decilitros. En una muestra de 30 botellas se ha observado que la varianza es 1,5 decilitros. ¿Puede decirse que la máquina está fuera de control?

88. Se está estudiando el absentismo laboral femenino de una empresa. Se han elegido al azar diez obreras de una determinada sección de la misma y se anota el número de días de falta al trabajo por diversos motivos durante los últimos cuatro meses. Estos fueron los resultados.

Empleada	1	2	3	4	5	6	7	8	9	10
Absentismo (días)	5	4	6	8	7	4	2	7	6	1

Calcular un intervalo de confianza para el promedio de días de absentismo de todas las empleadas en los últimos cuatro meses.

89. De acuerdo con un estudio, las viudas viven más tiempo que los viudos. Con los siguientes datos de supervivencia, obtenidos para 100 viudas y 100 viudos después de la muerte del cónyuge, ¿se puede concluir que las proporciones de viudas y viudos son iguales respecto a la diferencia de periodos de tiempo que sobreviven después de la muerte de su compañero?

Años vividos	Viuda	Viudo
Menos de 5	25	39
Entre 5 y 10	42	40
Más de 10	33	41

90. Calcule e interprete el coeficiente de correlación muestral para las siguientes calificaciones de 6 alumnos elegidos al azar. Estudie si puede considerarse el coeficiente de correlación poblacional significativamente distinto de cero:

Matemáticas	70 92 80 74 65 83
Inglés	74 84 63 87 78 90

91. Se elige una muestra de tamaño 2 de una población de Poisson para contrastar las hipótesis $H_0 : \lambda = 1$ frente a $H_1 : \lambda = 2$. Se considera como región crítica $\{\overline{X} \geq 1{,}5\}$. Hallar el nivel de significación y la potencia del contraste.

92. En una determinada marca de cigarrillos se efectúa un experimento para estudiar el contenido de alquitrán: se estudian 20 cigarrillos elegidos al azar de diferentes lotes. Los datos muestrales para el contenido de alquitrán son:

$$\overline{X} = 22 \text{ mg} \qquad S = 4 \text{ mg}.$$

Encontrar un intervalo de confianza al 90% para el contenido medio de alquitrán en un cigarrillo de la citada marca.

93. Una empresa farmacéutica ha desarrollado un medicamento nuevo que requiere el control estricto de la dosificación del ingrediente activo de cada cápsula. El fármaco puede ser peligroso si se excede de la dosis, e inútil si la dosificación es menor que la especificada. El error de estimación máximo para la media es igual a 0,2 mg. Si se sabe que la desviación típica es igual a 0,7 mg, ¿cuántas cápsulas del medicamento deben estudiarse para tener una confianza del 99% de que es correcta la cantidad media de ingrediente por cápsula?

94. En una urna hay cientos de bolas de tres colores: blancas, rojas y negras. Nos han dicho que hay el triple de rojas que de blancas y el doble de negras que de rojas. Extraemos 20 bolas al azar y salen 4 blancas, 7 rojas y el resto negras. ¿Podemos afirmar que en la urna las bolas no estaban en la proporción indicada?

95. Se quiere hacer un estudio para predecir la calificación en Sistemas Operativos de un alumno de 2° de ITIS de la ULPGC a partir de las calificaciones obtenidas por el mismo alumno en la asignatura de Lenguajes de Programación de primer curso. Las calificaciones de una muestra aleatoria de alumnos son las de la tabla:

LP	2,5	7	5	4	5	6	5	8,3	5
SO	2	6	8	8	5,5	9	7	9	6

LP	6,5	5,5	3,5	5	7	8	3	4,5	6,2
SO	7	5,5	6	6,3	6,5	9	5	3	7,2

a) Hacer un estudio lo más completo posible de la regresión lineal.

b) ¿Puede admitirse que la pendiente de la recta de regresión es mayor que 1?

96. En un estudio sobre la importancia de la lactancia materna en la producción de anticuerpos en los bebés, se realiza un experimento que consiste en medir la proporción de niños que presentan una concentración de anticuerpos menor que un nivel de referencia prefijado. Se van a estudiar dos muestras de igual tamaño de entre los niños con lactancia materna y sin ella. Cuántos niños debemos elegir de cada población para aproximar la verdadera diferencia de proporciones con un error menor del 2%, a nivel de significación del 98%.

97. Queremos hacer un estudio de la proporción de personas de tienen un nivel de audición menor que el 60%. Sabemos que dicha proporción es menor del 20% ¿Cuál es el tamaño mínimo de la muestra que debemos tomar para tener una seguridad del 95% de que la proporción verdadera no difiere de la proporción muestral en más de 0,02?

Soluciones 1.

Álgebra de sucesos, probabilidad y variables aleatorias

1. Diez amigos quieren sortear quién paga la cena: cada uno coge una moneda, que suponemos equilibrada, de su bolsillo y las lanzan todos a la vez. Si todos menos uno obtienen el mismo resultado, éste paga la cena; en caso contrario se repite el lanzamiento.

a) Probabilidad de que en una determinada jugada haya perdedor.

b) Distribución de probabilidad de las tiradas hasta para obtener un perdedor y el nº medio de tiradas necesarias.

Solución:

1) Sea X = número de caras cuando se lanzan las diez monedas, $X \in B(10,1/2)$.

$$P\text{ (haya perdedor)} = \mathrm{P}(\mathrm{X}=1)+\mathrm{P}(\mathrm{X}=9) = C_{10}^{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{9} + C_{10}^{9}\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right)^{1} = \frac{5}{2^8}$$

2) Sea Y = número de tiradas hasta que hay perdedor, $Y \in G(5/2^8)$, por tanto:

$$P(Y=k) = \left(\frac{2^8-5}{2^8}\right)^{k-1}\frac{5}{2^8} \qquad \text{y} \qquad E[Y] = \frac{2^8}{5} = 53.2$$

2. Dos de los 8 camiones de ocasión que va a comprar una compañía no tienen el motor en perfecto estado, hecho fácil de apreciar con una simple prueba. ¿Cuál de los tres casos siguientes es más probable?

a) Encontrar los dos camiones defectuosos eligiendo al azar, sin remplazamiento, cuatro de ellos.

b) Encontrar un camión defectuoso entre tres elegidos al azar, sin remplazamiento.

c) No encontrar ningún camión defectuoso entre dos elegidos al azar, y con remplazamiento.

Solución:

a) Sea X= nº de camiones defectuosos de los cuatro elegidos sin remplazamiento, X es una v. a. hipergeométrica H(8, 2; 4), por tanto la probabilidad pedida es:

$$P(X=2)=\frac{\binom{2}{2}\binom{6}{2}}{\binom{8}{4}}=\frac{\frac{6\cdot 5}{2!}}{\frac{8\cdot 7\cdot 6\cdot 5}{4!}}=\frac{6}{14}=\frac{3}{7}=\frac{48}{7\cdot 16}$$

b) Sea Y = nº de camiones defectuosos de los tres elegidos sin remplazamiento, Y es una v. a. hipergeométrica H(8, 2; 3) y por tanto la probabilidad pedida es:

$$P(Y=1)=\frac{\binom{2}{1}\binom{6}{2}}{\binom{8}{3}}=\frac{2\cdot\frac{6\cdot 5}{2!}}{\frac{8\cdot 7\cdot 6\cdot}{3!}}=\frac{30}{56}=\frac{15}{28}$$

c) Sea Z= nº de camiones defectuosos de los dos elegidos con remplazamiento, *Z* es una v. a. Binomial *B*(2, 0,25), por tanto:

$$P(Z=0)=\binom{2}{0}\cdot\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^2=\frac{9}{16}=\frac{63}{16\cdot 7}$$

Por tanto el caso más probable es el c)

3. Un estudio calcula en 0.34 la prob. de que un camión pase la ITV a la primera; si el camión tiene menos de dos años esta prob. aumenta a 0.82. Sabiendo que el 60% del parque de camiones tiene más de dos años, calcular la prob. de que un camión de más de dos años pase la ITV a la primera.

Solución:

$$\begin{cases} A = \text{pasar el I.T.V.} & P(A) = 0.34 \\ B = \text{tener menos de 2 años} & P(B) = 0.4 \\ P(A/B) = 0.82 & \end{cases} \qquad P\left(A/\overline{B}\right) = ?$$

$$A = A \cap \Omega = A \cap (B \cup \overline{B}) = (A \cap B) \cup (A \cap \overline{B}) \Rightarrow$$

$$\Rightarrow P(A) = P(A \cap B) + P(A \cap \overline{B}) \Rightarrow$$

$$(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(A/B)P(B) = 0.34 - 0.82 \cdot 0.4$$

$$P(A/\overline{B}) = \frac{P(A \cap \overline{B})}{P(\overline{B})} = \frac{0.012}{0.6} = 0.02$$

4. Supongamos que hay una prueba para diagnosticar el cáncer que da positivo el 95% de las veces cuando una persona tiene cáncer y da negativo el 95% de las veces cuando se aplica en personas que no tienen la enfermedad. Si la probabilidad de que una persona tenga cáncer es 0.005, ¿cuál es la prob. de que una persona tenga realmente cáncer cuando la prueba ha dado positiva?

Solución:

$$C = \text{tener cáncer} \qquad P(C) = 0.005$$

$$A = \text{dar positivo} \qquad \begin{cases} P(A/C) = 0.95 \\ P(\overline{A}/\overline{C}) = 0.95 \end{cases} \qquad P(C/A) = ?$$

$$P(C/A) = \frac{P(C \cap A)}{P(A)} = \frac{P(A/C)P(C)}{P(A/C)P(C) + P(A/\overline{C})P(\overline{C})} =$$

$$= \frac{0.95 \cdot 0.005}{0.95 \cdot 0.005 + 0.05 \cdot 0.995} = \frac{0.00475}{0.05450} = 0.08715$$

5. Sea un circuito con n componentes conectadas en paralelo, que funcionan de manera independiente; es decir, el sistema funciona si y solo si funciona al menos uno de los componentes. Si el tiempo de duración de cada componente se distribuye normalmente con media de 50 horas y desviación típica de 5 horas, ¿cuál debe ser el nº de componentes para que la prob. de que el sistema falle durante las primeras 55 horas sea aproximadamente 0.01?

Solución:

Sea X_i= duración del componente i-ésimo, i=1, 2,...n. X_i sigue una distribución normal N(50, 5), para toda i. Queremos calcular **n** para que P(tiempo de duración del sistema ≤ 55 horas) ≅ 0,01, pero:

$$P \text{ (tiempo de duración del sistema} \leq 55 \text{ horas)} =$$

$$P(X_1 < 55, \ \ldots, \ X_n < 55) = \prod_{i=1}^{n} P(X_i \leq 55) = \prod_{i=1}^{n} P\left(Z \leq \frac{55-50}{5}\right) =$$

$$= 0,9772^n \cong 0,01 \quad \Rightarrow n \log 0,9772 \cong \log 0,01 \quad \Rightarrow n = \frac{\log 0,01}{\log 0,972} = 199,66$$

Luego el número de componentes necesarias es 200.

6. En el jardinero del Sr. Rodríguez no se puede confiar: la prob. de que olvide regar el rosal durante la ausencia del Sr. Rodríguez es 2/3.

El rosal es muy delicado: si se lo riega tiene igual prob. de progresar que de secarse, pero si no es regado tiene solo un 0.25 de prob. de progresar.

A su regreso, el Sr. Rodríguez se encuentra con que el rosal está seco. ¿Cuál es la prob. de que el jardinero se haya olvidado de regarlo?

Solución:

$$\begin{cases} A = \text{el rosal es regado} & P(A) = 0.33 \\ B = \text{el rosal progresa} & \begin{cases} P(B/A) = 0.5 & \Rightarrow & P(\overline{B}/A) = 0.5 \\ P(B/\overline{A}) = 0.25 & \Rightarrow & P(\overline{B}/\overline{A}) = 0.75 \end{cases} \end{cases}$$

$$P(\overline{A}/\overline{B}) = \frac{P(\overline{B}/\overline{A}) \cdot P(\overline{A})}{P(\overline{B}/\overline{A}) \cdot P(\overline{A}) + P(\overline{B}/A) \cdot P(A)} = \frac{0.75 \cdot 0.66}{0.75 \cdot 0.66 + 0.5 \cdot 0.33} = 0.75$$

7. Sea un dado tal que la probabilidad de las distintas caras es proporcional al nº de puntos inscritos en ellas. Hallar la prob. de obtener con este dado un nº par.

Solución:

Sea $p_i = P$(salir i ptos. en el dado), tenemos entonces que:

$$\frac{p_1}{1}=\frac{p_2}{2}=\frac{p_3}{3}=\frac{p_4}{4}=\frac{p_5}{5}=\frac{p_6}{6}=\frac{\sum p_i}{21}=\frac{1}{21} \Rightarrow p_i=\frac{i}{21} \quad \text{para i=1,...,6.}$$

$$P(\text{salir un n}^{\circ}\text{ par})=p_2+p_4+p_6=\frac{2}{21}+\frac{4}{21}+\frac{6}{21}=\frac{12}{21}=\frac{4}{7}$$

8. Cierto profesor lleva siempre en el bolsillo dos cajas iguales de N fósforos. Cada vez que va a encender un cigarrillo toma al azar una de las cajas; al cabo de cierto tiempo una de las cajas está vacía. Calcular la prob. de que en ese momento la otra caja contenga r fósforos, supuesto que se da cuenta de que la caja se queda vacía al coger el último fósforo de la misma.

Solución:

Vamos a definir éxito = sacar un fósforo de la caja A, (entendemos que la otra caja es B). Tenemos P(éxito)=0.5 y esta prob. se mantiene a lo largo de todo el experimento, porque no influye para nada el número de fósforos que cada caja contiene en cada momento.

Sea X = nº de la extracción (de fósforos) en que se obtiene el N-ésimo éxito. Esta v. a. X sigue una distribución Binomial negativa.

Como queremos que al extraer el último fósforo de la caja A queden exactamente r fósforos en la caja B, debemos haber sacado de ésta N-r fósforos. Por lo tanto el nº total de extracciones debe ser N+(N-r)=2N-r. La última extracción debe ser de la caja A y de las 2N-r-1 primeras N-1 deben ser tambien de A.

Por tanto $X \in BN(N\ ,\ 0.5)$ y se tiene:

$$P(X=2N\text{-}r)=\binom{2N-r-1}{N-1}0.5^{N-1}0.5^{N-r}0.5=\frac{(2N-r-1)!}{(N-1)!(N-r)!}0.5^{2N-r}$$

Como además podemos repetir todo el razonamiento suponiendo que la caja B es la primera que se termina bastará multiplicar por 2 esa cantidad, y así la probabilidad pedida será:

$$P(\text{Hay } r \text{ fosforos en una caja al coger el último de la otra})=$$
$$=\frac{2\cdot(2N\text{-}r\text{-}1)!}{(N-1)!(N-r)!}0.5^{2N-r}$$

9. Las probabilidades de que tres hombres hagan diana en una competición de tiro son, respectivamente, 1/5, 1/4 y 1/3. Cada hombre dispara una vez. Calcular:

a) Probabilidad de que solo uno de ellos haga diana.

b) Si sabemos que solo uno de ellos ha hecho diana, probabilidad de que haya sido el tercero.

Solución:

$$\begin{cases} A = \text{el Sr.1 hace diana} & P(A)=1/5 \\ B = \text{el Sr.2 hace diana} & P(B)=1/4 \\ C = \text{el Sr.3 hace diana} & P(C)=1/3 \end{cases}$$

a) $P(\text{solo uno de ellos hace diana})=$

$$= \text{P}((\text{A}\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C))=$$
$$=P(A\cap\overline{B}\cap\overline{C})+P(\overline{A}\cap B\cap\overline{C})+P(\overline{A}\cap\overline{B}\cap C)=$$
$$=\frac{1}{5}\cdot\frac{3}{4}\cdot\frac{1}{2}+\frac{4}{5}\cdot\frac{1}{4}\cdot\frac{1}{2}+\frac{4}{5}\cdot\frac{3}{4}\cdot\frac{1}{2}=\frac{19}{40}$$

b) $P(C/\text{solo uno ha hecho diana})=\dfrac{P(C\cap\text{solo uno ha hecho diana})}{\text{P(solo uno ha hecho diana)}}=$

$$=\frac{P(\overline{A}\cap\overline{B}\cap C)}{P(\text{solo uno hizo diana})}=\frac{\frac{4}{5}\cdot\frac{3}{4}\cdot\frac{1}{2}}{\frac{19}{40}}=\frac{12}{19}$$

10. Un señor piensa comprar acciones de cierta compañía. La cotización de las acciones en bolsa durante los seis meses inmediatamente anteriores es un dato de interés y se observa que dicha cotización se relaciona con el Producto Interior Bruto (PIB). Si el PIB aumenta, la prob. de que las acciones aumenten de valor es de 0.8; si el PIB se mantiene esa prob. es de 0.2; si el PIB disminuye la prob. es de 0.1. Si para los próximos seis meses las probs. de que el PIB suba, se mantenga, o baje, son 0.4, 0.3 y 0.3, respectivamente, calcular la probabilidad de que esas acciones aumenten de valor en los próximos seis meses.

Solución:

Sea A_1 = El PIB aumenta en los próximos 6 meses P(A_1) = 0,4.

A_2 = El PIB se mantiene en los próximos 6 meses P(A_2) = 0,3.

A_3 = El PIB baja en los próximos 6 meses P(A_3) = 0,3.

B = aumenta el precio de las acciones en los próximos 6 meses.

$$\begin{cases} P(B/A_1)=0,8 \\ P(B/A_2)=0,2 \\ P(B/A_3)=0,1 \end{cases} \qquad P(B)=?$$

$$P(B)=\sum_{i=1}^{3} P(B/A_i)\cdot P(A_i)=0,4\cdot 0,8+0,3\cdot 0,2+0,3\cdot 0,1=0,41$$

11. Se sabe que el 40% de personas de cierto grupo grande estarían dispuestas a formar parte del comité ecologista. Para constituir el comité, formado por 5 personas, seleccionamos al azar una persona del grupo y le preguntamos si quiere formar parte del mismo, hasta encontrar las cinco personas buscadas. Calcular:

a) Probabilidad de que tras quince consultas aún no esté formado el comité.

b) Probabilidad de que sean necesarias doce consultas para formar el comité.

Solución:

Tenemos que P (aceptar formar parte del comité) = 0.4 = p.

a) Sea X= nº de personas que aceptan de las 15 consultadas, $X \in B(15, 0.4)$ ya que consideramos las extracciones independientes por ser un grupo grande, por tanto:

$$P(X\le 4)=\sum_{i=0}^{4} P(X=i)=\sum_{i=0}^{4}\binom{15}{i}\ 0.4^i 0.6^{15-i}=0.2173$$

b) Sea Y = nº de consultas hasta la 5ª aceptación, $Y \in BN(5, 0{,}4)$.

$$P(Y = 12)) = \binom{11}{4} \; 0.4^5 0.6^7 = 330 \cdot 0.4^5 0.6^7 = 0.0946$$

12. El gerente de un restaurante que sólo atiende mediante reserva, sabe por experiencia que el 15% de las personas que reservan una mesa, no asisten. El restaurante solo tiene 20 mesas, pero acepta 25 reservas. ¿Cuál es la probabilidad de que determinado día, a todas las personas que acuden al restaurante se les asigne mesa?

Solución:

Sea X = nº de personas que no acuden al restaurante de las 25 que tienen reserva, $X \in B(25, 0{,}15)$.

P(todas las personas que acuden tienen mesa) =

= $P(X \geq 5)$ = 1-$P(X \leq 4)$ = 1-0,6821= 0,3179.

13. A un sospechoso se le aplica un suero de la verdad del que se sabe que es fiable en un 90% de los casos cuando la persona es culpable y en un 99% cuando la persona es inocente. En otras palabras, el 10% de los culpables son considerados inocentes cuando se usa el suero y el 1% de los inocentes son considerados culpables. Si el sospechoso se escogió de un grupo del cual solo el 5% ha cometido alguna vez algún delito y el suero indica que la persona es culpable, ¿cuál es la probabilidad de que sea inocente?

Solución:

A = el suero de la verdad dice que es culpable.

C = el sospechoso es culpable.

$$P(A/C) = 0.90 \quad \Rightarrow \quad P(\overline{A}/C) = 0.10$$
$$P(\overline{A}/\overline{C}) = 0.99 \quad \Rightarrow \quad P(A/\overline{C}) = 0.01$$
$$P(C) = 0.05$$

$$P(\overline{C}/A)=\frac{P(\overline{C}\cap A)}{P(A)}=\frac{P(A/\overline{C})P(\overline{C})}{P(A/\overline{C})P(\overline{C})+P(A/C)P(C)}=$$

$$=\frac{0.01\cdot 0.95}{0.01\cdot 0.95+0.90\cdot 0.05}=0.6785$$

14. Un fabricante de medicamentos afirma que cierto fármaco cura una enfermedad de la sangre el 80% de las veces, en promedio. Para comprobar la afirmación una oficina gubernamental utilizó el producto medicinal en una muestra de 100 individuos y decidió aceptar tal afirmación si 75 o más eran curados.

a) ¿Cuál es la probabilidad de que la afirmación sea rechazada cuando la probabilidad de cura sea, de hecho, 0.8?

b) ¿Cuál es la probabilidad de que la afirmación sea aceptada cuando la probabilidad de cura sea solo de 0.7?

Solución:

a) $P(\text{rechazar el farmaco})=P(X<75)=P(X\leq 74.5)=P\left(\frac{X\text{-}80}{4}\leq\frac{74.5-80}{4}\right)=$

$$=P\left(Z\leq\frac{-5.5}{4}\right)=P(Z\leq -1.375)=0.0860$$

b) Si $p=0{,}7$ entonces $X\in B(100,\ 0{,}7)\cong N\left(70,\ \sqrt{21}\right)$

$$P(\text{aceptar el fármaco})=P(X\geq 75)=P(X\geq 74.5)=$$

$$=P\left(\frac{X-70}{\sqrt{21}}\geq\frac{74.5-70}{\sqrt{21}}\right)=P\left(Z\geq\frac{4.5}{\sqrt{21}}\right)=P(Z\geq 0.9819)=$$

$$=1-P(Z\leq 0.9819)=1-0.8365=0.1635$$

15. En la asignatura de Bioestadística de cada 3 mujeres que se presentan al examen aprueban 2, y de cada 5 hombres aprueban 3. En la clase hay 4 veces más mujeres que hombres. En estas condiciones:

a) Si elegimos al azar un alumno y resulta que está aprobado, ¿cuál es la probabilidad de que sea mujer?

b) Si elegimos al azar a un alumno y resulta que ha suspendido, ¿cuál es la probabilidad de que sea un hombre?

Solución:

Sea M= el alumno elegido es una mujer, $P(M)$ = 4/5

A = el alumno elegido ha aprobado, tenemos $\begin{cases} P(A/M) = 2/3 \\ P(A/\overline{M}) = 3/5 \end{cases}$

a) $$P(M/A) = \frac{P(M \cap A)}{P(A)} = \frac{P(A/M)P(M)}{P(A/M)P(M) + P(A/\overline{M})P(\overline{M})} =$$

$$= \frac{\frac{2}{3}\cdot\frac{4}{5}}{\frac{2}{3}\cdot\frac{4}{5}+\frac{3}{5}\cdot\frac{1}{5}} = \frac{40}{49}$$

b) $$P(\overline{M}/\overline{A}) = \frac{P(\overline{M} \cap \overline{A})}{P(\overline{A})} = \frac{P(\overline{A}/\overline{M})P(\overline{M})}{1 - P(A)} = \frac{\frac{2}{5}\cdot\frac{1}{5}}{1-\frac{49}{75}} = \frac{\frac{2}{25}}{\frac{26}{75}} =$$

$$= \frac{2\cdot 75}{25\cdot 26} = \frac{3}{13}$$

16. Un club de estudiantes extranjeros tiene entre sus miembros a dos canadienses, tres japoneses, cinco italianos y dos alemanes. Si se selecciona al azar un grupo de cuatro personas, calcular:

a) Probabilidad de que todas las nacionalidades estén representadas.

b) Probabilidad de que todas las nacionalidades excepto la italiana estén representadas.

Solución:

a) P (todas las nacionalidades representadas en el comité) =

$$= P(I, A, J, C) = \frac{5.2.3.2}{C_{12}^4} = \frac{5\cdot 2\cdot 3\cdot 2}{\frac{12\cdot 11\cdot 10\cdot 9}{4\cdot 3\cdot 2}} = \frac{4}{33}$$

b) P (todas las nacionalidades, salvo la italiana, representadas en el comité) =

$$=P(2A, J, C)+P(A,2J,C)+P(A,J,2C)=$$

$$=\frac{1\cdot 3\cdot 2}{C_{12}^{4}}+\frac{2\cdot 3\cdot 2}{C_{12}^{4}}+\frac{2\cdot 3\cdot 1}{C_{12}^{4}}=\frac{24}{495}=\frac{8}{165}$$

17. En el lejano reino de Falandia a los condenados a muerte se les concedía la gracia de que su vida dependiera de sacar bola blanca en el siguiente sorteo: se colocaban 50 bolas blancas en una urna y 50 negras en otra, entonces el condenado sacaba una bola, eligiendo la urna al azar con los ojos vendados.

En cierta ocasión, un reo pidió la gracia de que se le dejara distribuir las bolas de otro modo antes de hacer el sorteo. Tras alguna discusión de los expertos se le concedió la gracia, y colocó una bola blanca en una urna y en la otra 49 blancas y 50 negras. ¿Cuál resultó de este modo la probabilidad de salvar la vida? ¿Aumentó respecto a la que hubiera tenido sin la modificación o, por el contrario, se redujo?

Solución:

La probabilidad de salvarse con la antigua colocaación es 0.5.

Sean $U_1 = \{1b\}$ y $U_2 = \{49b,50b\}$ las dos urnas con la composición pedida por el reo.

Con la nueva disposición de las bolas en las urnas tenemos:

$$P(\text{salvarse}) = P(b/U_1)P(U_1)+P(b/U_2)P(U_2)=1\cdot\frac{1}{2}+\frac{49}{99}\cdot\frac{1}{2}=\frac{148}{198}=\frac{74}{99}$$

El reo eligió bien porque la probabilidad de salvarse aumentó en

$$\frac{74}{99}-\frac{1}{2}=\frac{49}{198}=0.2474$$

18. Se lanza un dado equilibrado y el número que resulta, es el número de bolas blancas que ponemos en una urna; se vuelve a lanzar el dado, poniendo esta vez tantas bolas negras en la misma urna anterior como indique el resultado del lanzamiento. De la urna así formada, se extrae una bola al azar. ¿Cuál es la probabilidad de que sea blanca?

Solución:

Sea E = lanzar dos dados. El espacio muestral asociado será:

$$\Omega = \{(1,1), (1,2)....(1,6), (2,1)....(2,6).....(6,1),....(6,6)\}$$

donde cada suceso elemental tiene probabilidad de ocurrir 1/36.

La composición de la urna de la que se extrae la bola está condicionada al resultado del lanzamiento de los dados. Así si el suceso elemental ocurrido fue el par (i , j), la probabilidad de blanca sería $\frac{i}{i+j}$ ya que en ese momento habría i+j bolas en la urna, de las cuales i serían blancas. Por tanto:

$$P(b) = P(b \cap \Omega) = P(b \cap \{(1,1).....(6,6)\}) = P(b \cap (1,1)) + ... +$$

$$+P(b \cap (6,6)) = \sum_{i=1}^{6}\sum_{j=1}^{6} P(b \cap (i,j)) = \sum_{i=1}^{6}\sum_{j=1}^{6} P(b/(i,j))\, P((i,j)) =$$

$$= \sum_{i=1}^{6}\sum_{j=1}^{6} \frac{i}{i+j}\frac{1}{36} = \frac{1}{36}\cdot\sum_{i=1}^{6}\sum_{j=1}^{6}\frac{i}{i+j} = \frac{1}{36}(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{2}{3}+\frac{2}{4}+$$

$$+\frac{2}{5}+\frac{2}{6}+\frac{2}{7}+\frac{2}{8}+\frac{3}{4}+\frac{3}{5}+\frac{3}{6}+\frac{3}{7}+\frac{3}{8}+\frac{3}{9}+\frac{4}{5}+\frac{4}{6}+\frac{4}{7}+\frac{4}{8}+\frac{4}{9}+$$

$$+\frac{4}{10}+\frac{5}{6}+\frac{5}{7}+\frac{5}{8}+\frac{5}{9}+\frac{5}{10}+\frac{5}{11}+\frac{6}{7}+\frac{6}{8}+\frac{6}{9}+\frac{6}{10}+\frac{6}{11}+\frac{6}{12}) =$$

$$= \frac{1}{36}(\frac{1}{2}+\frac{3}{3}+\frac{6}{4}+\frac{10}{5}+\frac{15}{6}+\frac{21}{7}+\frac{20}{8}+\frac{18}{9}+\frac{15}{10}+\frac{11}{11}+\frac{6}{12}) = \frac{18}{36} = \frac{1}{2}$$

19. Hallar la probabilidad de obtener al menos un 4 en dos lanzamientos de un dado honrado.

Solución:

Sea A = { todos los resultados en que hay al menos un cuatro, cuando se lanzan dos dados}= { (1, 4), (2, 4), (3, 4), (4, 4), (5, 4), (6, 4), (4, 1), (4, 2), (4, 3), (4, 5), (4, 6)}

P(al menos un cuatro) = P(A) = 11/36

20. Cada vez que un cliente compra un dentífrico elige la marca A o la marca B. Supóngase que la probabilidad de que elija la misma marca que en la compra anterior es 1/3. Si la primera marca que adquiere es A. ¿Cuál es la probabilidad de que la quinta compra sea de la marca B?

Solución:

El conjunto de sucesos favorables es:

Ω = {AAAAB, AAABB, AABAB, AABBB, ABAAB, ABABB, ABBAB, ABBBB}

$$P(\text{ 5}^{\text{a}}\text{ compra es B}/1^{\text{a}}\text{ compra es A})=\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}+$$

$$+\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}+\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}+\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}+$$

$$+\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}+\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{40}{81}$$

21. En un sistema de alarma, la probabilidad de que ocurra una situación de peligro es 0,1; si éste se produce, la probabilidad de que la alarma funcione es 0,95; la probabilidad de que la alarme se dispare sin haber habido peligro es 0,03. Hallar:

a) Probabilidad de que habiendo funcionado la alarma, no haya habido peligro.

b) Probabilidad de que haya peligro y la alarma no funcione.

c) Probabilidad de que no habiendo funcionado la alarma, haya habido peligro.

Solución:

Sea A = ocurre una situación de peligro, $P(A)$ = 0,1.

$$B = \text{la alarma funciona, } \begin{cases} P(B/A)=0{,}95 \\ P(\overline{B}/\overline{A})=0{,}03 \end{cases}$$

a) $P(\overline{A}/B)=\dfrac{P(\overline{A}\cap B)}{P(B)}=\dfrac{P(B/\overline{A})\cdot P(\overline{A})}{P(B/A)\cdot P(A)+P(B/\overline{A})\cdot P(\overline{A})}=$

$=\dfrac{0{,}97\cdot 0{,}9}{0{,}95\cdot 0{,}1+0{,}97\cdot 09}=0{,}90$

b) $P(A\cap\overline{B})=P(\overline{B}/A)\cdot P(A)=0{,}05\cdot 0{,}1=0{,}005$

c) $P(A/\overline{B})=\dfrac{P(A\cap\overline{B})}{P(\overline{B})}=\dfrac{0{,}005}{P(\overline{B}/A)P/(A)+P(\overline{B}/\overline{A})P(\overline{A})}=$

$$=\frac{0{,}005}{0{,}05\cdot 0{,}1+0{,}03\cdot 0{,}9}=0{,}65$$

22. Supongamos que a una piscina se ha arrojado cierta cantidad de peces de dos tipos: A y B, y que tenemos un método de extracción al azar, y con remplazamiento. Sabemos que la proporción de peces tipo A es p, con $0<p<1$. Sea X = nº de la extracción en que aparece por primera vez un pez tipo A. Calcular la función de prob. y la de distribución acumulada de X, su media y su varianza. Ídem con Y = nº de peces tipo B antes de la n-síma extracción tipo A.

Solución:

X = nº de la extracción en que aparece por 1ª vez un pez tipo A.

X sigue una distribución Geométrica de parámetro p, $G(p)$.

$$P(X=k)=(1-p)^{k-1}\,p\ ; \qquad F(x)=\begin{cases}0 & \text{si } x<1\\ \sum_{i=1}^{k}(1-p)^{i-1}p & \text{si } k\le x<k+1\end{cases};$$

$$\mu_X=\frac{1}{p}\ ; \qquad \sigma_X^2=\frac{1-p}{p^2}$$

Y = nº de peces tipo B antes de la n-sima extracción tipo A.

$$P(Y = k) = \binom{n+k-1}{k} p^n (1-p)^k;$$

$$F(y) = \begin{cases} 0 & si\ y < 0 \\ \sum_{i=0}^{k} \binom{n+k-1}{k} p^n (1-p)^k & si\ k \leq y < k+1 \end{cases}$$

Para calcular la esperanza y la varianza de Y, en vez de calcularlo por la definición hacemos el siguiente razonamiento: Y no sigue ninguno de nuestros modelos de v. a. discretas, pero si definimos Z = nº de peces hasta la n-sima extracción tipo A, esta variable Z sigue una distribución Binomial negativa BN(n , p) y además Y= Z-n, con lo que:

$$\mu_Y = E[Y] = E[Z-n] = E[Z] - n = \frac{n}{p} - n = \frac{n-n\ p}{p} = \frac{n(1-p)}{p}$$

$$\sigma_Y^2 = V[Z-n] = V[Z] = \frac{n(1-p)}{p^2}$$

23. Cuatro aspirinas y dos tabletas para purificar el agua son colocadas accidentalmente dentro del mismo bote. Una profesora de Estadística tiene una palpitante jaqueca, probablemente producida por sus alumnos. Selecciona apresuradamente dos pastillas del bote y las traga empleando un vaso de agua.

a) ¿Cuál es la probabilidad de que simplemente purifique el agua que bebe?

b) ¿Cuál es la probabilidad de que seleccione una aspirina y una tableta para purificar el agua?

Solución:

a) $P(\overline{A}, \overline{A}) = \dfrac{2}{\binom{6}{2}} = \dfrac{2}{\frac{6!}{2!4!}} = \dfrac{2}{15}$

b) $P(A, \overline{A}) = \dfrac{2 \cdot 4}{\binom{6}{2}} = \dfrac{2 \cdot 4 \cdot 2 \cdot 4!}{6!} = \dfrac{8}{15}$

24. Un jurado de tres miembros decide por mayoría. Dos de los miembros dan su veredicto según su criterio y tienen una probabilidad p de decidir correctamente. El tercer miembro se limita a lanzar una moneda y si sale cara votará culpable. Cada uno de los tres da su voto en un papel sin consultar con los demás. Otro tribunal está compuesto por un sólo juez que tiene probabilidad p de dar un veredicto correcto. ¿Cuál de los dos tribunales juzga mejor?

Solución:

Sea C_k = el juez k juzga correctamente, para k=1,2,3.

Sea $P(C_1) = P(C_2) = p$, $P(C_3) = 0{,}5$.

P (juzgar correctamente) =

= P (al menos dos de los tres jueces juzgan correctamente) =

$$= P(C_1, C_2, C_3) + P(C_1, C_2, \overline{C}_3) + P(C_1, \overline{C}_2, C_3) + P(\overline{C}_1, C_2, C_3) =$$

$$= p \cdot p \cdot 1/2 + p \cdot p \cdot 1/2 + p \cdot (1-p) \cdot 1/2 + (1-p) \cdot p \cdot 1/2 =$$

$$= p^2 + p \cdot (1-p) = p \cdot (p + 1 - p) = p$$

Vemos que es lo mismo elegir este método de dar un veredicto que dejar que un solo juez que tiene probabilidad p de juzgar correctamente, juzgue solo.

25. La probabilidad de un tipo de accidente industrial es 0,001. Una compañía de seguros propone a una empresa un seguro de accidentes cuyo coste anual es de 10 000 pts., comprometiéndose en caso de accidente a pagar 5 millones de pts. en concepto de indemnización. Calcular el beneficio esperado para la compañía de seguros.

Solución:

Sea N el número de seguros de este tipo que la aseguradora tiene contratados, sea X = número de accidentes en un año.

Sea B el beneficio, será B = 10.000 N-5.000.000 X,

$$E[B] = 10^4 N - 5\times10^6 E[X] = 10^4 N - 5\times10^6 \times N \times 0.001 = 5000N$$

Luego la ganancia esperada es de 5000 pts. por cada seguro contratado.

26. Una compañía aérea ha observado que el 12% de las plazas reservadas no se cubren y decide aceptar reservar por un 10% más de las plazas disponibles en aviones de 450 plazas. Calcular la proporción de vuelos en que algún pasajero con reserva, no tiene plaza. (Indicar las hipótesis que hay que hacer para resolver el problema).

Solución:

Sea A = una persona con reserva acude al aeropuerto, P(A) = 0,88.

Sea X = nº de personas de las 495 (450+10% de 450=495) que tienen reserva que acuden. Debemos suponer que el hecho de que un pasajero con reserva acuda o no al aeropuerto es independiente de lo que le ocurra a cualquier otro en tal caso $X \in B(495, 0{,}88) \cong N(435{,}6\ , 7{,}23)$. Así:

P(algún pasajero con reserva, no tenga plaza) = $P(X > 450)$ =

$$= P\left(\frac{X - 435{,}6}{7{,}23} > \frac{450{,}5 - 435{,}6}{7{,}23}\right) = P(Z > 2{,}06) = 0{,}0197$$

Vemos que algo menos del 2% de las veces ocurre que algún pasajero que tiene plaza reservada se queda sin asiento en el avión, lo que se denomina "overbooking".

27. Una empresa recibe piezas de un proveedor en lotes de 2000 que se someten al siguiente control de calidad: se toman 20 al azar y si hay más de una defectuosa se rechaza el lote; en otro caso se acepta. La calidad garantiza por el proveedor es un 8 por mil de defectuosas. Calcular la probabilidad de:

a) Aceptar un lote que contenga un 2% de defectuosas.

b) Rechazar un lote que debería ser aceptado al tener sólo el 8 por mil de defectuosas.

Solución:

Sea X = número de piezas defectuosas en las 20 observadas. Podemos considerar que la probabilidad de defectuosa se mantiene constante durante todo el experimento, al considerar que el número de piezas fabricadas es suficientemente grande, aunque realmente no sea así porque las extracciones se realizan sin remplazamiento. En este caso sería $X \in B(20\ , p)$ y se tiene:

a) $$P(X \leq 1/\ p = 0,02) = \left[P(X=0) + P(X=1)\right]_{p=0,02} =$$
$$= \binom{20}{0}(0,02)^0(0,98)^{20} + \binom{20}{1}(0,02)(0,98)^{19} = 0,94$$

b) $$P(X \geq 2/\ p = 0,008) = \left[1 - P(X=0) - P(X=1)\right]_{p=0,008} =$$
$$= 1 - \binom{20}{0}(0,008)^0(0,992)^{20} - \binom{20}{1}(0,008)(0,992)^{19} = 0,011$$

Vemos que aceptaremos casi todos los lotes, el 94%, con un 2% de defectuosas, pero casi no rechazaremos ningún lote que cumpla las especificaciones del fabricante (sólo el 1,1%). Esta regla de decisión es buena para el fabricante de las piezas, pero muy mala si lo que queremos es garantizar la detección de lotes de piezas en que la proporción de defectuosas sea superior a la garantizada por el fabricante. (Observemos que el 2% es 2,5 veces más que el 8 por mil).

28. Se desea capturar 14 gacelas. Para la expedición se dispone de tres vehículos que pueden transportar 500, 750 y 1000 kg de carga respectivamente. Decidir cuál sería el vehículo más idóneo para la expedición, si se desea usar el más ligero capaz de transportar las 14 gacelas. Se sabe que el peso de una gacela sigue una distribución N(50,6).

Solución:

Sea X_i = peso de la gacela i-ésima, $X \in N(50\ , 6)$

$$Y = \text{peso de las 14 gacelas} = \sum_{i=1}^{14} X_i, \quad Y \in N(700, 6\sqrt{14})$$

Ahora debemos calcular la probabilidad de que el peso de las 15 gacelas, no exceda de la capacidad de carga de los vehículos.

$$P(Y \leq 500) = P\left(\frac{Y-700}{6\sqrt{14}} \leq \frac{500-700}{6\sqrt{14}}\right) = P\left(Z \leq -8.91\right) \approx 0$$

$$P(Y \leq 750) = P\left(\frac{Y-700}{6\sqrt{14}} \leq \frac{750-700}{6\sqrt{14}}\right) = P(Z \leq 2.23) = 0.9868$$

$$P(Y \leq 1000) = P\left(\frac{Y-700}{6\sqrt{14}} \leq \frac{1000-700}{6\sqrt{14}}\right) = P(Z \leq 13.36) \approx 1$$

Está claro que el vehículo más idóneo es el furgón de 750 kg, ya que menos del 2% de las veces no podremos cargar las 14 gacelas en él.

29. La probabilidad de que un niño sea varón es de 0.51. Sin considerar partos múltiples, calcular:

a) Prob. de que una pareja tenga tres niñas antes de tener un varón.

b) Ídem. antes de tener dos varones.

c) Nº medio de hijos de una pareja para tener dos varones.

Solución:

a) $P(MMMV) = (0.49)^3\,(0.51)$

b) Sea X= número de hijos hasta tener el segundo varón.

Los casos favorables son MMMVV, MMVMV, MVMMV, VMMMV. Cada uno de ellos tiene probabilidad de ocurrir $(0.49)^3\,(0.51)^2$.

Por tanto: $P(X=5)=4\,(0.49)^3\,(0.51)^2 = 0{,}12240$.

c) X sigue una distribución Binomial Negativa con p=0,51 y k=2

$$E\left[X\right] = \frac{k}{p} = \frac{2}{0,49} = 3.92$$

que es el número esperado de hijos hasta tener el segundo varón.

30. Por un punto de una autopista pasan coches aleatoriamente, siguiendo una distribución de Poisson, con una media de 6 coches por minuto.

a) Prob. de que transcurran más de 20 s desde que pasa un coche hasta que pasan 5 más.

b) Si un perro tarda 10 s en cruzar la autopista y empieza a cruzar inmediatamente después de que pase un coche, prob. de que lo atropellen.

Solución:

Sea X = nº de coches que pasan en un minuto, $X \in P(6)$.

a) Sea Y = nº de coches en 20 s, $Y \in P(2)$ y, por tanto:

$$P(Y \leq 4) = 0.9473$$

b) Sea Z = tiempo (min) desde que pasa un coche hasta que pasa el siguiente, $Z \in \text{Exp}(6)$ y, por tanto:

$$P(\text{atropellar al perro}) = P\left(Z \leq \frac{1}{6}\right) = \int_0^{1/6} 6e^{-6x}\,dx = \left[-e^{-6x}\right]_0^{1/6} = 1 - 0.367 = 0.632$$

31. La vida de las plantas de determinada especie sigue una distribución exponencial de parámetro r=1/120.

a) Calcular la función de distribución de la variable.

b) ¿Qué proporción de plantas mueren en los cien primeros días?

c) Si una cierta planta ha vivido ya cien días, ¿qué prob. hay de que viva al menos otros 50 días más?

Solución:

Sea X= tiempo de duración de la vida de una planta de la especie en estudio (días), $X \in \text{Exp}(1/120)$.

a) $$F(x) = P(X \leq x) = \int_0^x \lambda \cdot e^{-\lambda t}dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x},\ \forall x \in R^+$$

En nuestro caso, como $\lambda = 1/120$, sería $F(x) = \begin{cases} 1 - e^{-\frac{x}{120}} & si\ x \geq 0 \\ 0 & \text{en el resto} \end{cases}$

b) $P(X \leq 100) = F(100) = 1 - e^{-\frac{100}{120}} = 1 - 0.4346 = 0.5654$

c) $P(X \geq 150 / X > 100) = \dfrac{P(X \geq 150 \cap X > 100)}{P(X > 100)} = \dfrac{P(X \geq 150)}{P(X > 100)} =$

$$= \frac{1 - F(150)}{1 - F(100)} = \frac{e^{-\frac{150}{120}}}{e^{-\frac{100}{120}}} = \frac{0.2865}{0.4346} = 0.659$$

32. La vida media de cierto tipo de motor pequeño es de 10 años, con una desviación estándar de 2 años. El fabricante repone gratis todos los motores que se estropeen siempre que estén en periodo de garantía. Si piensa reponer solo el 3% de los motores que fallen, ¿cuánto tiempo debe durar la garantía? (Suponemos que la vida útil de los motores sigue una distribución normal).

Solución:

Sea X = duración en años de un motor, $X \in N(10\ ,\ 2)$;

Queremos calcular t para que:

$$P(X < t) = 0.03 \Rightarrow 0.03 = P\left(\frac{X - 10}{2} < \frac{t - 10}{2}\right) = P\left(Z < \frac{t - 10}{2}\right) \Rightarrow$$

$$\Rightarrow \frac{t - 10}{2} = -1.88 \Rightarrow t = 6.24$$

Debe ofrecerse como periodo de garantía seis años y 3 meses.

33. Al inspeccionar 1450 juntas de soldadura realizadas por una máquina de soldar, se encontraron 148 uniones defectuosas. Al soldar 5 uniones, ¿cuál es la prob. de obtener X piezas defectuosas? Calcular la media.

Solución:

$$P(\text{soldadura defectuosa}) = \frac{148}{1450}$$

Sea X = número de soldaduras defectuosas de las 5, $X \in B(5, 148/1450)$

$$P(X = x) = \binom{5}{x} \cdot \left(\frac{148}{1450}\right)^x \left(\frac{1302}{1450}\right)^{5-x}$$

$$E[X] = n \cdot p = 5 \cdot \frac{148}{1450} = 0.5103$$

34. Un fabricante vende un artículo a un precio fijo de 100 pts. Si el peso del artículo es inferior a 8 g no lo puede vender y representa una pérdida total. La distribución de los pesos es una normal N(μ, 1) y el coste de producción de cada artículo viene dado por c=5x + 30. Determinar el peso medio μ que haga máximo el beneficio esperado.

Solución:

Sea X = peso de una unidad, sabemos que $X \in N(\mu, 1)$.

Sea C = coste de una unidad=$5X + 30$.

Sea B el beneficio que es una v.a. y viene dado por:

$$B = \begin{cases} 100 - (5X + 30) & \text{si } X > 8 \\ -(5X + 30) & \text{si } X \leq 8 \end{cases}$$

B es una variable dicotómica y toma sus valores con probabilidad. p y $(1-p)$, respectivamente, siendo $p = P(X>8) = P(Z>8-\mu)$. Por tanto:

$$E[B] = \left[100 - (5\mu - 30)\right] p - (5\mu + 30)(1 - p) = 100p - 5\mu - 30$$

El problema está en que la expresión que nos da la esperanza de B depende de μ, pero también de p, por ello lo resolvemos por tanteo:

$$Si\ \ \mu = 9{,}5 \Rightarrow C = 5\mu + 30 = 77{,}5 \Rightarrow$$

$$\Rightarrow B = \begin{cases} 22{,}5 & p = 0{,}9332 \\ -77{,}5 & 1-p = 0{,}0668 \end{cases} \Rightarrow E[B] = 15{,}82$$

$$Si\ \ \mu = 10 \ \Rightarrow C = 5\mu + 30 = 80 \ \cdot\Rightarrow$$

$$\Rightarrow B = \begin{cases} 20 & p = 0{,}9772 \\ -80 & 1-p = 0{,}0228 \end{cases} \Rightarrow E[B] = 17{,}72$$

$$Si\ \ \mu = 10{,}5 \Rightarrow C = 5\mu + 30 = 82{,}5 \Rightarrow$$

$$\Rightarrow B = \begin{cases} 17{,}5 & p = 0{,}9938 \\ -82{,}5 & 1-p = 0{,}0062 \end{cases} \Rightarrow E[B] = 16{,}88$$

$$Si\ \ \mu = 10{,}1 \Rightarrow C = 5\mu + 30 = 80{,}5 \Rightarrow$$

$$\Rightarrow B = \begin{cases} 19{,}5 & p = 0{,}9821 \\ -80{,}5 & 1-p = 0{,}0179 \end{cases} \Rightarrow E[B] = 17{,}71$$

$$Si\ \ \mu = 9{,}9 \Rightarrow C = 5\mu + 30 = 79{,}5 \Rightarrow$$

$$\Rightarrow B = \begin{cases} 20{,}5 & p = 0{,}9713 \\ -79{,}5 & 1-p = 0{,}0287 \end{cases} \Rightarrow E[B] = 17{,}63$$

Vemos que la aproximación con μ = 10 maximiza la esperanza del beneficio.

35. El peso de los botes de conserva fabricados en cierta industria se distribuye normalmente. Se han fabricado 4000 botes en un mes, de los cuales 800 pesaron menos de 1 kg y 1000 más de 2 kg. Hallar la media y la desviación típica de la distribución.

Solución:

Sea X = peso de un bote, $X \in N(\mu\ ,\ \sigma)$.

$$\left.\begin{aligned} \mathrm{P}(\mathrm{X} < 1) = \mathrm{P}\left(\frac{X-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right) = \frac{800}{4000} = 0.2 \Rightarrow \frac{1-\mu}{\sigma} = -0.846 \\ P(X > 2) = P\left(\frac{x-\mu}{\sigma} > \frac{2-\mu}{\sigma}\right) = \frac{1000}{4000} = 0.25 \Rightarrow \frac{2-\mu}{\sigma} = 0.674 \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left\{ \begin{array}{l} \mu = 1 + 0{,}846\,\sigma \\ \mu = 2 - 0{,}674\,\sigma \end{array} \right\} \Rightarrow (r.m.m.) \quad 0 = -1 + 1{,}52\,\sigma \Rightarrow$$

$$\Rightarrow \sigma = \frac{1}{1{,}52} = 0{,}6579 \Rightarrow \mu = 1 + 0{,}846 \cdot 0{,}6579 = 1{,}556$$

36. En una situación de selección de personal un psicólogo utiliza un test y pone como norma aceptar a los sujetos que se aparten más de 4 pero menos de seis puntos de la media de dicho test. Sabiendo que la distribución de las puntuaciones es normal, que la varianza es 16 y que el nº de sujetos elegidos resulta ser 58. Se pide:

a) Hallar el número de sujetos que realizó el test.

b) ¿Será seleccionado el sujeto que obtuvo el percentil 67? Justificar la respuesta.

Solución:

Sea X = puntuación de un test, $X \in N(\mu\,,4)$.

a) Queremos que $P\left(\mu + 4 < X < \mu + 6\right) = \frac{29}{n}$, pero

$$P(1 < \frac{X-\mu}{4} < 1.5) = P(1 < Z < 1.5) = 0.1587 - 0.0668 = 0.0919 \;\Rightarrow$$

$$\Rightarrow \; 0.0919 = \frac{29}{n} \;\Rightarrow\; n = \frac{29}{0.0919} = 315.56 \cong 316$$

b) Buscamos la puntuación obtenida por el sujeto que obtuvo el percentil 67.

$$P(X < X_{67}) = P\left(Z < \frac{X_{67} - \mu}{4} \right) = 0.67 \Rightarrow \frac{X_{67} - \mu}{4} = 0.44 \Rightarrow$$

$$\Rightarrow X_{67} = \mu + 4 \cdot 0.44 = \mu + 1.76$$

El sujeto que obtuvo el percentil 67 no será seleccionado porque su puntuación dista de la media 1,76 puntos, que no es un número comprendido entre 4 y 6.

37. En una centralita se reciben un promedio de cinco llamadas entre las 9:00 y las 10:00 horas, recibiéndolas al azar. Calcular:

a) Probabilidad de que se reciban una o más llamadas entre las 9:00 y las 10:00 horas de un día determinado.

b) Probabilidad de recibir exactamente 2 llamadas entre las 9:00 h y las 9:12 h.

c) Probabilidad de que durante una semana de 5 días, haya exactamente dos días en los que entre las 9:00 h y las 9:12 h no se reciba ninguna llamada.

Solución:

a) Sea X = nº de llamadas entre las 9:00 y las 10:00 de cierto día, por el contexto suponemos que $X\in P(\lambda)$ con $\lambda=5$.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{5^0}{0!}e^{-5} = 1 - 0.0067 = 0.9933$$

b) Sea Y = nº de llamadas entre las 9:00 h y las 9:12, que según sabemos será $Y\in P(1)$, puesto que 12 minutos es una quinta parte de una hora. Entonces:

$$P(Y = 2) = \frac{1^2}{2!}e^{-1} = 0.1839$$

c) Sea Z = nº de días de los cinco de la semana en que no se recibe ninguna llamada, $Z\in B(5\ ,p)$ con $p = P(Y=0)$.

$$p = P(Y = 0) = \frac{1^0}{0!}e^{-1} = 0.3678$$

$$P(Z = 2) = C_5^2 \cdot (0.3678)^2 (1 - 0.3678)^3 =$$

$$= \frac{5 \cdot 4}{2} \cdot (0.3678)^2 \cdot (0.6322)^3 = 0{,}04624$$

38. En un laboratorio se dispone de una población de ratas. Se definen dos variables aleatorias X e Y de la siguiente forma:

$$X = \begin{cases} 0 & \text{si el animal no ha recibido entrenamiento.} \\ 1 & \text{si el animal ha recibido entrenamiento.} \end{cases}$$

Y = nº de éxitos cuando se le hace realizar cierta tarea dos veces consecutivas.

La función de distribución conjunta viene dada por la tabla siguiente:

X\Y	0	1	2
0	0.2	0.4	0.5
1	0.25	0.75	1

a) Calcular la prob. de que una rata elegida al azar haya recibido entrenamiento.
b) Calcular la prob. de que una rata elegida al azar haya obtenido algún acierto al realizar la tarea dos veces consecutivas.
c) Calcular la Cov(X,Y) y estudiar la independencia de las variables.

Solución:

A partir de la tabla de distribución conjunta acumulada, que nos da $F(x_i, y_j)$, escribimos la de probabilidad conjunta, teniendo en cuenta que:

$$F(x_i \ , \ y_j) = \sum_{\substack{x_r \leq x_i \\ y_s \leq y_j}} P(X = x_r \ , \ Y = y_s)$$

X\Y	0	1	2	total
0	0,2	0,2	0,1	0,5
1	0,05	0,3	0,15	0,5
total	0,25	0,5	0,25	

a) $P(\text{entrenamiento}) = P(X = 1, Y = 0,1,2) = 0.5$

b) $P(\text{algún acierto}) = P(X = 0,1, Y = 1,2) = 0,5 + 0,25 = 0.75$

c) $COV(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y]$

$$E[X] = 0 \cdot P(X=0) + 1 \cdot P(X=1) = 0.5;$$

$$E[Y] = 0 \cdot P(Y=0) + 1 \cdot P(Y=1) + 2 \cdot P(Y=2) = 0,5 + 2 \cdot 0,25 = 1$$

$$E[XY] = 0 \cdot P(XY=0) + 1 \cdot P(XY=1) + 2 \cdot P(XY=2) = 0,6$$

$$\Rightarrow COV[X,Y] = 0,6 - 0,5 = 0,1$$

Las variables X e Y no son independientes porque su covarianza es distinta de cero.

39. Consideremos dos distribuciones normales con desviaciones típicas de 5 unidades. Una de ellas tiene media -8. Sabiendo que existe un punto A que deja a su derecha un 10% del área bajo la densidad que tiene por media -8 y a su izquierda un 15% del área bajo la densidad de la otra. Calcular la media desconocida.

Solución:

Sean $X \in N(-8, 5)$ e $Y \in N(\mu, 5)$ las dos variables. Tenemos:

$$P(X > A) = P\left(\frac{X+8}{5} > \frac{A+8}{5}\right) = P\left(Z > \frac{A+8}{5}\right) = 0.1 \Rightarrow$$

$$\Rightarrow \frac{A+8}{5} = 1.282 \Rightarrow A = -1.59$$

$$P(Y < A) = P\left(\frac{Y-\mu}{5} < \frac{A-\mu}{5}\right) = P\left(Z < \frac{-1.59-\mu}{5}\right) = 0.15 \Rightarrow$$

$$\Rightarrow \frac{-1.59-\mu}{5} = -1.037 \Rightarrow \mu = 3.595$$

40. Una licorería da servicio tanto a clientes que llegan en automóvil como a los que llegan caminando. En un día elegido al azar sean X e Y, respectivamente, las proporciones de tiempo en que se utilizan las instalaciones para atender a los automovilistas y a los que llegan caminando, y su función de densidad conjunta la siguiente:

$$f(x,y)=\begin{cases}\dfrac{2(x+2y)}{3} & \text{si } \ 0\le x\le 1,\ 0\le y\le 1\\ 0 & \text{en otro caso}\end{cases}$$

a) Calcular la densidad marginal de X.

b) Calcular la densidad marginal de Y.

c) Calcular la probabilidad de que las instalaciones para atender a los automovilistas estén ocupadas menos de la mitad del tiempo.

Solución:

a) $f_1(x)=\int_{y=0}^{1}\frac{2}{3}(x+2y)\,dy=\frac{2}{3}\left[xy+y^2\right]_{y=0}^{1}=\frac{2}{3}(x+1)$

b) $f_2(y)=\int_{x=0}^{1}f(x,y)\,dx=\int_{x=0}^{1}\frac{2}{3}(x+2y)\,dx=\frac{2}{3}\left[\frac{x^2}{2}+2xy\right]_0^1=\frac{2}{3}\left(\frac{1}{2}+2y\right)$

c) $P(X\le 0.5,\ \ 0\le Y\le 1)=\int_{x=0}^{0.5}\int_{y=0}^{1}f(x,\ \ y)\,dxdy=\int_{x=0}^{0.5}\int_{y=0}^{1}\frac{2}{3}(x+2y)\,dxdy=$

$$=\int_{x=0}^{0.5}\frac{2}{3}(x+1)\,dx=\left[\frac{x^2}{3}+\frac{2x}{3}\right]_{x=0}^{0.5}=\frac{1}{12}+\frac{1}{3}=\frac{5}{12}$$

41. El peso neto de una lata de conservas es aleatorio, con una distribución uniforme entre 485 y 535 gramos. Si el producto sale al mercado garantizando un peso neto de por lo menos medio kilo, ¿cuál es la probabilidad de que al inspeccionar seis latas, se detecte que más de una no cumple la garantía?

Solución:

Sea X = peso de una lata de conservas, $X\in U(485\ ,\ 535)$

$$f(x)=\begin{cases}\dfrac{1}{50} & \text{si} \quad 485 \le x \le 535 \\ 0 & \text{en el resto}\end{cases}$$

$$P(X<500)=\int_{485}^{500}\frac{1}{50}\,dx=\left[\frac{x}{50}\right]_{485}^{500}=\frac{15}{50}=0.3$$

Sea Y = nº de latas que pesan menos de ½ Kg de las 6 inspeccionadas, $Y\in B(6\,,\,0{,}3)$ y tenemos:

$$P(Y\ge 2)=1-P(Y=0)-P(Y=1)=1-0{,}7^6-6\cdot 0{,}3\cdot 0{,}7^5=0{,}58$$

42. La variable aleatoria bidimensional (X,Y) es tal que:

a) X tiene distribución exponencial de parámetro δ.

b) Supuesto que X = x, Y sigue una distribución exponencial de parámetro x.

Calcular la distribución de Y.

Solución:

Queremos calcular $f_2(y)$, que viene dada por:

$$f_2(y)=\int_0^{+\infty} f(x,y)\,dx$$

Tenemos $\begin{cases} f_1(x)=\delta\cdot e^{-\delta x} \\ f(y/X=x)=x\cdot e^{-xy}\end{cases}$

$$f(y/X=x)=\frac{f(x,y)}{f_1(x)}\Rightarrow f(x,y)=f(y/X=x)\cdot f_1(x)=\delta x e^{-(\delta+y)x}$$

Por tanto:

$$f_2(y) = \int_0^{+\infty} \delta x e^{-(\delta+y)x} dx = * \quad \left\{ \begin{array}{l} \delta x = u \Rightarrow \delta dx = du \\ e^{-(\delta+y)x} dx = dv \Rightarrow v = -\dfrac{1}{\delta + y} e^{-(\delta+y)x} \end{array} \right\}$$

Por tanto:

$$* = -\frac{\delta}{\delta + y} \cdot \left[x e^{-(\delta+y)x} \right]_0^{\infty} + \int_0^{\infty} \frac{\delta}{\delta + y} e^{-(\delta+y)x} dx =$$

$$= -\frac{\delta}{(\delta + y)^2} \left[e^{-(\delta+y)x} \right]_0^{\infty} = \frac{\delta}{(\delta + y)^2}$$

43. Como parte de un experimento se decide inyectar a cada ratón de una muestra aleatoria de 25, con un fármaco cuya dosificación es de 0.004 mg/g de peso. Se sabe que el peso de esta cepa de ratones se distribuye normalmente con una media de 19 g y una desviación típica de 4 g.

a) Si el investigador consigue para sus pruebas un total de 2 mg del fármaco, ¿qué probabilidad hay de quedarse sin él?

b) ¿De qué cantidad de fármaco debería disponer para que el riesgo de quedarse sin él sea inferior al 1%?

Solución:

Sea X_i = peso del ratón i-ésimo (gramos), $X_i \in$ N(19 , 4), $\forall i$ = 1,...25.

Sea Y = peso de los 25 ratones (gramos) = $\sum X_i$, y por tanto, según sabemos

$$Y \in N\left(25 \cdot 19, 4\sqrt{25}\right) = N(475 \ , \ 20)$$

a) $$P(0.004Y \geq 2) = P(Y \geq \frac{2}{0.004}) = P(Y \geq 500) = P\left(Z \geq \frac{500 - 475}{20}\right) =$$

$$= P(Z \geq 1.25) = 0.1056$$

Algo más del 10% de las veces los 2 mg de fármaco no serán suficientes para inyectar a los 25 ratones.

b) $P(0.004Y \geq A) \cong 0.01 \Rightarrow$

$$\Rightarrow 0.01 = P(Y \geq \frac{A}{0.004}) = P(Z \geq \frac{\frac{A}{0.004} - 475}{20}) \Rightarrow$$

$$\Rightarrow \frac{\frac{A}{0.004} - 475}{20} = 2.326 \Rightarrow A = (475 + 20 \cdot 2.326) \times 0.004 = 2,0864$$

Por tanto debería tener 2,0864 mg del fármaco para lograr que el riesgo de quedarse sin él fuera inferior al 1%.

44. Se lanzan un dado y una moneda al aire, y se consideran las variables aleatorias:

$$X = \begin{cases} 1 & \text{si sale cara} \\ -1 & \text{si sale cruz} \end{cases} \qquad Y = \begin{cases} 1 & \text{si sale impar} \\ 2 & \text{si sale par} \end{cases}$$

Hallar la función de probabilidad, media, varianza y desviación típica de las variables:

a) X+Y.

b) XY.

c) Y/X (Y condicionada a X).

Solución:

a) $P(X+Y=0) = P(X=-1, Y=1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(X+Y=1) = P(X=-1, Y=2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(X+Y=2) = P(X=1, Y=1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(X+Y=3) = P(X=1, Y=2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$$E[X+Y]=\frac{1}{4}\cdot(0+1+2+3)=\frac{6}{4}=1.5$$

$$V[X+Y]=E\left[(X+Y)^2\right]-\left(E[X+Y]\right)^2=\frac{1}{4}(0+1+4+9)-1,5^2=1.25$$

$$\sigma_{X+Y}=\sqrt{\frac{5}{4}}=\frac{\sqrt{5}}{2}$$

b) $P(X\cdot Y=-2)=P(X=-1,Y=2)=\frac{1}{4}$

$$P(X\cdot Y=-1)=P(X=-1,Y=1)=\frac{1}{4}$$

$$P(X\cdot Y=1)=P(X=1,Y=1)=\frac{1}{4}$$

$$P(X\cdot Y=2)=P(X=1,Y=2)=\frac{1}{4}$$

$$E[X\cdot Y]=\frac{1}{4}\cdot(-2-1+1+2)=0$$

$$V[X\cdot Y]=E\left[(XY)^2\right]-\left(E[XY]\right)^2=\frac{10}{4}=2.5$$

$$\sigma_{X\cdot Y}=\sqrt{\frac{10}{4}}=\frac{\sqrt{10}}{2}$$

c) X e Y son dos v. a. claramente independientes, por tanto la distribución de probabilidad de Y condicionada a X es la distribución de Y, por tanto:

$P(Y=1)=P(Y=2)=0.5;$

$E[Y]=0.5(1+2)=1.5;$

$V[Y]=0.5(1+4)-(1.5)^2=0.25;\quad \sigma_Y=\sqrt{0.25}=0.5$

45. Los accidentes de trabajo que se producen en una fábrica por semana, siguen una ley de Poisson tal que la probabilidad de que haya 5 accidentes es 16/15 de la probabilidad de que haya 2 accidentes. Calcular:

a) El parámetro de la ley de Poisson.
b) Nº máximo de accidentes semanales, con seguridad del 90%.
c) Probabilidad de que no haya ningún accidente en 4 semanas.

Solución:

Sea X = número de accidentes por semana, $X \in P(\lambda)$ y nos dicen que:

$$P(X=5)=\frac{16}{15}P(X=2)$$

a) $$\left.\begin{array}{l} P(X=5)=\dfrac{\lambda^5}{5!}e^{-\lambda} \\ P(X=2)=\dfrac{\lambda^2}{2!}e^{-\lambda} \end{array}\right\} \Rightarrow \frac{\lambda^5}{5!}e^{-\lambda}=\frac{16}{15}\frac{\lambda^2}{2!}e^{-\lambda} \Rightarrow 30\lambda^2\left(\lambda^3-64\right)=0 \Rightarrow$$

$$\Rightarrow \lambda=4$$

b) $$P(X\le k)=\sum_0^k \frac{\lambda^k}{k!}e^{-\lambda}=0.9 \Rightarrow k\cong 7$$

c) Sea Y = número de semanas de las cuatro siguientes en que no hay ningún accidente, $Y\in B(4\,,p)$ siendo

$$p=P\left(X=0\right)=\frac{\lambda^0}{0!}e^{-\lambda}=e^{-4}=0.0183$$

$$P(Y=4)=\binom{4}{4}(0.0183)^4=1.1215\cdot 10^{-7}$$

46. Una moneda es lanzada hasta que aparece dos veces seguidas el mismo resultado. Describir el espacio muestral y hallar la probabilidad de que:
a) El experimento termine antes del sexto lanzamiento.
b) El experimento termine en un lanzamiento par.

Solución:

El espacio muestral sería:

$$\{CC,\ ++,\ +CC,\ C++, C+CC,\ +C++, \ldots\ldots,$$
$$(+C)..\overset{k}{}..(+C)++,\ (C+)....\overset{k}{}..(C+)CC, \ldots\ldots,$$
$$(C+).....\overset{k}{}...(C+)+,\ (+C)..\overset{k}{}..(+C)C, \ldots\ldots\ \}$$

a) Sea X = número de tiradas hasta concluir el experimento

$$P(X<6) = P(X=2) + P(X=3) + P(X=4) + P(X=5) =$$
$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{30}{32}$$

b) $P(X = \text{n}^\circ \text{ par}) = \sum_{k=1}^{\infty} P(X=2k) = \sum_{k=1}^{\infty} 2\frac{1}{2^{2k}} = \sum_{k=1}^{\infty} \frac{1}{2^{2k-1}} =$

$$= \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7} + \ldots = \frac{\frac{1}{2}}{1-\frac{1}{4}} = \frac{2}{3}$$

47. Para obtener el permiso de conducir se realiza un test con 20 ítems. Se sabe que una determinada persona tiene una probabilidad 0.8 de contestar bien cada ítem. Calcular:

a) Probabilidad de que la primera pregunta que contesta bien sea la tercera que hace.

b) Probabilidad de que apruebe al contestar el doceavo ítem, sabiendo que para aprobar el test es necesario contestar 10 ítems bien.

Solución:

a) Si la primera pregunta que contesta bien es la 3ª que contesta, el suceso que ha ocurrido es MMB, por tanto:

$$P\ (\text{MMB}) = 0.2^2 \cdot 0.8 = 0.032$$

b) Para que apruebe al contestar el duodécimo ítem, debe ocurrir que conteste dicho ítem correctamente y que entre los 11 anteriores que contesta, 9 estén correctos. Un caso favorable sería BBBBBBBBBMMB cuya probabilidad de ocurrencia es de $0.8^{10}0.2^{2}$. Cada uno de los casos favorables tiene esa misma probabilidad de ocurrir, basta contar cuántos son. Por tanto:

$$P(\text{apruebe al contestar el duodécimo ítem}) = \binom{11}{2}\cdot(0.8)^{10}\cdot(0.2)^{2} = 0{,}02536$$

48. Un abogado se traslada diariamente de su casa en las afueras a su oficina en el centro de la ciudad. En promedio el viaje dura 24 minutos con una desviación estándar de 3.8 minutos y suponemos que los tiempos de traslado siguen una distribución normal.

a) Probabilidad de que cierto día tarde al menos media hora en llegar a la oficina.

b) Si la oficina abre a las 9:00 a.m. y él sala de su casa a las 8:45 a.m. diariamente, ¿qué porcentaje de días llega tarde?

c) Si sale de su casa a las 8:35 a.m. y en la oficina se sirve un café entre las 8:45 y las 9:00 ¿cuál es la probabilidad de que se pierda el café?

d) Encontrar el tiempo por encima del cual duran el 15% de los traslados más lentos.

e) Encontrar la probabilidad de que 2 de los 3 siguientes traslados duren menos de media hora.

Solución:

Sea X = tiempo (minutos) que tarda en llegar a su oficina, $X \in$ N(24 , 3,8):

a) $P(X \geq 30) = P(\dfrac{X - 24}{3.8} \geq \dfrac{30-24}{3.8}) = P(Z \geq 1.5789) = 0.0580$

b) $P(X \geq 15) = P\left(\dfrac{X-24}{3.8} \geq \dfrac{15-24}{3.8}\right) = P(Z > -2.3684) =$

$= 1 - 0.0089 = 0.9911$

c) $P(X \geq 25) = P\left(\frac{X-24}{3.8} \geq \frac{25-24}{3.8}\right) = P(Z > 0.2632) = 0.3962$

d) $P(X \geq T_L) = 0.15 \Rightarrow P\left(Z \geq \frac{T_L - 24}{3.8}\right) = 0.15$

$\Rightarrow T_L = 24 + 3.8 \cdot 1.036 = 27.94$ min.

c) Sea Y = número de traslados de los 3 siguientes que duran menos de 30 minutos, Y$\in$ B(3, p), siendo

$$p = 1 - P(X \geq 30) = 1 - 0{,}058 = 0{,}942 \text{ según se calculó en a).}$$

$$P(Y = 2) = \binom{3}{2} \cdot p^2 \cdot (1-p) = 3 \cdot 0.058 \cdot 0.942^2 = 0.1544$$

49. Una moneda se lanza dos veces. Sea Z el número de caras en el primer lanzamiento y W el número total de caras en los dos lanzamientos. La moneda no está equilibrada, siendo la probabilidad de cara 0.4, encontrar:

a) La distribución de probabilidad conjunta de W y Z.

b) La distribución marginal de W.

c) La distribución marginal de Z.

d) La probabilidad de que ocurra al menos una cara.

Solución:

El conjunto de los posibles valores de (Z,W) es { (0,0), (0, 1), (1,1), (1,2)}

a) $P(Z=0, W=0) = P(++) = 0.6^2 = 0.36$

$P(Z=0, W=1) = P(+C) = 0.6 \cdot 0.4 = 0.24$

$P(Z=1, W=1) = P(C+) = 0.4 \cdot 0.6 = 0.24$

$P(Z=1, W=2) = P(CC) = 0.4^2 = 0.16$

b) $P(W=0) = 0.36$

$P(W=1) = 0.48$

$P(W=2) = 0.16$

c) $P(Z=0)=0.6$

$P(Z=1)=0.4$

d) P(al menos una cara) = 1- P(ninguna cara)= 1 - $P(Z = 0, W = 0)$ = 1 - 0.36 = 0.64.

50. La longitud de una cierta pieza se distribuye en el intervalo [1,3] con función de densidad f(x) = k (x-1) (3-x) con k constante y f(x) =0 en el resto. Sólo es válida una pieza si su longitud está comprendida entre 1.7 y 2.4.

a) ¿Cuál es la probabilidad de que una determinada pieza sea útil?

b) Si las piezas se empaquetan en lotes de 5 unidades, y se acepta el lote si contiene menos de dos piezas defectuosas, ¿cuál es la probabilidad de que un cierto lote sea rechazado?

Solución:

Sea X = longitud de una pieza, nos dicen que su función de densidad es:

$$f(x)=\begin{cases} k(x-1)(3-x) & \text{si } 1 \le x \le 3 \\ 0 & \text{en el resto} \end{cases}$$

Tenemos que calcular k, sabiendo que se debe cumplir: $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\int_1^3 k(-x^2+4x-3)dx = k\left[-\frac{1}{3}x^3+2x^2-3x\right]_1^3 =$$

$$= k\left[(-9+18-9)-\left(-\frac{1}{3}+2-3\right)\right] = k\frac{4}{3} = 1 \Rightarrow k = \frac{3}{4}$$

a) $P(1.7 \le x \le 2.4) = \int_{1.7}^{2.4} k(-x^2+4x-3)dx = k\left[-\frac{x^3}{3}+2x^2-3x\right]_{1.7}^{2.4} =$

$$= 0.75\left[(-4.608+11.52-7.2)-(-1.6376+5.78-5.1)\right] =$$
$$= 0.75 \cdot 0.6696 = 0.5022 \cong 0.5$$

b) Sea Y = número de piezas útiles de las 5 del lote, $Y \in B(5\ , p)$ con p = 0,502.

$$P(\text{rechazar un lote}) = P(Y > 3) = 1 - P(Y \leq 3) = 1 - \sum_{k=0}^{3} \binom{5}{k} 0.502^{k} 0.498^{5-k} =$$
$$= 1 - 0.8125 = 0.1875$$

51. Se informó a dos estudiantes que habían recibido puntuaciones estándar de 0.8 y -0.4, respectivamente, en una prueba de inglés. Si sus puntuaciones fueron 88 y 64, respectivamente, hallar la media y la desviación típica de las puntuaciones de esa prueba, sabiendo que dichas puntuaciones siguen una distribución normal.

Solución:

Sea X = puntuación en la prueba de inglés, $X \in N(\mu, \sigma)$, la puntuación estándar de un alumno que sacó la nota X vendrá dada por $\dfrac{X-\mu}{\sigma}$ y por tanto:

$$\begin{cases} \dfrac{88-\mu}{\sigma} = 0.8 \\ \dfrac{64-\mu}{\sigma} = -0.4 \end{cases} \Rightarrow \begin{cases} \mu = 88 - 0.8\sigma \\ \mu = 64 + 0.4\sigma \end{cases} \Rightarrow 1.2\sigma = 24 \Rightarrow \quad \sigma = 20 \quad \text{y} \quad \mu = 72$$

52. Cinco personas lanzan simultáneamente monedas, para determinar quién ha de comprar los refrescos para todos. El sistema es el siguiente: el primero que obtenga un resultado (ya sea cara o cruz) distinto de cada uno de los resultados obtenidos por los demás, debe pagar los refrescos de todos. Sea X el número de ensayos requeridos para concluir el juego. Se pide:

a) Determinar su función de probabilidad.

b) Determinar su función de distribución.

c) Calcular P(X>2), P(X<2) y P(3≤X≤6).

d) Calcular P(X>5 / X>2).

e) Calcular E[X] y V(X).

Solución:

a) Sea X= número de ensayos para terminar el juego, X sigue una distribución geométrica G(p) siendo que el juego consiste en lanzar 5 monedas equilibradas repetidamente y termina cuando exactamente uno de los resultados es distinto de los otros cuatro, y por tanto p = P(haya perdedor).

Sea Y= número de caras cuando se lanzan las cinco monedas, $Y \in B(5\ ,\ 0{,}5)$.

$$p = P(\text{haya perdedor}) = P(Y=1) + P(Y=4) = \frac{5}{2^5} + \frac{5}{2^5} = \frac{5}{2^4}$$

$$\Rightarrow P(X=k) = (1-p)^{k-1}\, p = \left(\frac{11}{2^4}\right)^{k-1}\left(\frac{5}{2^4}\right) = \frac{5\times 11^{k-1}}{2^{4k}}$$

b)
$$F(t) = P(X \le t) = \begin{cases} 0 & si \quad t<1 \\ 5\sum_{1}^{k} \frac{11^{k-1}}{2^{4k}} & si \quad k \le t < k+1, \quad k=1,\cdots,4 \\ 1 & si \quad t \ge 5 \end{cases}$$

c) $$P(X>2) = 1 - P(X \le 2) = 1 - F(2) = 1 - 5\frac{1}{2^4} - 5\frac{11}{2^8} =$$
$$= 1 - \frac{5}{2^4} - \frac{55}{2^8} = \frac{2^8 - 5\cdot 2^4 + 5\cdot 11}{2^8} = 0{,}4336$$

$$P(X<2) = P(X=1) = \frac{5}{2^4}$$

$$P(3 \le X \le 6) = \sum_{i=3}^{6} P(X=i) = 5\left(\frac{11^2}{2^{12}} + \frac{11^3}{2^{16}} + \frac{11^4}{2^{20}} + \frac{11^5}{2^{24}}\right) = 0{,}367$$

d) $$P(X>5/X>2) = \frac{P(X>5)}{P(X>2)} = \frac{1-P(X\le 5)}{1-P(X \le 2)} =$$

$$= \frac{1 - \frac{5}{2^4}\left(1 + \frac{11}{2^4} + \frac{11^2}{2^8} + \frac{11^3}{2^{12}} + \frac{11^4}{2^{16}}\right)}{1 - \frac{5}{2^4}\left(1 + \frac{11}{2^4}\right)} = 0{,}325$$

e) $E[X]=\frac{1}{p}=\frac{2^4}{5}=\frac{16}{5}=3,2$ $\qquad V[X]=\frac{1-p}{p^2}=\frac{\frac{11}{2^4}}{\frac{5^2}{2^8}}=\frac{176}{25}=7,04$

53. En un cierto hospital se comprobó que el peso en kilos de los niños al nacer era una variable aleatoria cuya función de densidad es:

$$f(x)=\begin{cases} kx & si\ 2\le x\le 4 \\ 0 & en\,otro\,caso \end{cases}$$

Se pide:

a) Hallar k para que f(x) sea función de densidad.

b) Hallar la Función de distribución, media y varianza.

c) Probabilidad de que un niño elegido al azar pese más de 3 kilos.

d) Probabilidad de que pese entre 2 y 3.5 kilos.

e) ¿Cuánto debe pesar un niño para tener inferior o igual a su peso al 90% de los niños?

Solución:

a) Para que sea función de densidad se debe cumplir:

$$\int_{-\infty}^{+\infty} f(x)dx=1,$$

$$\int_2^4 kxdx=\frac{k}{2}\left[x^2\right]_2^4=\frac{k}{2}(16-4)=6k=1\Leftrightarrow k=\frac{1}{6}$$

b)

$$F(x)=P(X\le x)=\begin{cases} 0 & si\,x<2 \\ \frac{(x^2-4)}{12} & si\,2\le x<4 \\ 1 & si\,x>4 \end{cases}$$

$$E[X]=\int_2^4 \frac{x^2}{6}dx=\frac{1}{18}\left[x^3\right]_2^4=\frac{64-8}{18}=\frac{28}{9}$$

$$V[X]=E\left[X^2\right]-(E[X])^2=\frac{4^4-2^4}{24}-\left(\frac{28}{9}\right)^2=0,3209$$

c) $P(X>3)=1-F(3)=1-\frac{1}{12}(3^2-4)=1-\frac{5}{12}=\frac{7}{12}$

d) $P(2\leq X\leq 3,5)=F(3,5)-F(2)=\frac{\left(3,5^2-4\right)-\left(2^2-4\right)}{12}=0,69$

e) $P(X\leq A)=0,9\Rightarrow F(A)=0,9\Rightarrow \frac{A^2-4}{12}=0,9\Rightarrow$

$$\Rightarrow A^2=14,8\Rightarrow A=\sqrt{14,8}=3,85 \text{ kg}$$

54. Los empresarios de una fábrica de automóviles deciden otorgar un premio entre los distribuidores si venden 320 automóviles o más por día. El número de automóviles vendidos al día por los distribuidores A y B está normalmente distribuido de la forma siguiente:

Distribuidor	Media	Desviación Típica
A	290 automóviles	20 automóviles
B	300 automóviles	10 automóviles

Se pide:

a) ¿Qué porcentaje de los días obtendrá premio el distribuidor A?

b) ¿Qué porcentaje de los días obtendrá premio el distribuidor B?

c) ¿A quién de los dos distribuidores beneficia la decisión de la empresa?

d) Si se asociaran los dos distribuidores A y B, ¿qué porcentaje de los días obtendrían premio?

Solución:

a) $P(X \geq 320) = P(Z \geq \frac{320 - 290}{20}) = P(Z \geq 1.5) = 0.0668$

El 6,68% de los días el distribuidor A obtendrá premio.

b) $P(Y \geq 320) = P\left(Z \geq \frac{320 - 300}{10}\right) = P(Z \geq 2) = 0.0228$

El 2,28% de los días el distribuidor B obtendrá premio.

c) La decisión de la empresa beneficia al distribuidor A; B casi nunca llega a superar el mínimo de los 320 coches al día.

d) Suponemos que se trata de que se asocien pero ante la empresa siguen figurando como dos vendedores de manera que en el cómputo de ventas de cada uno de ellos siempre figura la mitad de lo que hayan vendido entre los dos, la misma cantidad para ambos. En tal caso, la variable será:

$$X + Y \in N\left(590, \quad \sqrt{20^2 + 10^2}\right) = N(590, \quad 22.36)$$

$$P\left(\frac{X+Y}{2} \geq 320\right) = P(X + Y \geq 640) = P\left(Z \geq \frac{640 - 590}{22,36}\right) =$$
$$= P(Z \geq 2,23) = 0,0132$$

Si se asocian sólo el 1,32% de los días obtendrán premio.

55. En el niquelado de ciertas láminas metálicas se producen desperfectos que se distribuyen aleatoriamente sobre toda la superficie niquelada. El nº de defectos por lámina es una v. a. de Poisson de media 2.

a) Si se eligen al azar 10 láminas de la cadena de montaje, ¿cuál es la probabilidad de que exactamente 3 tengan defectos?

b) Se toman láminas de la cadena de montaje hasta obtener 7 sin defectos, ¿cuál es la probabilidad de que haya habido que elegir 10 láminas?

c) ¿Qué porcentaje de láminas tendrá 4 defectos o más?

Solución:

a) Sea X= número de defectos en una lámina, $X\in$ P(2).

Sea Y = número de láminas con algún defecto de las 10, $Y\in B(10,p)$, siendo:

$$p = P(X\geq 1) = 1-P(X=0)=1-0,136 = 0,864$$

Y por tanto:

$$P(Y=3)=\binom{10}{3}0,864^3 0.135^7 = 63\times 10^{-6}$$

b) Sea Z = número de la extracción que nos da la 7ª sin defectos, $Z\in$ BN(7, 0,135).

$$P(Z=10)=\binom{9}{6}0,135^3 0,864^7 = 0,07$$

c) $$P(X\geq 4)=1-P(X\leq 3)=1-e^{-2}\left(\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}\right)=0,1428$$

El porcentaje de láminas con 4 defectos ó más es 14,28%.

56. El tiempo, en horas, empleado diariamente en transporte por los trabajadores de cierta ciudad, sigue una distribución Gamma con parámetros a=2, p=2.

a) Calcular la probabilidad de que un trabajador emplee más de media hora en transporte.

b) Calcular el tiempo mínimo que emplea el 50% de los trabajadores que más tiempo pierden en transporte.

Solución:

Sea X= tiempo en horas que emplea un trabajador en llegar al trabajo, $X\in$ G(2,2) y su densidad vendrá dada por:

$$f(x)=\frac{a^p}{\Gamma(p)}x^{p-1}e^{-ax}=\frac{2^2}{\Gamma(2)}xe^{-2x}=\frac{4}{1!}xe^{-2x}=4xe^{-2x}\text{ para } x>0$$

$$\text{a) } P(X\geq 0,5)=\int_{0,5}^{+\infty}4xe^{-2x}dx=2\int_{0,5}^{+\infty}2xe^{-2x}dx=(*)$$

$$=2\left[-xe^{-2x}\right]_{0,5}^{+\infty}+2\int_{0,5}^{+\infty}e^{-2x}dx=e^{-1}+\left[-e^{-2x}\right]_{0,5}^{+\infty}=2e^{-1}$$

$$(*)\begin{cases} x=u \Rightarrow dx=du \\ 2e^{-2x}dx=dv \Rightarrow v=-e^{-2x}\end{cases}$$

$$\text{b) } 0,5=P(X\geq A)=\int_{A}^{+\infty}4xe^{-2x}dx=2\left[-xe^{-2x}\right]_{A}^{+\infty}+\left[-e^{-2x}\right]_{A}^{+\infty}=$$

$$=Ae^{-2A}+e^{-2A}=e^{-2A}(1+A)\Rightarrow -2A+Ln(1+A)=Ln0,5\Rightarrow$$

$$\Rightarrow A=0,8$$

El 50% de los trabajadores que más tardan en llegar al trabajo invierte a partir de unos 50 minutos.

57. La empresa Olis, encargada de la fabricación y el mantenimiento de ascensores, desea mejorar la calidad de sus productos y el servicio al cliente. A tal fin, se lleva a cabo un análisis de la situación actual del que se concluye que la distribución conjunta de la variable aleatoria X, tiempo transcurrido, en días, desde la instalación de un ascensor, hasta la primera avería, e Y, tiempo, en días, que tarda la empresa en acudir al lugar del problema, una vez se ha dado el aviso, viene dada por:

$$f(x,y)=\begin{cases}\frac{1}{100}e^{-\frac{x+900y}{300}} & si\ x>0,\ y>0\\ 0 & en\ otro\ caso\end{cases}$$

a) Calcular la función de densidad del tiempo de funcionamiento hasta la primera avería.

b) El jefe de escalera de un vecindario está indignado porque hace 3 días que dio parte a la empresa de una avería y todavía no han acudido a arreglarla. ¿Cuál es la probabilidad que se dé una situación como ésta?

c) ¿Son X e Y independientes?

Solución:

a) $$f_1(x)=\frac{1}{100}\int_{y=0}^{+\infty} e^{-\frac{x+900y}{300}}dy=-\frac{1}{300}\left[e^{-\frac{x+900y}{300}}\right]_{y=0}^{+\infty}=\frac{1}{300}e^{-\frac{x}{300}}$$

b) $$P(Y>3)=\frac{1}{100}\int_{x=0}^{+\infty}\int_{y=3}^{+\infty} e^{-\frac{x+900y}{300}}dx\,dy=\frac{1}{100}\int_{y=3}^{+\infty}\left(\int_{x=0}^{+\infty} e^{-\frac{x+900y}{300}}dx\right)dy=$$

$$=\frac{1}{100}\int_{x=0}^{+\infty}\left(-300\left[e^{-\frac{x+900y}{300}}\right]_{x=0}^{+\infty}\right)dy=\frac{1}{100}\int_{y=3}^{+\infty}e^{-3y}dy=$$

$$=\frac{1}{100}\left[-e^{-3y}\right]_{y=3}^{+\infty}=\frac{1}{100}e^{-9}$$

Que la empresa tarde más de 3 días en venir a arreglar el ascensor es casi imposible.

c) Las variables son independientes si y sólo si $f(x,y)=f_1(x)\,f_2(y)$. Ya hemos calculado $f_1(x)$, calculamos ahora $f_2(y)$

$$f_2(y)=\frac{1}{100}\int_{x=0}^{\infty} e^{-\frac{x+900y}{300}}dx=-\frac{300}{100}\left[e^{-\frac{x+900y}{300}}\right]_{x=0}^{\infty}=3e^{-3y}$$

Y de aquí:

$$f_1(x)f_2(y)=\frac{1}{300}e^{-\frac{x}{300}}3e^{-3y}=\frac{1}{100}e^{-\frac{x+900y}{300}}=f(x,y)$$

Por tanto las variables son independientes.

58. Se estudia un proceso de fabricación que tiene unos límites de tolerancia superior e inferior, y se sabe que la probabilidad de rechazo correspondiente a estos límites es, respectivamente:

$$p_1 = 0{,}06 \qquad p_2 = 0{,}02$$

En una muestra de 100 elementos cierto nº de ellos serán rechazados por exceder las especificaciones y otros por no alcanzarlas. La función de probabilidad conjunta del nº de unidades rechazadas es:

$$P(x_1, x_2, x_3) = \frac{100}{x_1!\,x_2!\,x_3!}(0{,}06)^{x_1}(0{,}02)^{x_2}(0{,}92)^{x_3}$$

a) ¿Cual es la probabilidad de que sean rechazadas 2 piezas por defecto y 5 por exceso de tolerancia?

b) ¿Cuál es la probabilidad de que sean rechazadas cuatro piezas?

c) ¿Cual es la probabilidad de que sean rechazadas dos piezas?

Solución:

a) $P(5,2,93) = \dfrac{100!}{5!2!93!}(0,06)^5(0,02)^2(0,92)^{93} = 0,0448$

b) P(rechazar 4 piezas) =

$$=P(4,0,96)+P(3,1,96)+P(2,2,96)+P(1,3,96)+P(0,4,96)=$$

$$=\frac{100!}{4!96!}(0,06)^4(0,92)^{96}+\frac{100!}{3!96!}(0,06)^3(0,02)(0,92)^{96}+$$

$$+\frac{100!}{2!\,2!96!}(0,06)^2(0,02)^2(0,92)^{96}+\frac{100!}{3!96!}(0,06)(0,02)^3(0,92)^{96}+$$

$$+\frac{100!}{4!96!}(0,02)^4(0,92)^{96}=0,333$$

c) P(rechazar a lo sumo 2 piezas) = P(2, 0, 98)+P(1, 1, 98)+P(0, 2, 98)=

$$=\frac{100!}{2!98!}(0,06)^2(0,92)^{98}+\frac{100!}{98!}(0,06)(0,02)(0,92)^{98}+$$

$$+\frac{100!}{2!98!}(0,02)^2(0,92)^{98}=0,008$$

59. El tiempo que se tarda en reparar una máquina tiene la función de distribución siguiente:

$$F(x)=\begin{cases}0 & si\ x\le 0\\ x/2 & si\ 0<x\le 1\\ 1/2 & si\ 1<x\le 2\\ x/4 & si\ 2<x\le 4\\ 1 & si\ x>4\end{cases}$$

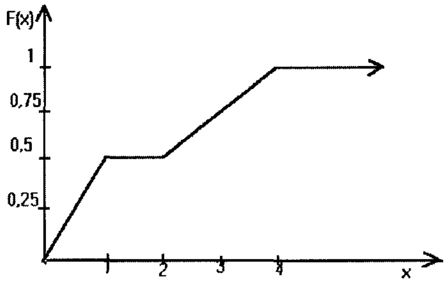
a) Dibujar la función de distribución.

b) Obtener la función de densidad e interpretarla.

c) Si el tiempo de reparación de cierta máquina es superior a una hora, ¿cuál es la probabilidad de que sea superior a tres horas y media?

Solución:

a)



b) Sabemos que:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt \quad \text{de donde} \quad f(t) = \frac{\partial F}{\partial t}$$

$$f(x) = \begin{cases} 0 & \text{si } x \le 0 \\ 1/2 & \text{si } 0 < x \le 1 \\ 0 & \text{si } 1 < x \le 2 \\ 1/4 & \text{si } 2 < x \le 4 \\ 0 & \text{si } x > 4 \end{cases}$$

Luego tenemos: $f(x) = \begin{cases} 1/2 & \text{si } 0 < x \le 1 \\ 1/4 & \text{si } 2 < x \le 4 \\ 0 & \text{en el resto} \end{cases}$

Vemos que la variable, que es tiempo que se tarda en reparar cierta máquina según dice el enunciado, no toma valores sino entre 0 y 1 , y entre 2 y 4. Esto significa que o la reparación dura cualquier cantidad de tiempo hasta una hora (la mitad de las veces) o dura cualquier cantidad de tiempo entre dos y cuatro horas (la otra mitad de las veces).

c) $P(X>3,5/x>1)=\dfrac{P(X>3,5)}{P(X>1)}=\dfrac{1-F(3,5)}{1-F(1)}=\dfrac{1-3,5/4}{1-1/2}=0,25$

60. Una máquina de empaquetado automático deposita por término medio en cada paquete 81,5 g de producto, con desviación típica de 8 g. El peso medio del paquete vacío es de 14,5 g, con desviación típica de 6 g. Ambas variables son normales e independientes. Se pide:

a) Calcular la distribución del peso de los paquetes llenos.

b) Escribir la función de densidad conjunta de la variable:

(X , Y) = (peso producto, peso envase)

Solución:

a) Sean X = peso de producto, $X\in N(81.5\ ,\ 8)$
Y = peso paquete vacío, $Y\in N(14.5\ ,\ 6)$

Sea Z = peso paquete lleno = $X+Y$ entonces

$$Z\in N(81.5+14.5\ ,\ \sqrt{8^2+6^2})=N(96,10)$$

b) La variable bidimensional $(X\ ,\ Y)$ tendrá como función de densidad conjunta el producto de las dos funciones de densidad de las componentes, puesto que el enunciado dice que son independientes, y se tiene:

$$f(x,y)=f_1(x)\cdot f_2(y)=\frac{1}{8\sqrt{2\pi}}e^{-\frac{(x-81,5)^2}{8^2}}\frac{1}{6\sqrt{2\pi}}e^{-\frac{(y-14,5)^2}{6^2}}=$$

$$=\frac{1}{96\pi}e^{-\left[\frac{(x-81,5)^2}{64}+\frac{(y-14,5)^2}{36}\right]}$$

61. En dos grupos de 2º año de carrera de una Facultad se ha medido el cociente de inteligencia de los alumnos. En el grupo A la media fue 100 y la desviación típica 10, y en el grupo B fueron 105 y 12, respectivamente. Ambos grupos tienen el mismo nº de alumnos, y escogido un alumno al azar se comprueba que su cociente de inteligencia es superior a 120. Suponiendo normalidad, ¿cuál es la probabilidad de que el citado alumno provenga del grupo B?

Solución:

Sea X_A= C. I. de un alumno del grupo A, $X_A \in N(100\ ,\ 10)$.

X_B= C. I. de un alumno del grupo B, $X_B \in N(105\ ,\ 12)$.

Sea C = un alumno tiene C.I. superior a 120.

A = un alumno pertenece al grupo A.

B = un alumno pertenece al grupo B.

Tenemos:

$$P(B/C) = \frac{P(B \cap C)}{P(C)} = \frac{P(C/B)\cdot P(B)}{P(C/A)\cdot P(A) + P(C/B)\cdot P(B)} =$$

$$= \frac{P(X_B > 120)\cdot 0,5}{P(X_A > 120)\cdot 0,5 + P(X_B > 120)\cdot 0,5},$$

$$\text{pero} \begin{cases} P(X_A > 120) = P(Z > \dfrac{120-100}{10}) = P(Z > 2) = 0,0228 \\ P(X_B > 120) = P(Z > \dfrac{120-105}{12}) = P(Z > 1,25) = 0,1056 \end{cases}$$

por tanto: $P(B/C) = \dfrac{0,1056}{0,0228 + 0,1056} = \dfrac{0,1056}{0,1284} = 0,8224$

62. En un laboratorio el 40% de los ratones sometidos a un cierto estímulo son machos. El tiempo, en minutos, que un ratón tarda en responder al estímulo es una variable aleatoria exponencial, siendo el tiempo medio de respuesta el doble en los machos que en las hembras.

Se toma un ratón al azar y se le somete al experimento. Si la probabilidad de que tarde más de 4 minutos en responder es 0,8, obténgase el tiempo medio de respuesta para los machos y para las hembras.

Solución:

A= ratón macho P(A)= 0,4

B = tardar más de 4 min en responder al estímulo P(B)= 0,8

X = tiempo que tarda un macho responder al estímulo $X \in Exp(\lambda)$

Y = tiempo que tarda una hembra responder al estímulo $Y \in Exp(\theta)$

$$E[X]=2E[Y] \Rightarrow \frac{1}{\lambda}=\frac{2}{\theta} \Rightarrow \theta=2\lambda$$

$$P(B)=P(B/A)P(A)+P(B/\bar{A})P(\bar{A})$$

$$\text{pero}\begin{cases} P(B/A)=P(X>4)=\int_4^\infty \lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_4^\infty=e^{-4\lambda} \\ P(B/\bar{A})=P(Y>4)=\int_4^\infty \theta e^{-\theta x}dx=\left[-e^{-\theta x}\right]_4^\infty=e^{-4\theta}=e^{-8\lambda} \end{cases}$$

y se tiene:

$0,8=0,4e^{-4\lambda}+0,6e^{-8\lambda}$ y haciendo $x=e^{-4\lambda}$ tenemos

$0,6x^2+0,4x-0,8=0$ cuyas raíces son $\begin{cases} x_1=0,87 \\ x_2=-1,54 \end{cases}$ y de aquí

$$e^{-4\lambda}=0,87 \quad\Rightarrow\quad \lambda=\frac{\ln 0,87}{-4}=0,35 \quad\Rightarrow\quad \theta=0,70 \Rightarrow\begin{cases} \frac{1}{\lambda}=2,88 \\ \frac{1}{\theta}=1,44 \end{cases}$$

Respuesta: el tiempo medio que tardan los ratones machos en reaccionar ante ese estímulo es 2,88 min y las hembras 1,44 min.

63. Supongamos que la probabilidad de seleccionar una palabra de la oración LA NIÑA SE PUSO SU PRECIOSO SOMBRERO ROJO es proporcional al nº de letras que forman dicha palabra. Si X es el nº de letras de la palabra seleccionada. Decir cuáles son: el experimento; el espacio muestral; la función de probabilidad asociada; la variable aleatoria, diciendo de qué tipo es; el recorrido de la variable; la función de probabilidad de la variable; y cuál es el valor de E[X] y V[X].

Solución:

** Experimento: elegir una palabra de entre las que forman la oración LA NIÑA SE PUSO SU PRECIOS0 SOMBRERO ROJO.

** Espacio muestral:

Ω={ la, se, su, niña, puso, rojo, precioso, sombrero}.

** Función de probabilidad asociada: nos dice que la probabilidad de elegir una palabra es proporcional al número de letras que la forman, por tanto:

$$\begin{cases} P(\text{la})=P(\text{se})=P(\text{su})=p \\ P(\text{niña})=P(\text{puso})=P(\text{rojo})=2p \Rightarrow \\ P(\text{precioso})=P(\text{sombrero})=4p \end{cases}$$

$$\Rightarrow 3p+3(2p)+2(4p)=1 \Rightarrow p=\frac{1}{17}$$

Y se tiene:

ω_i	la	se	su	niña	puso	rojo	precioso	sombrero
$P(\omega i)$	1/17	1/17	1/17	2/17	2/17	2/17	4/17	4/17

** Variable aleatoria: X= número de letras de la palabra elegida, X es una v. a. discreta.

** Recorrido de la variable: $X(\Omega)=\{2, 4, 8\}$.

** Función de probabilidad de la v. a. X.

En general $P(X=K)=P(\{\omega_i / X(\omega_i)=k\})$

k	2	4	8
P(X = k)	3/17	6/17	8/17

** Esperanza y varianza de la variable X:

$$E[X]=\sum_{k} k\,P(X=k)=2\cdot\frac{3}{17}+4\cdot\frac{6}{17}+8\cdot\frac{8}{17}=\frac{94}{17}=5,53$$

$$V[X]=E\left[X^2\right]-\left(E[X]\right)^2=\left(4\cdot\frac{3}{17}+16\cdot\frac{6}{17}+64\cdot\frac{8}{17}\right)-\left(\frac{94}{17}\right)^2=5,9$$

64. El n° de coches que diariamente pasan por cierta aldea se distribuye según una ley de Poisson de intensidad λ=20. La probabilidad de que uno de estos coches se detenga en la taberna de la aldea es p=0,2. Estudiar la distribución del n° de coches que paran en la taberna en un día determinado y calcular la probabilidad de que en un día determinando se detengan 5 coches.

Solución:

Sea X = número de coches que pasan por la aldea cierto día, $X\in P(\lambda)$ con $\lambda=20$.

Sea $(Y\,/\,X=n)$ = número de coches que paran en la taberna supuesto que ese día pasan n coches por la aldea $Y/x\in B(n\,,\,p)$ con $p=0,2$.

Queremos estudiar la distribución de Y = número de coches que paran en la taberna cierto día, o sea debemos buscar una ley general que nos dé

$P(Y=k)$ para $k=0, 1, 2, \ldots\ldots.$

$$P(Y=k)=\sum_{i=0}^{\infty}P(Y=k/X=k+i)P(X=k+i)=$$

$$=\sum_{i=0}^{\infty}\binom{k+i}{k}p^k(1-p)^i\frac{\lambda^{k+i}}{(k+i)!}e^{-\lambda}=$$

$$=\sum_{i=0}^{\infty}\frac{(k+i)!}{k!\,i!}p^k(1-p)^i\frac{\lambda^{k+i}}{(k+i)!}e^{-\lambda}=$$

$$=\frac{p^k\lambda^k e^{-\lambda}}{k!}\sum_{i=0}^{\infty}\frac{(1-p)^i\lambda^i}{i!}=\frac{p^k\lambda^k e^{-\lambda}}{k!}e^{(1-p)\lambda}=\frac{(p\lambda)^k}{k!}e^{-p\lambda}$$

Este resultado pone de manifiesto que la variable Y sigue una distribución de Poisson de parámetro pλ, que en nuestro caso es 0,2*20=4. Por tanto, respondiendo a la segunda pregunta:

$$P(Y=5)=\frac{4^5}{5!}e^{-4}=0{,}1562$$

65. Al efectuar la suma de 300 números reales se aproxima cada uno de los sumandos por el entero más próximo. Así si uno de los sumandos fuera el número 13,845..., para efectuar la suma este sumando se tomaría como 14.

Suponiendo que la distribución de decimales es uniforme en el intervalo [-0,5 , 0,5] se pide calcular la probabilidad de que la diferencia entre la suma de los números reales S y la suma S* que resulta de los redondeos no sea superior a cuatro unidades.

Solución:

Sean X_i i=1, 2,...,300 los sumandos reales y X_i^* los sumandos enteros que son aproximaciones de los Xi. La relación entre ellos vendrá dada por $X_i = X_i^* + d_i$ con $d_i \in$ [-0,5 , 0,5].

Llamando S a la suma de los 300 sumandos reales y S^* a la suma de los 300 sumandos enteros aproximaciones de los reales, se tiene:

$$\sum_{i=1}^{300} X_i=\sum_{i=1}^{300} X_i^*+\sum_{i=1}^{300} d_i \Rightarrow S-S^*=\sum_{i=1}^{300} d_i$$

Lo que nos piden es $P(\left|S-S^*\right|<4)=P(-4<S-S^*<4)$.

Para poder calcularlo necesitamos conocer la distribución de la variable de interés que es S-S*.

Cada una de las $d_i \in U$ (-0,5 , 0,5) por tanto :

$$E\left[d_i\right]=\frac{-0{,}5+0{,}5}{2}=0 \;;\qquad V\left[d_i\right]=\frac{b-a}{12}=\frac{1}{12}$$

S-S* es suma de 300 v. a. idénticamente distribuidas e independientes, entonces por el teorema central del límite.

$$S-S^* \cong N\left(0\,,\sqrt{300\cdot\frac{1}{12}}\right)=N(0\,,5)$$

y ya estamos en condiciones de calcular la probabilidad pedida.

$$P(-4 < S - S^* < 4) = P(-\frac{4}{5.} < Z < \frac{4}{5}) = 0,7881 - 0,2119 = 0,5762$$

66. Se está estudiando si interesa fabricar una máquina automática nueva para soldar. Se considerará exitosa si tiene una efectividad del 99% en sus soldaduras, y si no, no se fabricará. Se lleva a cabo una prueba con un prototipo y se realizan 100 soldaduras. La máquina se aceptará para su fabricación si no son defectuosas más de tres soldaduras.

a) ¿Cuál es la probabilidad de que una máquina eficiente sea rechazada?

b) ¿Cuál es la probabilidad de que una máquina ineficiente sea rechazada? (Supongamos que el verdadero porcentaje de soldaduras buenas fuera el 95%).

Solución:

Sea X = número de soldaduras defectuosas de las 100, $X \in$ B(100, p), con p sin determinar:

a) Si la máquina es eficiente p = 0.01, entonces X $\in$ B(100, 0.01); como n·p =1<5 aproximamos X por X' siendo X' $\in$ Poisson(1).

$$P(\text{rechazar máquina eficiente}) = P(X \geq 4) = 1 - P(X \leq 3) =$$

$$= 1 - P(X' \leq 3) = 1 - \sum_{i=1}^{3} \frac{1^i}{i!} \cdot e^{-1} = 1 - 0,981 = 0.019$$

b) Si la máquina no es eficiente y p = 0,05, entonces $X \in$ B(100, 0.05), y como n>30 y np= 5 aproximamos la binomial por la normal

$$X'' \in N\left(np, \sqrt{np(1-p)}\right) = N\left(5, \sqrt{4.75}\right)$$

y tenemos

$$P(\text{rechazar máquina ineficiente}) =$$

$$= P(X > 3) \cong P(X'' \geq 3,5) = P(\frac{X'' - 5}{\sqrt{4,75}} > \frac{3,5 - 5}{\sqrt{4,75}}) = 0,74$$

67. En un campeonato de ajedrez entre dos jugadores, las tablas no se tienen en cuenta y el juego continúa hasta que uno de los contrincantes alcanza seis puntos (ganar = 1 punto, perder = 0 puntos). Considerando a los participantes igualmente expertos, hallar la probabilidad de que, con tales reglas, en el momento de finalizar la confrontación, el perdedor tenga k puntos, k = 0, 1, 2, 3, 4, 5.

Solución:

Sea A = el jugador A gana una partida, B = el jugador B gana una partida.

Dado que los dos jugadores son igualmente expertos y que las tablas no se tienen en cuenta, se tiene $P(A) = P(B) = 1/2$.

Si A gana el torneo, la última partida la ha ganando A

Sea Y = número de puntos que tiene B cuando A gana la 6ª vez.

$$P(Y=k)=\binom{6+k-1}{k}\ 0.5^{6+k}$$

La probabilidad pedida es el doble de esta cantidad, ya que ahora debemos sumar la probabilidad de que A tenga k puntos al finalizar un torneo en que B sea ganador. El razonamiento es exactamente igual.

68. Un bolso contiene tres monedas, dos de las cuales son normales y equilibradas y la otra tiene dos caras. Se escoge una moneda al azar y se tira cuatro veces. Si las cuatro veces sale cara, ¿qué probabilidad hay de que se haya escogido la moneda con dos caras?

Solución:

Sea A = la moneda elegida es una de las normales, P(A) = 2/3.

B = al tirar cuatro veces la moneda elegida, salen 4 caras.

Lo que se pide es:

$$P(\bar{A}/B)=\frac{P(\bar{A}\cap B)}{P(B)}=\frac{P(B/\bar{A})P(\bar{A})}{P(B/A)P(A)+P(B/\bar{A})P(\bar{A})}=$$

$$=\frac{1\cdot\frac{1}{3}}{\left(\frac{1}{2}\right)^4\cdot\frac{2}{3}+1\cdot\frac{1}{3}}=\frac{24}{27}$$

69. La duración de cierto tubo de radio es una variable aleatoria continua con función de densidad:

$$f(x)=\begin{cases}\frac{100}{x^2} & \text{si x>100}\\ 0 & \text{en el resto}\end{cases}$$

a) ¿Cuál es la probabilidad de que un tubo dure menos de 200 horas si se sabe que todavía funciona después de 150 horas?

b) ¿Cuál es la probabilidad de que si se instalan 3 de tales tubos, exactamente uno deba ser sustituido después de 150 horas?

c) ¿Cuál es el nº máximo de tubos que se debe poner en un sistema para que haya una probabilidad de al menos 0,5 de que tras 150 horas de servicio todos funcionen?

Solución:

a) Sea X = duración de un tubo de radio (horas)

$$P\left(X<200/X>150\right)=\frac{P(X<200\cap X>150)}{P(X>150}=\frac{P(150<X<200)}{P(X>150)}=$$

$$=\frac{\int_{150}^{200}\frac{100}{x^2}dx}{\int_{150}^{+\infty}\frac{100}{x^2}dx}=\frac{\left[-\frac{100}{x}\right]_{150}^{200}}{\left[-\frac{100}{x}\right]_{150}^{+\infty}}=\frac{150}{600}=0,25$$

b) Sea Y = número de componentes que se estropean antes de 150 h de los tres instalados en el sistema, Y es una binomial B(3 , p) donde P = P(X<150) = 1-2/3 = 1/3 por el apartado 1).

$$P(Y=1)=\binom{3}{1}\frac{1}{3}\frac{4}{9}=\frac{4}{9}$$

c) Sea Z = número de tubos que funcionan más de 150 horas de los n instalados, Z es una binomial B(n, 2/3).

$$P(Z=n)=\left(\frac{2}{3}\right)^n \Rightarrow \; P(Z=n)>0,5 \Leftrightarrow \; n\cdot\ln\frac{2}{3}>\ln 0,5$$

$$\Rightarrow \; n>1,71 \Rightarrow n=2$$

70. Tenemos dos tipos de urnas: U y V. Cada una contiene bolas blancas y negras. En cada una de las urnas tipo U hay 20 bolas blancas y 80 negras. En cada una de las urnas tipo V hay 40 bolas blancas y 60 negras. Calcular en qué proporción se encuentran ambos tipos de urnas, sabiendo que la probabilidad de que una bola blanca proceda de una urna tipo U es 0,2.

Solución:

Está claro que no hay el mismo número de urnas U que de urnas V, al menos en principio. Si llamamos U = sacar bola de una urna tipo U, tenemos P(U)=p. Si llamamos V= sacar bola de una urna tipo V, tendremos P(V)=1-p. Además:

$$P(B/U)=\frac{20}{100}=0,2 \quad P(B/V)=\frac{40}{100}=0,4 \quad P(U/B)=0,2$$

De donde:

$$P(U/B)=\frac{P(B\cap U)}{P(B)}=\frac{P(B/U)P(U)}{P(B/U)P(U)+P(B/V)P(V)}=$$

$$=\frac{0,2p}{0,2p+0,4(1-p)}=0,2 \Rightarrow 0,2p=0,04p+0,08-0,08p \Rightarrow$$

$$\Rightarrow 0,24p=0,08 \Rightarrow p=\frac{1}{3} \quad y \quad 1-p=\frac{2}{3}$$

71. La probabilidad de que la media aritmética de 64 variables aleatorias independientes, igualmente distribuidas, con media α y varianza 16α, esté comprendida entre $-\alpha$ y 3α es 0,6. Calcular α.

Solución:

Sean $X_1,...,X_{64}$ las v. a. independientes e idénticamente distribuidas donde:

$$E[X_i]=\alpha \quad \text{y} \quad Var[X_i]=16\alpha, \ \forall i=1,2,...,64$$

La variable $\sum_{i=1}^{64} X_i$ puede considerarse que sigue una distribución normal, por el teorema central del límite. Entonces la media muestral $\bar{X}=\frac{1}{64}\sum X_i$ también sigue una distribución normal y sus parámetros son, según sabemos:

$$E[\bar{X}]=\alpha \qquad\qquad V[\bar{X}]=\frac{16\alpha}{64} \Rightarrow \sigma_{\bar{X}}=\frac{\sqrt{\alpha}}{2}$$

$$P(-\alpha<\bar{X}<3\alpha)=P(\frac{-\alpha-\alpha}{\sqrt{\alpha}/2}<\frac{\bar{X}-\alpha}{\sqrt{\alpha}/2}<\frac{3\alpha-\alpha}{\sqrt{\alpha}/2})=$$

$$=P(\frac{-4\alpha}{\sqrt{\alpha}}<Z<\frac{4\alpha}{\sqrt{\alpha}})=0,6 \Rightarrow \alpha=\frac{0,84^2}{16}=0,0441$$

72. Se observa una fuente radioactiva durante 5 intervalos de tiempo no solapados de 8 segundos de duración cada uno. Las partículas son emitidas según una ley de Poisson con una media de 0,5 partículas por segundo y registradas en un contador. Hallar la probabilidad de que:

a) En cada uno de los cinco intervalos se cuenten 4 o más partículas.

b) Se cuenten 4 o más partículas al menos en uno de los intervalos.

Solución:

Sea X = número de partículas emitidas en un intervalo de 8 s, X sigue una distribución de Poisson de parámetro $\lambda = 4$, puesto que el promedio de partículas por segundo es 0,5. Sea Y = número de veces que se emiten 4 o más partículas de los 5 intervalos de 8 s estudiados, Y sigue una distribución binomial B(5, p), siendo

$$p = P(X \geq 4) = \frac{4^4}{4!}e^{-4} = 0,193$$

a) $P(Y = 5) = 0.193^5 = 0,00027$

b) $P(Y \geq 1) = 1 - P(Y = 0) = 1 - 0.807^5 = 0,342$

73. Un empleado de un banco sustituye cautelosamente un billete bueno por uno falso en cada fajo de cien billetes. Si el interventor del banco toma un billete al azar de cada uno de 50 fajos, ¿cuál es la probabilidad de que tope con algún billete falso?

Solución:

Sea A = sacar un billete falso de un fajo, $P(A) = 0,01$.

Sea X = número billetes falsos de los 50 extraídos, $X \in B(50, 0,01)$.

P(algún billete falso) = $P(X \geq 1) = 1-P(X = 0) = 1-0,99^{50} = 1-0,6050 = 0,395$.

74. En Alemania, durante el siglo XVI, la unidad de longitud se determinaba así: Un domingo, los primeros dieciséis hombres que salían de la iglesia eran puestos en fila. Se tomaban a continuación las medidas de sus pies izquierdos, se sumaban y se dividían entre dieciséis. El resultado era el *pie correcto y legal* a utilizar en las transacciones.

Se sabe que la longitud (en mm) del pie de un hombre adulto de la época es aleatoria con distribución N(262,5 , 12).

Encontrar la probabilidad de que dos valores del *pie correcto y legal*, determinados en dos diferentes iglesias, difieran entre sí en más de 5 mm.

Solución:

Sea X_i = longitud pie hombre i-ésimo iglesia 1, $X_i \in N(262,5\ , 12)$, i =1,...16

Sea Y_i = longitud pie hombre i-ésimo iglesia 2, $Y_i \in N(262,5\ , 12)$, i =1,...16

Según sabemos:

$$\bar{X}=\frac{1}{16}\sum_{i=1}^{16}X_i\ ,\quad \bar{Y}=\frac{1}{16}\sum_{i=1}^{16}Y_i$$

y se tiene $\bar{X}\in N(262,5\ ,3)$ e $\bar{Y}\in N(262,5\ ,3)$

ya que

$$E\left[\bar{X}\right]=E\left[\frac{1}{16}\Sigma X_i\right]=\frac{1}{16}\Sigma E\left[X_i\right]=\frac{1}{16}\cdot 16\cdot 262,5=262,5$$

$$V\left[\bar{X}\right]=V\left[\frac{1}{16}\Sigma X_i\right]=\frac{1}{16^2}\Sigma V\left[X_i\right]=\frac{12^2}{16}\Rightarrow \sigma_{\bar{X}}=\frac{12}{4}=3$$

Sea W = $\bar{X}-\bar{Y}$, W$\in$N(0 , $3\sqrt{2}$), ya que:

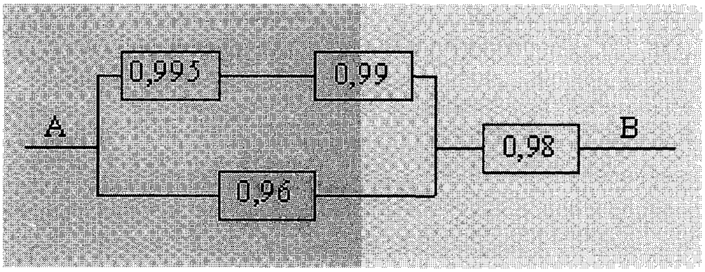
$$E[W]=E\left[\bar{X}-\bar{Y}\right]=E\left[\bar{X}+(-1)\bar{Y}\right]=E\left[\bar{X}\right]-E\left[\bar{Y}\right]=0$$
$$V[W]=V\left[\bar{X}-\bar{Y}\right]=V\left[\bar{X}+(-1)\bar{Y}\right]=V\left[\bar{X}\right]+V\left[\bar{Y}\right]=9+9=18$$
$$\Rightarrow \sigma_W=3\sqrt{2}$$

Ya podemos calcular lo que nos piden, que será:

$$P(|W|>5)=P(W<-5)+P(W>5)=2\cdot P(W<-5)=2\cdot P(Z<\frac{5}{3\sqrt{2}})=$$

$$= 2.P(Z< -1,18) = 2. =,1190 = 0,238$$

75. Consideremos una red de comunicaciones como la de la figura. Dos puntos están comunicados si hay alguna línea que los una, todos cuyos repetidores (representados en el diagrama mediante rectángulos) funcionan. Sobre cada repetidor se indica la probabilidad de que esté en servicio. ¿Cuál es la probabilidad de que A y B queden incomunicados?



Solución:

Llamemos C, D, E y F a los cuatro repetidores.

Sea C = el repetidor C funciona, $P(C) = 0{,}995$.

D = el repetidor D funciona, $P(D) = 0{,}99$.

E = el repetidor E funciona, $P(E) = 0{,}96$.

F = el repetidor F funciona, $P(F) = 0{,}98$.

Para que los puntos A y B estén comunicados tiene que ocurrir el suceso $[[(C\cap D)\cup E]\cap F]$, y por tanto:

$P(\text{A y B estén incomunicados}) = 1 - P([(C\cap D)\cup E]\cap F) =$

$= 1 - P[(C\cap D)\cup E]P(F) =$

$= 1 - [P(C\cap D) + P(E) - P(C\cap D\cap E)]\, P(F) =$

$= 1 - [P(C)P(D) + P(E) - P(C)\, P(D)\, P(E)]\, P(F) =$

$= 1 - (0{,}995\text{x}0{,}99 + 0{,}96 - 0{,}995\text{x}0{,}99\text{x}0{,}96\text{x}0{,}98 = 0{,}02058$

76. Cinco personas juegan a los chinos. ¿Cuál es la probabilidad de que el total de monedas escondidas sea 7? (El juego de los chinos consiste en adivinar el nº total de monedas que un grupo de jugadores tienen escondidas, sabiendo que cada jugador puede tener 0, 1, 2 ó 3 monedas, a voluntad).

Solución:

Cada persona puede elegir sacar 0, 1, 2, ó 3 monedas, suponemos que con probabilidad 0,25 cada caso. Como son 5 personas el número de casos posibles es $VR_4^5 = 4^5 = 1.024$. De todos los casos posibles serán casos favorables todas las posibles maneras de poner 5 sumandos que sólo pueden tomar el valor 0, 1, 2 ó 3 y de manera que sumen 7.

Los únicos tipos de sumandos que cumplen estas condiciones son los de los seis tipos siguientes:

i)	3	3	1	0	0
ii)	3	2	2	0	0
iii)	3	2	1	1	0
iv)	3	1	1	1	1
v)	2	2	2	1	0
vi)	2	2	1	1	1

Ahora hay que contar cuántos casos hay de cada tipo, teniendo en cuenta que cada combinación (p.e. 2 treses, 1 uno y 2 ceros) se puede tener de maneras distintas, según qué jugador sea el que tiene cada cantidad de monedas en la mano.

Para contar los casos que hay de tipo i) razonamos de la forma siguiente:

De los cinco jugadores dos de ellos deben tener 3 monedas: hay $\binom{5}{2}$ formás posibles. De los tres jugadores restantes dos de ellos deben tener 0 monedas: Hay $\binom{3}{2}$ casos posibles. El jugador que queda será el que tenga una sola moneda. De este modo el número total de casos tipo i) que pueden presentarse son:

Número de casos tipo i):

$$\binom{5}{2}\binom{3}{2} = \frac{5!}{2!\ 3!}\ \frac{3!}{2!\ 1!} = 30$$

Siguiendo un razonamiento análogo en cada caso tenemos:

Nº de casos tipo ii): $\binom{5}{1}\binom{4}{2} = \dfrac{5!}{1!\ \ 2!} = 30$

Nº de casos tipo iii): $\binom{5}{1}\binom{4}{1}\binom{3}{2} = \dfrac{5!}{2!} = 60$

Nº de casos tipo iv): $\binom{5}{1} = 5$

Nº de casos tipo v): $\binom{5}{3}\binom{2}{1} = \dfrac{5!}{3!} = 20$

Nº de casos tipo vi): $\binom{5}{2} = \dfrac{5!}{2!\ \ 3!} = 10$

Total casos favorables = 155

$$P(\text{sumar } 7) = \frac{155}{1.024}$$

77. El número de hamburguesas que se venden semanalmente en un *burguer*, en cientos de unidades, es una v. a. con función de densidad:

$$f(x) = \begin{cases} \dfrac{x^2}{9} & \text{si } 0 < x < 3 \\ 0 & \text{en el resto} \end{cases}$$

Si en una semana se venden x hamburguesas, la cantidad de ellas que llevan queso, Y, se distribuye según la siguiente función de densidad:

$$f(y/X = x) = \begin{cases} \dfrac{1}{x} & \text{si } 0 < y < x < 3 \\ 0 & \text{en el resto} \end{cases}$$

a) Sabiendo que cierta semana se han vendido cien hamburguesas, ¿cuál es la probabilidad de que por lo menos la mitad de ellas tuvieran queso?

b) Calcular la probabilidad de que en una semana el nº de hamburguesas sin queso sea superior a cien?

Solución:

Sea X = número hamburguesas por semana (cientos)

Y = número hamburguesas con queso por semana (cientos)

a) Para X = 1 se tiene $f(y_{/X=1}) = 1$ y de aquí

$$P(Y > 0{,}5_{/X=1}) = \int_{y=0,5}^{1} f(y_{/X=1})dy = \int_{y=0,5}^{1} dy = [y]_{0,5}^{1} = 0{,}5$$

b) Sea X-Y = número hamburguesas sin queso (cientos)

$$f(x\ ,\ y) = f(y/X=x)\ \ f(x) = \begin{cases} \dfrac{x^2}{9}\ \dfrac{1}{x} = \dfrac{x}{9} & \text{si } 0<x<3\ ,\ 0<y<x<3 \\ 0 & \text{en otro caso} \end{cases}$$

$$P(X-Y>1) = \int_{x=1}^{x=3}\int_{y=0}^{y=x-1} f(x\ ,\ y)\ dx\ dy = \int_{x=1}^{x=3}\left[\int_{y=0}^{y=x-1}\frac{x}{9}\ dy\right]dx =$$

$$= \int_{x=1}^{x=3}\left[\frac{xy}{9}\right]_{y=0}^{y=x-1} dx = \int_{x=1}^{x=3}\frac{x\ \ (x-1)}{9}\ dx = \frac{1}{9}\left[\frac{x^3}{3}-\frac{x^2}{2}\right]_{x=1}^{x=3} =$$

$$= \frac{1}{9}\left[\left(9-\frac{9}{2}\right)-\left(\frac{1}{3}-\frac{1}{2}\right)\right] = \frac{28}{54} = \frac{14}{27}$$

78. La demanda diaria, en unidades de cierto producto durante 30 días laborables fue:

38	35	76	58	48	59
67	63	33	69	53	51
30	21	32	32	59	62
49	78	48	42	72	52
47	66	52	44	44	56

a) Agrupar en cinco intervalos y construir las tablas de frecuencias agrupadas. Dibujar el histograma de frecuencias y el polígono de frecuencias acumuladas.

b) Calcular la mediana y la desviación mediana para los datos sin agrupar y agrupados.

Solución:

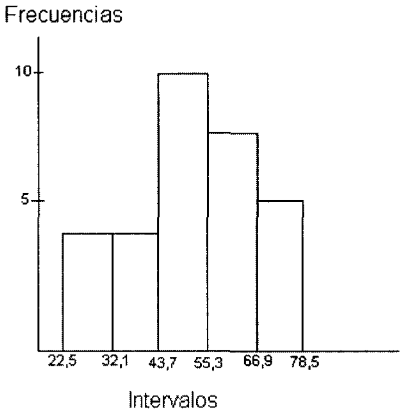
a) Rango = 78-21 =57

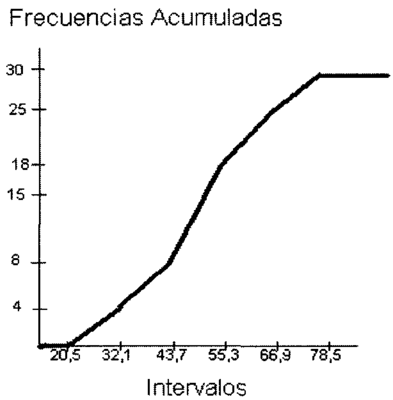
Añadimos una unidad de precisión, total 58.

Calculamos la longitud de cada grupo 58/5 = 11,6.

Escribimos la tabla de frecuencias absolutas y relativas y las correspondientes acumuladas, con los datos ya agrupados.

$(x_{i-1}, x_i]$	x_{M_i}	f_i	$f\,ac._i$	$f\,r_i$	$f\,r\,ac._i$
(20,5 , 32,1]	26,3	4	4	4/30	4/30
(32,1 , 43,7]	37,9	4	8	4/30	8/30
(43,7 , 55,3]	49,5	10	18	10/30	18/30
(55,3 , 66,9]	61,1	7	25	7/30	25/30
(66,9 , 78,5]	72,7	5	30	5/30	30/30





b) Para los datos sin agrupar:

$$\text{Mediana}=\frac{x_{15}+x_{16}}{2}=\frac{51+52}{2}=51,5$$

$$\text{Desviación mediana}=\frac{\sum|x_i-Md|}{N}=$$

$$= \frac{|38-51,5|+|67-51,5|+...+|56-51,5|}{30} = 11,6$$

Para los datos agrupados:

Mediana:

i) localizamos el intervalo que verifique

$$\begin{cases} f\,ac(x_{i-1}) < \dfrac{N}{2} \\ f\,ac.(x_i) \geq \dfrac{N}{2} \end{cases} \text{ que es } (43{,}7\ ,\ 55{,}3]$$

$$\text{ii)}\ \frac{N}{2} - f\ ac(x_{i-1}) = 15 - 8 = 7 = c$$

$$\text{iii)}\ x_i - x_{i-1} = 11,6; \quad \frac{x_i - x_{i-1}}{f_i} = \frac{11,6}{10} = 1,16$$

$$\text{iv)}\ Md = x_{i-1} + c \cdot \frac{x_i - x_{i-1}}{f_i} = 43{,}7 + 7 \cdot 1{,}16 = 51{,}8$$

$$\text{Desviación mediana} = \frac{\sum |x_{M_i} - Md| \cdot f_i}{N} =$$

$$= \frac{1}{30}(|26{,}3 - 51{,}82| \cdot 4 + |37{,}9 - 51{,}82| \cdot 4 + |49{,}5 - 51{,}82| \cdot 10 +$$

$$+ |61{,}1 - 51{,}82| \cdot 7 + |72{,}7 - 51{,}82| \cdot 5) = 11{,}67$$

79. La tasa de suicidios de un país es de uno al mes por cada millón de habitantes.

Se pide:

a) Calcular el número medio de meses sin suicidios al año.

b) Calcular, en una ciudad de dos millones y medio de habitantes, la probabilidad de que en un mes haya más de cuatro suicidios.

Solución:

Sea X = número de suicidios al mes (por millón de habitantes). Consideramos que la variable sigue un modelo de Poisson con $\lambda=1$.

a) Sea Y = número de meses, de los 12 de un año, en que no hay suicidios (seguimos considerando que hablamos de un millón de habitantes).

Y sigue una distribución binomial siendo $p = P(X=0) = 1/e$.

Por tanto:

$$E[Y] = n \cdot p = \frac{12}{e} = 4,41$$

b) Sea X' = número de suicidios al mes en una ciudad de 2,5 millones de habitantes, $X' \in P(2,5)$. Entonces:

$$P(X' > 4) = 1 - P(X' \leq 4) = 1 - 0,8912 = 0,1088$$

Algo más del 10% de los meses ocurren más de 4 suicidios.

80. Supongamos que los diámetros de los pernos de una caja grande siguen una distribución normal con una media de 2 cm y desviación típica de 0,03 cm. Además supongamos que los diámetros de los agujeros de las tuercas de otra caja grande siguen una distribución normal de media 2,02 cm. Y desviación típica de 0,04 cm. Un perno y una tuerca ajustarán si el diámetro del agujero de la tuerca es mayor que el diámetro del perno y la diferencia entre ambos diámetros no es mayor que 0,05 cm. Si se seleccionan al azar un perno y una tuerca, ¿cuál es la probabilidad de que ajusten?

Solución:

Sea X = diámetro de un perno, $X \in N(2\ ,\ 0,03)$.

Sea Y = diámetro de una arandela, $Y \in N(2,02\ ,\ 0,04)$.

Para que escogidos un perno y una arandela al azar, encajen, debe ocurrir

$$0 < Y\text{-}X < 0,05.$$

Pero,

$$Y\text{-}X \in N(0{,}02\ ,\sqrt{0{,}03^2+0{,}04^2}\,) = N(0{,}02\quad ,0{,}05)$$

Por tanto:

$$P(0<Y-X<0{,}05)=P(-\frac{0{,}02}{0{,}05}<Z<\frac{0{,}03}{0{,}05})=$$
$$=P(-0{,}4<Z<0{,}6)=0{,}7257-0{,}3446=0{,}3811$$

Sólo en algo más del 38% de los casos un perno y una tuerca elegidos al azar encajarán.

81. La llegada del correo a una empresa se produce, con igual probabilidad en cualquier instante entre las 10 y las 11 horas. El conserje lo entrega a la secretaria del director, también con igual probabilidad, en cualquier instante desde el momento que llega pero nunca más tarde de las 11. ¿A qué hora, en promedio, recibirá la secretaria el correo? ¿Cuál es la probabilidad de recibirlo antes de las 10 y media?

Solución:

Sea X= momento en que el conserje recibe el correo, $X \in$ U(10, 11) y por tanto:

$$f(x)=\begin{cases}1 & \text{si } \ 10\le x\le 11\\ 0 & \text{en el resto}\end{cases}$$

Sea $Y_{/X=x}$ momento en que la secretaria recibe el correo, supuesto que el conserje lo recibió en el momento X=x, $Y_{/X=x} \in$ U(x, 11) y por tanto:

$$f(y_{/X=x})=\begin{cases}\dfrac{1}{11-x} & \text{si } \ x<y<11\\ 0 & \text{en el resto}\end{cases}$$

De aquí tenemos:

$$f(x,\ y) = f\left(y_{/X=x}\right) f(x) = \begin{cases} \dfrac{1}{11-x} & \text{si } 10 \le x \le y \le 11 \\ 0 & \text{en el resto} \end{cases}$$

Por definición:

$$E\left[g(x,y)\right] = \iint_{(X,Y)(\Omega)} g(x,y) f(x,y)\, dx\, dy$$

En particular para g(x,y)=y, y con esta f será:

$$E[y] = \int_{x=10}^{11} \int_{y=x}^{11} \left(\frac{y}{11-x} dy \right) dx = \int_{x=10}^{11} \frac{1}{2(11-x)} \left[y^2 \right]_{y=x}^{11} dx =$$

$$= \int_{x=10}^{11} \frac{11^2 - x^2}{2(11-x)} dx = \int_{x=10}^{11} \frac{11+x}{2} \, dx = \left[\frac{11}{2} x + \frac{1}{4} x^2 \right]_{x=10}^{11} =$$

$$= \frac{121}{2} + \frac{121}{4} - \frac{110}{2} - \frac{100}{4} = \frac{43}{4} = 10,75$$

En promedio la secretaria recibirá el correo a las 11 menos cuarto.

$$P(Y < 10,5) = \int_{10}^{10,5} \int_{10}^{x} \frac{1}{11-x} dy \quad dx = \int_{10}^{11} \frac{1}{11-x} [y]_{10}^{x} dx =$$

$$= \int_{10}^{10,5} \frac{x-10}{11-x} dx = \left[-x - L(11-x) \right]_{10}^{10,5} = 0,1931$$

La secretaría recibirá el correo antes de las 10:30 algo más del 19% de las veces.

82. Una persona tira dos dados, uno de ellos es un cubo y el otro un tetraedro regular, tomando el número de la cara inferior cuando se trata del tetraedro, ¿cuál es la probabilidad de que la suma de los números obtenidos no sea menor que 5?

Solución:

El espacio muestral del experimento, denotando en primer lugar el resultado del dado, será:

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), (2,1),..., (2,4),..., (6,1),..., (6,4)\}$$

en total 24 casos posibles.

Sea A = la suma de puntos no es menor que 5 = suma de puntos ≥ 5.

$$P(A) = 1 - P(\overline{A}) = 1 - P(\{(1,\ 1),\ (1,\ 2),\ (1,\ 3),\ (2,\ 1),\ (2,\ 2),\ (3,\ 1)\}) =$$
$$= 1 - \frac{6}{24} = 0,75$$

83. En función de su prioridad un programa para ordenador espera en la fila de entrada cierto tiempo, después del cual lo ejecuta el procesador central en un determinado tiempo. La función de densidad conjunta para los tiempos de espera y de ejecución viene dada por:

$$f(t_1,t_2) = \begin{cases} 2\ e^{-(\frac{t_1}{5}+10t_2)} & \text{si } t_1,\ t_2 > 0 \\ 0 & \text{en el resto} \end{cases}$$

Se pide:

a) Obtener la probabilidad conjunta de que el tiempo de espera no sea mayor de 8 minutos y el de ejecución no sea mayor de 12 segundos.

b) Obtener las funciones de densidad marginales y analizar la independencia de ambas variables.

c) Determinar el tiempo esperado de ejecución de un programa, si esperó en la fila de entrada un tiempo t_1^*.

Solución:

a) $$P(X \le 8,\ Y \le 12) = \int_{t_1=0}^{8} \int_{t_2=0}^{12} 2\ e^{-(\frac{t_1}{5}+10t_2)} dt_1\ \ dt_2 =$$

$$= \int_{t_1=0}^{8} \left[-\frac{2}{10}\ e^{-\left(\frac{t_1}{5}+10t_2\right)} \right]_{t_2=0}^{12} dt_1 = -\frac{1}{5} \int_{t_1=0}^{8} (e^{-\frac{t_1}{5}-120} - e^{-\frac{t_1}{5}})\ \ dt_1 =$$

$$= -\frac{1}{5}\left[-5\ e^{-\frac{t_1}{5}-120}\right]_{t_1=0}^{8} + \frac{1}{5}\left[-5\ e^{-\frac{t_1}{5}}\right]_{t_1=0}^{8} =$$

$$= e^{-\frac{8}{5}-120} - e^{-120} - e^{-\frac{8}{5}} + 1 = 0,7981$$

b) $$f_1(t_1) = \int_{t_2=0}^{\infty} 2\ e^{-(\frac{t_1}{5}+10t_2)}\, dt_2 = -\frac{2}{10}\left[e^{-(\frac{t_1}{5}+10t_2)}\right]_{t_2=0}^{\infty} =$$

$$= \frac{1}{5}\ e^{-\frac{t_1}{5}}, \quad t_1 \in Exp(\frac{1}{5})$$

$$f_2(t_2) = \int_{t_1=0}^{\infty} 2\ e^{-(\frac{t_1}{5}+10t_2)}\, dt_1 = -10\left[e^{-(\frac{t_1}{5}+10t_2)}\right]_{t_1=0}^{\infty} =$$

$$= 10\ e^{-10t_2}, \quad t_2 \in Exp(10)$$

Además:

$$f_1(t_1)\cdot f_2(t_2) = \frac{1}{5}\ e^{-\frac{t_1}{5}} \cdot 10\ e^{-10t_2} = 2\ e^{-(\frac{t_1}{5}+10t_2)} = f(t_1,\ t_2)$$

y concluimos que las variables son independientes.

c) Nos piden:

$$E\left[T_2 / T_1 = t_1^*\right] = \int_0^{\infty} t_2 \cdot f(t_2 / T_1 = t_1^*)\ dt_2$$

pero, como acabamos de concluir que las dos variables son independientes, las condicionadas coinciden con las marginales. Además la función de densidad marginal de t_2 es la de una variable exponencial de parámetro 10, según ya señalamos, y por tanto su esperanza es la inversa del parámetro, es decir 1/10= 0,1.

84. Supongamos que cuando cierta máquina está correctamente ajustada, el 50% de los artículos que produce son de alta calidad y el otro 50% son de calidad media. Supongamos, sin embargo, que la máquina está mal ajustada durante el 10% del tiempo y que, en estas condiciones, el 25% de los artículos producidos por ella son de alta calidad y el 75 % de los artículos son de calidad media.

a) Supongamos que 5 artículos producidos por la máquina en un tiempo determinado son seleccionados al azar e inspeccionados. Si cuatro de ellos son de alta calidad y uno de calidad media, ¿cuál es la probabilidad de que la máquina estuviera correctamente ajustada durante ese tiempo?

b) Supongamos que se selecciona un artículo adicional , que fue producido por la máquina al mismo tiempo que los otros cinco, y resulta ser de calidad media. ¿Cuál es la nueva probabilidad final de que la máquina estuviera correctamente ajustada?

Solución:

a) Sea A = la máquina está bien ajustada, $P(A)$=0,90

$$B = \text{un artículo es de alta calidad}\begin{cases} P(B/A)=0,5 \\ P(B/\bar{A})=0,25 \end{cases}$$

Sea X= número de artículos de alta calidad cuando se extraen 5 de la cadena.

$$X_{/A} \in B(5,\ 0.5)\ \ y\ \ X_{/\bar{A}} \in B(5,\ 0.25)$$

y por tanto:

$$P(X=4)=P(X=4_{/A})\cdot P(A)+P(X=4_{/\bar{A}})\cdot P(\bar{A})=$$
$$=5\cdot 0,5^5\cdot 0,9+5\cdot 0,25^4\cdot 0,75\cdot 0,1=0,142$$

Nos piden:

$$P(A_{/X=4})=\frac{P(A\cap X=4)}{P(X=4)}=\frac{P(X=4_{/A})P(A)}{P(X=4)}=\frac{5\cdot 0,5^5\cdot 0,9}{0,142}=0,9896$$

b) Consideramos la sexta extracción como única realización aleatoria, puesto que sabemos el resultado de las cinco anteriores.

$$\text{Sea } Y=\begin{cases} 0 & \text{si el } 6^{\circ} \text{ artículo es de calidad media} \\ 1 & \text{si el } 6^{\circ} \text{ artículo es de calidad alta} \end{cases}$$

Entonces:

$$Y_{/A} \in Bernouilli(0,5) \quad \text{y} \quad Y_{/\bar{A}} \in Bernouilli(0,25)$$

y por tanto:

$$P(Y=0)=P(Y=0_{/A})\cdot P(A)+P(Y=0_{/\bar{A}})\cdot P(\bar{A})=$$
$$=0,5\cdot 0,9+0,75\cdot 0,1=0,525$$

Nos piden:

$$P(A_{/Y=0})=\frac{P(A\cap Y=0)}{P(Y=0)}=\frac{P(Y=0_{/A})\cdot P(A)}{P(Y=0)}=\frac{0,5\cdot 0,9}{0,525}=0,8571$$

85. Una persona apuesta a que tirando tres veces dos dados, obtendrá en una de ellas un total de 5 puntos y en otra un total de 7 puntos. ¿Qué probabilidad tiene de ganar la apuesta?

Solución:

Sea 5 = la suma de los puntos de los dados es 5.
7 = la suma de los puntos de los dados es 7.
* = la suma de los puntos de los dados no es 5 ni 7.

Se tiene:

$$P(5)=P(\{(1,\ 4),\ (2,\ 3),\ (3,\ 2),\ (4,\ 1)\})=\frac{4}{36}$$

$$P(7)=P(\{(1,\ 6),\ (2,\ 5),\ (3,\ 4),\ (4,\ 3),\ (5,\ 2),\ (6,\ 1)\})=\frac{6}{36}$$

$$P(*)=1-P(5)-P(7)=\frac{26}{36}$$

El jugador ganará la apuesta si el resultado de las tres tiradas es de uno de los tipos (5, 7, *) o (5, 7, 7) o (5, 5, 7) y cada uno de estos tipos en todas las formas posibles.

Número de casos de tipo (5, 7,*) = 3!

Número de casos de tipo (5, 7, 7) = 3

Número de casos de tipo (5, 5, 7) = 3

Por tanto:

$$P(\text{ganar apuesta}) = 6\frac{4}{36}\frac{6}{36}\frac{26}{36}+3\frac{4}{36}\frac{6}{36}\frac{6}{36}+3\frac{4}{36}\frac{4}{36}\frac{6}{36}=\frac{31}{324}=0,095$$

86. Supongamos que la distribución del número de defectos en un determinado rollo de tela sigue una distribución de Poisson con media 5 y que para una muestra aleatoria de 125 rollos se cuenta el nº de defectos de cada rollo. Calcular la probabilidad de que el número promedio de defectos por rollo en la muestra sea menor que 5,5.

Solución:

Sea X_i= número de defectos del rollo i-ésimo, i=1, 2, 3,.... $X_i \in$ P(5).

Nos piden $P(\bar{X} < 5,5)$, pero $\bar{X} = \frac{1}{125}\sum_{i01}^{125} X_i$.

Si llamamos $X' = \sum_{i=1}^{125} X_i$ entonces $\bar{X} = \frac{1}{125}X'$.

X' es suma de 125 variables independientes e idénticamente distribuidas, en virtud del teorema central del límite podemos considerar que sigue una distribución normal de parámetros:

$$E[X'] = E\left[\sum X_i\right] = \sum E[X_i] = 125\cdot 5$$
$$V[X'] = V\left[\sum X_i\right] = \sum V[X_i] = 125\cdot 5$$

Entonces tenemos que $\bar{X}$ también sigue una distribución normal y sus parámetros son:

$$E[\bar{X}] = E\left[\frac{1}{125}X'\right] = \frac{1}{125}E[X'] = 5$$

$$V[\bar{X}] = V\left[\frac{1}{125}X'\right] = \left(\frac{1}{125}\right)^2 V[X'] = \frac{1}{5} \Rightarrow \sigma_{\bar{X}} = \frac{\sqrt{5}}{5\sqrt{5}} = \frac{1}{5}$$

Estamos en condiciones de calcular la probabilidad pedida:

$$P(\overline{X} < 5,5) = P(\frac{\overline{X}-5}{0,2} < \frac{5,5-5}{0,2}) = P(Z < 2,5) = 0,9938$$

87. Sea X una variable aleatoria con media μ y desviación típica σ.

Calcular $E\lfloor (X\text{-}c)^2 \rfloor$ en función de μ y σ, siendo c una constante. ¿Qué valor de c hace mínima esa expresión?

Solución:

$$E\left[(X-c)^2\right] = E\left[X^2 - 2cX + c^2\right] = E\left[X^2\right] - 2cE[X] + c^2 =$$
$$= (E\left[X^2\right] - \mu^2) + \mu^2 - 2cE[X] + c^2 = \sigma^2 + \mu^2 - 2c\mu + c^2$$

Para buscar el valor de c que hace mínima esa expresión debemos derivarla respecto a c e igualarla a cero.

$$\frac{\partial E\left[(X-c)^2\right]}{\partial c} = -2\mu + 2c = 0 \Leftrightarrow \; c = \mu$$

y además $\dfrac{\partial^2 E\left[(X-c)^2\right]}{\partial c^2} = 2 > 0$ luego $c\;0\;\mu$ es efectivamente el valor de c que hace que $E[X]=\sigma^2$ que es su valor mínimo.

88. En un campeonato de tenis usted tiene la opción de escoger la secuencia de partidos A-B-A o la B-A-B donde A y B indican sus oponentes. Para clasificarse debe usted ganar dos partidos consecutivos. El jugador A es mejor que el B. ¿Qué secuencia elegiría usted?

Solución:

Sea A = usted gana a A, $p_A = P(A)$

B = usted gana a B, $p_B = P(B)$

Se tiene $p_A < p_B$ puesto que es más difícil ganar a A que a B

Serie A-B-A:

$$P(\text{usted gana el torneo}) = P(AB\bar{A},\ \bar{A}BA,\ ABA) =$$
$$= p_A p_B (1-p_A) + (1-p_A) p_B p_A + p_A p_B p_A = p_A p_B (2-p_A)$$

Serie B-A-B:

$$P(\text{usted gana el torneo}) = P(BA\bar{B},\ \bar{B}AB,\ BAB) =$$
$$= p_B p_A (1-p_B) + (1-p_B) p_A p_B + p_B p_A p_B = p_A p_B (2-p_B)$$

Dado que $p_A < p_B$ se tiene $2 - p_A > 2 - p_B$ y de aquí:

$$p_A\ p_B\,(2 - p_A) > p_A\ p_B\,(2 - p_B)$$

Por tanto, bebe elegir la serie A-B-A porque es con la que tiene mayor probabilidad de ganar el torneo.

89. Se admite que las retribuciones percibidas en una empresa siguen una distribución normal. Por las relaciones de los seguros sociales, se sabe que el 1% son superiores a 5 800 000 pts. y el 10% son inferiores a 1 200 000 pts. Calcular qué proporción de las retribuciones son superiores a 3 000 000 pts.

Solución:

Sea X = retribución de un trabajador (millones) $X \in N(\mu\ , \sigma)$.

Tenemos que $P(X>5{,}8) = 0{,}01$ y $P(X<1{,}2) = 0{,}1$.

$$0,01 = P(X > 5,8) = P(\frac{X-\mu}{\sigma} > \frac{5,8-\mu}{\sigma}) = P(Z > \frac{5,8-\mu}{\sigma}) \Rightarrow$$
$$\Rightarrow \frac{5,8-\mu}{\sigma} = 2,33$$

$$0,1 = P(X < 1,2) = P(\frac{X-\mu}{\sigma} < \frac{1,2-\mu}{\sigma}) = P(Z < \frac{1,2-\mu}{\sigma}) \Rightarrow$$
$$\Rightarrow \frac{1,2-\mu}{\sigma} = -1,28$$

$$\begin{cases} 5,8-\mu = 2,33\sigma \\ 1,2-\mu = -1,28\sigma \end{cases} \Rightarrow 4,6 = 3,61\sigma \Rightarrow \sigma = 1,274 \Rightarrow \mu = 2,831$$

$$P(X > 3) = P(\frac{X-2,831}{1,274} > \frac{3-2,831}{1,274}) = P(Z > 0,1326) = 0,4483$$

El 44,8% de los trabajadores de esa empresa cobra más de 3 000 000 pts.

90. En un aparato de control actúan dos variables, X e Y independientes, ambas con una distribución uniforme, la primera entre 1 y 9, la segunda entre 1 y *a*. El aparato funciona bien cuando $x<4y^2$. Calcular el valor de *a* para que:

$$P(X > 4Y^2) \leq 0,01.$$

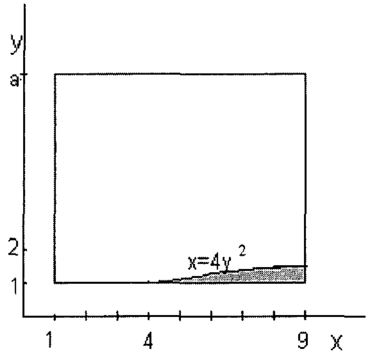
Solución:

$$X \in U(1,\ 9) \Rightarrow f(x) = \begin{cases} \frac{1}{8} & \text{si} \quad 1 \leq x \leq 9 \\ 0 & \text{en el resto} \end{cases}$$

$$Y \in U(1\ ,\ a) \Rightarrow f(y) = \begin{cases} \frac{1}{a-1} & \text{si} \quad 1 \leq y \leq a \\ 0 & \text{en el resto} \end{cases}$$

Por ser X e Y v. a. independientes, la variable aleatoria bidimensional (X ,Y), tendrá como función de densidad conjunta:

$$f(x\ ,\ y) = f(x)\ f(y) = \begin{cases} \dfrac{1}{8(a-1)} & \text{si}\quad 1 \leq x \leq 9,\quad 1 \leq y \leq a \\ 0 & \text{en el resto} \end{cases}$$



El rectángulo de la figura, cuyo borde superior es todavía indeterminado, es el dominio de la variable bidimensional. El volumen determinado por la densidad conjunta sobre la zona sombreada, que corta al rectángulo en los puntos (4 , 1) y (9 , 1.5) es la probabilidad pedida.

$$0{,}01 \geq \int_{y=1}^{1,5} \left(\int_{x=4y^2}^{9} \frac{1}{8(a-1)}\ dx \right) dy = \int_{y=1}^{1,5} \frac{1}{8(a-1)} \left[x\right]_{4y^2}^{9} dy =$$

$$= \int_{1}^{1,5} \frac{1}{8(a-1)} (9-4y^2) dy = \frac{1}{8(a-1)} \left[9y - \frac{4}{3} y^3 \right]_1^{\frac{3}{2}} =$$

$$= \frac{1}{8(a-1)} \left(\frac{27}{2} - \frac{4}{3}\frac{27}{8} - 9 + \frac{4}{3} \right) = \frac{1}{8(a-1)} \frac{81-27-54+8}{6} =$$

$$= \frac{1}{8(a-1)} \frac{8}{6} = \frac{1}{6(a-1)} \quad \Rightarrow \quad 0{,}01 \geq \frac{1}{6(a-1)} \quad \Rightarrow \quad a \geq 18{,}83$$

Análogamente se podría haber resuelto calculando:

$$0,01 \geq \int_{x=4}^{9}\left(\int_{y=1}^{\sqrt{x}/2}\frac{1}{8(a-1)}dy\right)dx = \int_{x=4}^{9}\frac{1}{8(a-1)}\left[y\right]_1^{\sqrt{x}/2}dx =$$

$$= \int_{x=4}^{9}\frac{1}{8(a-1)}\left(\frac{\sqrt{x}}{2}-1\right)dx = \frac{1}{8(a-1)}\left[\frac{1}{3}\sqrt{x^3}-x\right]_4^9 =$$

$$= \frac{1}{8(a-1)}\left(\frac{27}{3}-9-\frac{8}{3}+4\right) = \frac{1}{8(a-1)}\frac{4}{3} = \frac{1}{6(a-1)}$$

Como vemos esto nos lleva a resolver la misma desigualdad anterior, como debía ocurrir.

91. Una moneda cuyas caras están marcadas con un 2 y un 3, respectivamente, se tira 5 veces. La probabilidad de salir la cara que tiene el 2 es 0,6. ¿Cuál es la probabilidad de que la suma de los puntos obtenidos en las 5 tiradas sea 12?

Solución:

La única manera de sumar 12 con 5 tiradas de esta moneda es saliendo tres veces un dos y dos veces un tres.

Sea X = número de doses en los 5 lanzamientos $X \in B$ (5 , 0,6) por tanto:

$$P(\text{sumar doce}) = P(X=3) = \binom{5}{3} 0,6^3\, 0,4^2 = 0,3456$$

92. Un portamonedas contiene cuatro monedas que sólo pueden ser pesetas o duros. Se sacan dos monedas y resultan ser dos pesetas. Si las monedas se vuelven a guardar en el portamonedas, y a continuación se extrae de nuevo una moneda, ¿cuál es la probabilidad de que sea un duro?

Solución:

Puesto que sabemos que dos de las monedas son pesetas, sólo podemos estar en uno de los tres casos siguientes:

A = el monedero contiene 4 pesetas.

B = el monedero contiene 3 pesetas y 1 duro.

C = el monedero contiene 2 pesetas y 2 duros.

Consideramos estos tres casos equiprobables, pues no tenemos ninguna información que nos haga suponer lo contrario.

Sea D = sacar un duro en la segunda extracción.

$$P(D) = P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) =$$
$$= 0 \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} + \frac{2}{4} \cdot \frac{1}{3} = \frac{3}{12} = \frac{1}{4} = 0,25$$

93. Una empresa fabrica ascensores con capacidad para 16 personas. Cuando el peso de las personas que pretenden utilizar el ascensor excede del que puede elevar el motor, suena una alarma ordenando el desalojo. Si el peso de una persona (en kg) se distribuye como una normal de media 74 kg y varianza de 144, ¿cuál debe ser la potencia del motor para que la alarma suene no más del 1% de las veces en que 16 personas ocupan el ascensor?

Solución:

Sea X_i = peso de la persona i-ésima $X_i \in N(74, 12)$.

Sea $X = \sum_{i=1}^{16} X_i$ = peso de las 16 personas,

$$X \in N(n\mu, \sigma\sqrt{n}) = N(1184, 48)$$

Tenemos que encontrar un valor A tal que:

$$P(X > A) = 0,01$$

$$0,01 = P(X > A) = P\left(\frac{X - 1184}{48} > \frac{A - 1184}{48}\right) = P\left(Z > \frac{A - 1184}{48}\right) \Rightarrow$$
$$\Rightarrow \frac{A - 1184}{48} = 2,33 \Rightarrow A = 1184 + 2,33 \cdot 48 = 1295,84$$

94. Sea X una variable exponencial.

a) ¿Cuál es la probabilidad de que X tome un valor mayor que su media?

b) ¿Cuáles son las probabilidades de que X tome un valor que se encuentre como máximo a una distancia de una desviación estándar de la media, primero, y a dos desviaciones de la media, después?

Solución:

Si es una v. a. exponencial su función de densidad viene dada por:

$$f(x)=\begin{cases}\lambda\ e^{-\lambda x} & \text{si}\ \ x>0\\ 0 & \text{en el resto}\end{cases} \qquad \text{siendo } E[X]=\frac{1}{\lambda}$$

a) $P(X>E[X])=P(X>\frac{1}{\lambda})=1-P(X\le\frac{1}{\lambda})=\int_0^{1/\lambda}\lambda e^{-\lambda x}dx=$

$$=\left[-e^{-\lambda x}\right]_{x=0}^{1/\lambda}=e^{-\frac{2}{\lambda}}$$

b) Sabemos que $V[X]=\frac{1}{\lambda^2}\Rightarrow\sigma_x=\frac{1}{\lambda}$

$$P(\frac{1}{\lambda}-\frac{1}{\lambda}<X<\frac{1}{\lambda}+\frac{1}{\lambda})=\left[-e^{-\lambda x}\right]_{x=0}^{2/\lambda}=-e^{-2}+1=0{,}86$$

$$P(\frac{1}{\lambda}-\frac{2}{\lambda}<X<\frac{1}{\lambda}+\frac{2}{\lambda})=\left[-e^{-\lambda x}\right]_{0(*)}^{3/\lambda}=-e^{-3}+1=0{,}95$$

(*) Recordemos que la exponencial es positiva luego la variable no puede distar de su media por la izquierda dos desviaciones típicas.

95. Sean X e Y dos variables aleatorias continuas con función de densidad conjunta:

$$f(x\ ,\ y)=\begin{cases}x\ \ e^{-x(y+1)} & \text{si}\ \ x,\ \ y>0\\ 0 & \text{en otro}\ \ \text{caso}\end{cases}$$

a) Demostrar que f(x , y) es función de densidad.
b) Calcular P(X<2,Y<1).
c) Obtener las funciones de densidad marginales de X y de Y.
d) ¿Son X e Y estadísticamente independientes?

Solución:

a) $$\int_{x=0}^{\infty}\int_{y=0}^{\infty} x \quad e^{-x(y+1)} \quad dx \quad dy = \int_{0}^{\infty}\left[-e^{-x(y+1)}\right]_{y=0}^{\infty} dx = \int_{x=0}^{\infty} e^{-x} dx = = \left[-e^{-x}\right]_{0}^{\infty} = 1$$

b) $$P(X<2 \ , \ Y<1) = \int_{x=0}^{2}\int_{y=0}^{1} x \quad e^{-x(y+1)} dx \quad dy = \int_{x=0}^{2}\left[-e^{-x(y+1)}\right]_{y=0}^{1} dx =$$

$$= \int_{x=0}^{2}(e^{-x} + e^{-2x})dx = \left[-e^{-x}\right]_{x=0}^{2} + \left[-\frac{1}{2}e^{-2x}\right]_{x=0}^{2} =$$

$$= 1 - e^{-2} + \frac{1}{2} - \frac{1}{2}e^{-4} = \frac{3}{2} - e^{-2}(1 + \frac{1}{2}e^{-2})$$

c) $$f_1(x) = \int_{y=0}^{\infty} x \quad e^{-x(y+1)} dy = \left[- \quad e^{-x(y+1)}\right]_{y=0}^{\infty} = e^{-x}$$

$$f_2(y) = \int_{x=0}^{\infty} x\, e^{-x(y+1)} dx = \text{Por partes} = (*)$$

$$(*)\left\{\begin{array}{l} x = u \qquad \Rightarrow \qquad dx = du \\ e^{-x(y+1)} dx = dv \Rightarrow v = \int e^{-x(y+1)} dx = -\dfrac{1}{y+1} e^{-x(y+1)} \end{array}\right\}$$

$$(*) = \left[-\frac{x}{y+1} e^{-x(y+1)}\right]_{x=0}^{\infty} + \int_{x=0}^{\infty} \frac{1}{y+1} e^{-x(y+1)} dx =$$

$$= \left[-\frac{1}{(y+1)^2} e^{-x(y+1)}\right]_{x=0}^{\infty} = \frac{1}{(y+1)^2}$$

d) $$f_1(x) \cdot f_2(y) = e^{-x} \cdot \frac{1}{(y+1)^2} \neq f(x \quad , \quad y)$$

Las dos variables no son independientes.

96. Un fabricante de escapes para automóviles desea garantizar su producto durante un tiempo igual a la vida del vehículo. El fabricante supone que el tiempo de duración de su producto es una variable aleatoria con una distribución normal, con una vida promedio de tres años y una desviación estándar de seis meses. Si el costo de reemplazo de una unidad es de 10 $, ¿cuál puede ser el costo de reemplazo para dos años si se instalan 1 000 000 de unidades?

Solución:

Sea X = duración de un escape (años) $X \in N(3, 0{,}5)$.

$$P(X < 2) = P(\frac{X-3}{0,5} < \frac{2-3}{0,5}) = P(Z < -2) = 0,0228$$

Se espera que aproximadamente el 2,28% de los tubos de escape se rompan antes de los dos años. Por tanto el gasto será:

$$Gasto = 10 \times 10^6 \times 0{,}0228 = 228.000\$$$

97. Cierta enfermedad puede ser producida por tres tipos de virus, A, B y C. En un laboratorio hay tres tubos con el virus A, dos con el virus B y cinco con el virus C. La probabilidad de que el virus A produzca la enfermedad citada es $P\{|X| < 4\}$ siendo $X \in N(3,5)$. La probabilidad de que el virus B produzca la enfermedad es $P\{Y \leq 3\}$, siendo $Y \in B(5, 0{,}7)$. Por último la probabilidad de que el virus C produzca la enfermedad es $P\{W \leq 5\}$ siendo $W \in P(4)$ (variable de Poisson).

Se elige un tubo al azar y se inocula el virus a un animal, éste contrae la enfermedad; probabilidad de que el virus inoculado sea del tipo C.

Solución:

Sean A = elegir tubo con virus A P(A) = 0,3.

B = elegir tubo con virus B P(B) = 0,2.

C = elegir tubo con virus C P(C) = 0,5.

Sea D = contraer la enfermedad.

$$P(D/A) = P(|X| < 4) = P(-4 < X < 4) = P(\frac{-4-3}{5} < \frac{X-3}{5} < \frac{4-3}{5}) =$$
$$= P(-1,4 < Z < 0,2) = 0,5793 - 0,0808 = 0,4985$$

$$P(D/B) = P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - 0,1631 = 0,8369$$

$$P(D/C) = P(W \leq 5) = 0,7851$$

De donde:

$$P(C/D) = \frac{P(D/C) \cdot P(C)}{P(D/A) \cdot P(A) + P(D/B) \cdot P(B) + P(D/C) \cdot P(C)} =$$
$$= \frac{0,7851 \cdot 0,5}{0,4985 \cdot 0,3 + 0,8369 \cdot 0,2 + 0,7851 \cdot 0,5} = 0,5533$$

98. La calificación promedio en una prueba de Estadística fue de 62,5 con una desviación estándar de 10. El profesor sospecha que el examen fue difícil. De acuerdo con lo anterior desea ajustar las calificaciones de manera que el promedio sea de 70 y la desviación estándar de 8. ¿Qué ajuste de tipo aX+b debe hacer?

Solución:

Tenemos X = calificación en el examen. Sabemos que:

$$\begin{cases} E[X] = 62,5 \\ \sigma_X = 10 \end{cases}$$

Buscamos una variable $X' = aX + b$, tal que:

$$\begin{cases} E[X'] = 70 \\ \sigma_{X'} = 8 \end{cases}$$

De las propiedades de las esperanzas y varianzas tenemos:

$$E[X'] = E[aX+b] = a \quad E[X]+b = 62,5 \quad a+b \Rightarrow \ 62,5 \quad a+b = 70$$
$$V[X'] = V[aX+b] = a^2 V[X] = a^2 \quad 100 \ \Rightarrow \ 100a^2 = 64$$

Resolviendo el sistema tenemos $a = 0.8$, $b = 20$

Luego la nueva calificación X' se obtendrá multiplicando la calificación antigua por a = 0,8 y sumándole b = 20.

99. Se lanza una moneda, si sale cara se echan *a* bolas blancas en una urna y si sale cruz son *2a* bolas blancas las que se ponen en la urna. El mismo procedimiento se sigue con una segunda tirada, poniendo *b* bolas negras si sale cara y *2b* bolas negras si sale cruz. De la urna así compuesta se extrae una bola al azar. ¿Cuál es la probabilidad de que sea negra?

Solución:

Al lanzar la moneda dos veces los resultados posibles son, según sabemos: CC, C+, +C, CC. Cada uno de ellos tendrá 0,25 de probabilidad de ocurrir puesto que debemos suponer la moneda equilibrada. Según el posible resultado de los lanzamientos de la moneda, será la composición de la urna:

Moneda: CC	Urna:	a bolas blancas, b bolas negras.
C+		a bolas blancas, 2b bolas negras.
+C		2a bolas blancas, b bolas negras.
++		2a bolas blancas, 2b bolas negras.

Por tanto:

$$P(N) = P(N/cc)P(cc) + P(N/c+)P(c+) + P(N/+c)P(+c) +$$
$$+P(N/++)P(++) = \frac{b}{a+b}\cdot 0,25 + \frac{2b}{a+2b}\cdot 0,25 + \frac{b}{2a+b}\cdot 0,25 +$$
$$+\frac{2b}{2a+2b}\cdot 0,25 = 0,25 \quad b \quad \left(\frac{2}{a+b} + \frac{2}{a+2b} + \frac{1}{2a+b}\right)$$

100. El peso de cereal que contiene una caja se aproxima a una distribución normal con una media de 600 g. El proceso de llenado de las cajas está diseñado para que, de entre 100 cajas, el peso de una se encuentre fuera del intervalo 590-610 gramos. ¿Cuál es el valor máximo de la desviación estándar para alcanzar este requerimiento?

Solución:

Sea X = peso de cereal de una caja, $X \in N(600, \sigma)$.

Queremos que $P(590 < X < 610) = 0{,}99$ y tenemos que determinar la desviación típica.

$$P(590 < X < 610) = P\left(\frac{590-600}{\sigma} < Z < \frac{610-600}{\sigma}\right) = 0{,}99$$

$$\Rightarrow\ P\left(Z < \frac{-10}{\sigma}\right) = 0{,}005 \Rightarrow \frac{-10}{\sigma} = -2{,}575 \Rightarrow\ \sigma = 3{,}88$$

101. Los jurados para una corte federal de distrito se seleccionan de manera aleatoria entre la lista de votantes del distrito. En un determinado mes se selecciona una lista de 25 candidatos. Ésta contiene los nombres de 20 hombres y cinco mujeres.

a) Si la lista de votantes se encuentra igualmente dividida por sexo, ¿cuál es la probabilidad de obtener una lista de 20 hombres y cinco mujeres?

b) ¿Cuál es la probabilidad de elegir de entre esa lista un jurado de 12 personas, de las cuales sólo una sea mujer?

c) Si el alumno fuera el abogado de la defensa, ¿qué podría argumentar mediante el empleo de las respuestas de las partes a) y b)?

Solución:

a) Consideramos que la elección de los 25 miembros del jurado consiste en realizaciones independientes por suponer que la población entre la que se elige es suficientemente numerosa, por tanto si X = número de varones de entre las 25 personas elegidas, será $X \in B(25\ ,\ 0{,}5)$ y por tanto:

$$P(X=20)=\binom{25}{20}0,5^{25}=1,28\cdot 10^{-28}$$

Observamos que la probabilidad de tal composición es remota, lógico puesto que partimos de una población de mitad hombres y mitad mujeres.

b) Supongamos que ya ha ocurrido el hecho poco probable de que hayan sido elegidos 20 hombres y 5 mujeres, ahora debemos formar el jurado con sólo 12 de ellos.

Sea Y = nº de mujeres entre los 12 seleccionados, $Y \in$ *Hg* (25,5;12) porque las realizaciones ya no son independientes ya que se escogen entre un grupo cerrado de 25 personas de las cuales sabemos que sólo 5 son mujeres, y se tiene:

$$P(Y=1)=\frac{\binom{5}{1}\cdot\binom{20}{11}}{\binom{25}{12}}=0,16$$

c) $P(X=20,\ Y=1)=P(Y=1_{/X=20})P(X=20)=$

$$=016\cdot 1,28\cdot 10^{-28}=2,05\cdot 10^{-29}$$

La probabilidad de que ocurran las dos cosas simultáneamente es remota: el defensor debe recusar al jurado.

102. Se tiene una moneda con una probabilidad de 2/3 de que al tirarla el resultado sea cara. Se lanza la moneda: si el resultado es cara, se saca una bola aleatoriamente de una urna que contiene dos bolas rojas y tres verdes; si el resultado es cruz se saca una bola de otra urna que contiene dos bolas rojas y dos verdes. ¿Cuál es la probabilidad de sacar una pelota roja?

Solución:

Sean C = salir cara. P(C) = 2/3.

R = salir bola roja. P(R/C)=2/5 P(R/$\bar{C}$) = 1/2.

$$P(R)=P(R/C)\cdot P(C)+P(R/\bar{C})\cdot P(\bar{C})=\frac{2}{5}\cdot\frac{2}{3}+\frac{1}{2}\cdot\frac{1}{3}=\frac{13}{30}$$

103. ¿Coinciden la media y la mediana en una variable que presenta una distribución exponencial?

Solución:

Sabemos que la media de la distribución exponencial es la inversa del parámetro que denotamos λ, o sea 1/λ.

La mediana es el valor de la variable que deja a su izquierda la mitad de la población, es decir:

$$0,5 = P(X < M_d) = \int_{x=0}^{M_d} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{x=0}^{M_d} = 1 - e^{-\lambda M_d} \Rightarrow$$

$$\Rightarrow e^{-\lambda M_d} = 0,5 \Rightarrow -\lambda M_d = Ln \ \ 0,5 \Rightarrow M_d = -\frac{Ln 0,5}{\lambda} = \frac{0,69}{\lambda}$$

Es decir: $M_d = 0{,}69E[X]$

Esto pone de manifiesto que la media y la mediana en una exponencial no coinciden.

104. Considérense las dos apuestas siguientes:

a) Se tira 50 veces una moneda equilibrada, usted gana si el número de caras oscila entre 20 y 30; pierde en caso contrario.

b) Se tira 5000 veces la misma moneda, usted gana si el número de caras oscila entre 2000 y 3000; pierde en caso contrario.

Respóndase razonada y detalladamente qué apuesta prefiere y por qué.

Solución:

Está claro que debemos decidir comparando las probabilidades de ganar la apuesta en cada uno de los dos casos:

a) Sea X = número de caras cuando tiro la moneda 50 veces, $X \in B(50, 0,5)$

$$P(20 \leq X \leq 30) = P(19{,}5 \leq X' \leq 30{,}5) = P(\frac{19{,}5-25}{\sqrt{12{,}5}} < Z < \frac{30{,}5-25}{\sqrt{12{,}5}}) =$$

$$= P(-1{,}55 < Z < 1{,}55) = 1 - 0{,}1212 = 0{,}8788$$

Siendo,

$$X \cong X' \in N(50\cdot 0,5,\ \sqrt{50\cdot 0,5\cdot 0,5}) = N(25,\ \sqrt{12,5})$$

b) Sea X = número de caras cuando tiro 5000 veces la moneda, $X \in B(5000,\ 0,5)$.

$$P(2000 \le X \le 3000) = P(1999,5 < X' < 3000,5) =$$
$$= P(\frac{1999,5-2500}{\sqrt{1250}} < Z < \frac{3000,5-2500}{\sqrt{1250}}) =$$
$$= P(-14,15 < X < 14,5) \approx 1$$

Siendo,

$$X \cong X' \in N(5000\cdot 0,5\ ,\ \sqrt{5000\cdot 0,5\cdot 0,5}) = N(2500\ ,\sqrt{1250})$$

Está claro que en la segunda opción prácticamente hay seguridad de ganar. El número de tiradas influye proporcionalmente sobre la oscilación (desviación) de los resultados.

105. En una oficina de empleo se aplica un test a los aspirantes a cierto puesto de trabajo. De los 250 presentados, 210 obtienen calificación inferior a 5,5. Se sabe que la distribución de las calificaciones es normal con desviación típica 1. Si hay 150 plazas, ¿cuál es la mínima puntuación que tiene que alcanzar un candidato para obtener una plaza?

Solución:

Sea X = puntuación de un aspirante, $X \in N(\mu\ ,\ 1)$.

Según lo que nos dice el enunciado se tiene:

$$P(X<5,5) = 210/250 = 0,84$$

La proporción de alumnos que pueden ser aprobados es 150/250=0,6, que será el 60% que mayor calificación haya sacado.

O sea queremos calcular la calificación A, tal que:

$$P(X<A) = 0,4$$

$$0,4 = P(X < A) = P(X - \mu < A - \mu) = P(Z < A - \mu) = 0,4 \Rightarrow$$
$$\Rightarrow A - \mu = -0,25$$

Pero podemos calcular μ:

$$0,84 = P(X < 5,5) = P(X - \mu < 5,5 - \mu) = P(z < 5,5 - \mu) \Rightarrow$$
$$\Rightarrow 5,5 - \mu = 0,99 \Rightarrow \mu = 5,5 - 0,99 = 4,01$$

Por tanto aprobarán la oposición los alumnos con nota a partir de:

$$A = \mu - 0,25 = 3,76$$

106. Consideremos una situación ideal de la Bolsa. Al empezar un año parte de 100 enteros y la probabilidad de que suba un entero en un día es de un tercio y de que baje un entero de dos tercios. Supongamos que las alzas y bajas son independientes de un día a otro. Sea X la variable que da la posición de la Bolsa después de cuatro sesiones consecutivas. Se pide:

a) Hallar la distribución de X.

b) Calcular su esperanza.

Solución:

Sea X = valor de la Bolsa después del 4º día.

a) Los valores que puede tomar X son:

X = 104 (si sube los 4 días).
X = 102 (si sube 3 días y baja 1 día).
X= 100 (si sube 2 días y baja 2 días).
X = 98 (si sube 1 día y baja 3 días).
X = 96 (si baja los 4 días).

Sea Y = número de días que sube la Bolsa, de los 4 en estudio, $Y \in B(4, {}^1/_3)$.

$$P(X=104)=P(Y=4)=\left(\tfrac{1}{3}\right)^4$$

$$P(X=102)=P(Y=3)=\binom{4}{3}\left(\frac{1}{3}\right)^3\frac{2}{3}=\frac{8}{3^4}$$

$$P(X=100)=P(Y=2)=\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2=\frac{24}{3^4}$$

$$P(X=98)=P(Y=1)=\binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3=\frac{32}{3^4}$$

$$P(X=96)=P(Y=0)=\left(\frac{2}{3}\right)^4=\frac{16}{3^4}$$

b) La esperanza de un v. a. discreta viene dada por:

$$E[X]=\sum_k x_k\cdot P(X=x_k)$$

$$E[X]=\frac{1}{3^4}(104+102\cdot 8+100\cdot 24+98\cdot 32+96\cdot 16)=98,66$$

107. Sea una urna con 1 bola blanca, 2 rojas y 3 negras. Se extraen dos bolas sin remplazamiento. Denotamos X_1, X_2, X_3, las variables que representan el número de bolas blancas, rojas y negras, respectivamente:

a) Calcular las distribuciones marginales de las tres variables.

b) Calcular la esperanza y la desviación típica de cada una de las variables.

Solución:

X_1 = número de bolas blancas de las dos extraídas

X_2 = número de bolas rojas de las dos extraídas

X_3 = número de bolas negras de las dos extraídas

$X_1 = k$	$P(X_1 = k)$
0	$\binom{5}{2} \Big/ \binom{6}{2} = 2/3$
1	$1 \Big/ \binom{6}{2} = 1/3$

$X_2 = k$	$P(X_2 = k)$
0	$\binom{4}{2} \Big/ \binom{6}{2} = 6/15$
1	$2 \cdot 4 \Big/ \binom{6}{2} = 8/15$
2	$1 \Big/ \binom{6}{2} = 1/15$

$X_3 = k$	$P(X_3 = k)$
0	$\binom{3}{2} \Big/ \binom{6}{2} = 3/15$
1	$3 \cdot 3 \Big/ \binom{6}{2} = 9/15$
2	$\binom{3}{2} \Big/ \binom{6}{2} = 3/15$

3) $E[X_1] = \dfrac{1}{3}$; $V[X_1] = \dfrac{3}{9} - \dfrac{1}{9} = \dfrac{2}{9} \Rightarrow \sigma_{X_1} = \dfrac{\sqrt{2}}{3}$

$E[X_2] = \dfrac{2}{3}$; $V[X_2] = \dfrac{12}{15} - \dfrac{4}{9} = \dfrac{16}{45} \Rightarrow \sigma_{X_2} = \dfrac{4}{3\sqrt{5}}$

$E[X_3] = 1$; $V[X_3] = \dfrac{22}{15} - 1 = \dfrac{7}{15} \Rightarrow \sigma_{X_3} = \sqrt{\dfrac{7}{15}}$

108. En una mesa de juego, en 1654, Meré propuso a Pascal la siguiente afirmación: “Es más probable obtener al menos un as en una sola tirada con cuatro dados, que al menos un doble as en veinticuatro tiradas con dos dados". Demostrar que Meré tenía razón.

Solución:

X = número de ases al tirar cuatro dados $X \in B(4, 1/6)$.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \left(\frac{5}{6}\right)^4 = 0,5177$$

Y = número de veces que sale un doble as en 24 tiradas con dos dados $Y \in B(24, 1/36)$.

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \left(\frac{35}{36}\right)^{24} = 0,4914$$

Por tanto Meré tenía razón.

109. Supongamos que la distribución de estaturas, en cm, de los jóvenes en edad militar en cierto país es una v. a. con distribución N (175,100) y que son rechazados aquellos cuya estatura es inferior a 155 cm o superior a 195 cm. ¿Cuál es la probabilidad de que un soldado en la mili mida entre 172 y 181 cm?

Solución:

X = estatura de un joven en edad militar, $X \in N(175, 100)$.

P(un joven en edad militar es aceptado en la mili)=

$$= P(155 < X < 195) = P(-0,2 < Z < 0,2) == 1 - 0,8414 = 0,1586$$

$$P(172 < X < 181 / \text{está en la mili}) = \frac{P(172 < X < 181)}{0,1586} = \frac{0,0359}{0,1586} = 0,2263$$

110. Un individuo tiene un aparato que necesita para funcionar una válvula C cuya duración en horas viene dada por la función de densidad:

$$f(x)=\frac{1}{10}e^{-\frac{x}{10}},\quad x\geq 0$$

Tiene dos válvulas de ese tipo y quiere saber cuál es la probabilidad de que el aparato funcione más de 21 horas, suponiendo que los demás componentes del sistema tienen garantizada una duración de al menos 100 horas y que si se estropea una válvula empieza a funcionar la otra automáticamente.

Solución:

Nos piden $P(X + Y > 21)$ siendo:

X = duración de la primera válvula

Y = duración de la segunda válvula.

Como las válvulas son independientes la variable bidimensional (X, Y) tiene como función de densidad conjunta el producto de las marginales.

$$f(x,\ y)=\frac{1}{10}e^{-\frac{x}{10}}\frac{1}{10}e^{-\frac{y}{10}}=\frac{1}{100}e^{-\frac{x+y}{10}},\quad x,y\geq 0$$

$$P(X+Y>21)=P(X\leq 21,\ Y\geq 21-X)+P(X>21)=$$

$$=\int_0^{21}\left(\int_{21-x}^{\infty}\frac{1}{100}e^{-\frac{x+y}{10}}dy\right)dx+\int_{21}^{\infty}\frac{1}{10}e^{-\frac{x}{10}}dx=$$

$$=2,1e^{-2,1}+e^{-2,1}=0,257+0,122=0,379$$

111. Un profesor realiza un test con diez ítems a una clase, teniendo cada ítem cuatro posibles respuestas: A, B, C y D, de las cuales sólo una es correcta. Suponiendo que un alumno no se ha preparado el test y tiene la misma probabilidad de responder A, B, C o D, hallar la probabilidad de los siguientes sucesos:

a) Que conteste todos los ítems mal.

b) Que conteste al menos 5 ítems bien.

c) Que el séptimo ítem sea el tercero que contesta bien.

d) ¿Cuántos ítems se espera que conteste bien?

Solución:

Sea X = número de ítems bien contestados de los 10, $X \in B$ (10, 0.25).

a) $P(X = 0) = (0.75)^{10} = 0{,}0563$

b) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0{,}9219 = 0.0781$

c) Y= número de ítems que contesta hasta contestar el tercero correcto,

$Y \in BN$ (3, 0.25) y por tanto:

$$P(Y=7) = \binom{6}{2} 0{,}25^3 0{,}75^4 = 0{,}07415$$

d) $E[X] = 2.5$

112. Un país está habitado por dos grupos étnicos: A (alpinos) y N (nórdicos), que se encuentran en las proporciones 0,75 y 0,25. Se sabe que la talla de los individuos adultos varones es N (170,7) en cm, para el grupo A y N (176,7) para el grupo N. Se dice que un individuo es "alto" si su talla es superior a 180 cm. Calcular:

a) La proporción de altos de los grupos A y N.

b) Probabilidad de que un individuo sea nórdico, si se sabe que es alto.

c) Probabilidad de que de seis individuos elegidos al azar, 4 de ellos sean "altos".

Solución:

a) Tenemos $P(A) = 0.75$ y $P(N) = 0.25$.

Sea X = altura de un individuo (cm), tenemos

$$X_{/A} \in N(170, 7) \text{ y } X_{/N} \in N(176, 7)$$

$$P(\text{alto}/A) = P(X > 180/A) = P\left(Z > \frac{180-170}{7}\right) = P(Z > 1{,}43) = 0.0764$$

$$P(\text{alto}/N) = P(X > 180/N) = P\left(Z > \frac{180-176}{7}\right) = P(Z > 0.57) = 0.2843$$

El 7.64% de los alpinos y el 28.43% de los nórdicos son "altos".

b) $$P(N/\text{alto}) = \frac{P(\text{alto}/N)P(N)}{P(\text{alto}/N)P(N) + P(\text{alto}/A)P(A)} = 0.5536$$

c) Sea Y = número de personas "altas" de las 6, $Y \in B(6, p)$, siendo $p=P(\text{alto})=0.1284$,

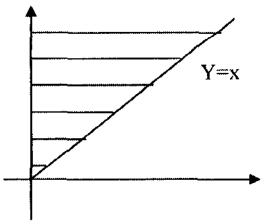
$$P(Y=4) = \binom{6}{4} 0.1284^4 0.8716^2 = 0.0031$$

113. Sea (X,Y) una v. a. bidimensional con función de densidad conjunta:

$$f(x,\ y) = \begin{cases} e^{-y} & \text{si } 0 < x < y < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Describir el recinto donde está definida la variable bidimensional. Calcular la covarianza de las dos variables y decir si son independientes.

Solución:



El recinto de definición de la variable es la parte del cuadrante positivo del plano XY por encima de la diagonal principal, puesto que Y se tiene que mantener siempre mayor que X. Es decir el área rayada de la figura.

Para calcular la covarianza:

$$E[X] = \int_{x=0}^{\infty} \int_{y=x}^{\infty} x e^{-y} dx\, dy = \int_{x=0}^{y} x\left[-e^{-y}\right]_{y=x}^{\infty} dx = \int_{x=0}^{\infty} x e^{-x} dx = 1$$

$$E[Y] = \int_{x=0}^{\infty} \int_{y=x}^{\infty} y e^{-y} dy\, dx = \int_{x=0}^{\infty} \left\{ \left[-y e^{-y}\right]_{y=x}^{\infty} + \int_{y=x}^{\infty} e^{-y} dy \right\} dx =$$

$$= \int_{x=0}^{\infty} \left(x e^{-x} + e^{-x}\right) dx = 2$$

$$E[XY] = \int_{x=0}^{\infty} \int_{y=x}^{\infty} x y e^{-y} dy\, dx = \int_{x=0}^{\infty} x \left\{ \int_{y=x}^{\infty} y e^{-y} dy \right\} dx = \int_{x=0}^{\infty} x\left(x e^{-x} + e^{-x}\right) dx$$

$$= \int_{x=0}^{\infty} x\left(x e^{-x} + e^{-x}\right) dx = \int_{0}^{\infty} x^2 e^{-x} dx + \int_{x0}^{\infty} x e^{-x} dx = 3$$

Tenemos entonces:

$$Cov\,[XY] = E[XY] - E[X]E[Y] = 3 - 2 = 1$$

Las variables no son independientes, cosa que esperábamos puesto que ya en el recorrido Y tiene que ser mayor que X, y por tanto depende de X.

$$f_1(x) = \int_{y=x}^{\infty} e^{-y}dy = \left[-e^{-y}\right]_x^{\infty} = e^{-x} \Rightarrow X \in Exp(1) \Rightarrow E[X] = 1$$

$$f_2(y) = \int_{x=0}^{y} e^{-y}dx = e^{-y}\left[x\right]_0^y = ye^{-y} \Rightarrow E[Y] = \int_0^{\infty} yf_2(y)dy = \int_0^{\infty} y^2e^{-y}dy = 2$$

114. Supóngase que el 75% de las personas de cierta área metropolitana viven en la ciudad y el 25% restante en las afueras. Si las 1200 personas que asisten a un concierto son consideradas una muestra aleatoria ¿cuál es la probabilidad de que el número de personas que asisten al concierto y que viven en las afueras sea menor de 270?

Solución:

Sea X = número de personas que viven en las afueras de las 1200 que asisten al concierto; $X \in B\,(1200,\, 0{,}25) \cong N\,(300,\, 15)$.

$$P(X < 270) = P\left(Z < \frac{-30}{15}\right) = P(Z < -2) = 0{,}0228$$

115. Una urna contiene 5 bolas cuyo color se desconoce. Se extrae una bola y se devuelve a la urna. La bola extraída es de color rojo. A continuación se sacan simultáneamente dos bolas. ¿Cuál es la probabilidad de que las dos bolas sean también rojas?

Solución:

Supongamos que n de las 5 bolas son rojas, n = 1,2,3,4,5. Supongamos que cada una de las posibles composiciones de la urna es tan probable como cualquier otra, es decir:

$$P(n=1)=P(n=2)=P(n=3)=P(n=4)=P(n=5)=1/5$$

$$P(RR/n)=\frac{\binom{n}{2}}{\binom{5}{2}}=\frac{n(n-1)}{20}$$

Por tanto:

$$P(RR)=\sum_{i=1}^{5}P(RR/n=i)\ P(n=i)=\frac{1}{5}(0+\frac{2}{20}+\frac{6}{20}+\frac{12}{20}+\frac{6}{20}+1)=0,26$$

116. Uno de cada cuatro habitantes de una región ha sido vacunado contra cierta enfermedad. En el curso de una epidemia se observa que entre los enfermos hay uno vacunado por cuatro no vacunados.

a) ¿Es la vacuna de alguna eficacia?

b) Se sabe, además, que hay un enfermo de entre doce vacunados. ¿Cuál es la probabilidad de caer enfermo de un individuo no vacunado?

Solución:

Sea $E =$ padecer la enfermedad y $V =$ estar vacunado. Se tiene

$$P(V) = 1/4,\ P(V/E) = 1/5.$$

a) Para estudiar si la vacuna es efectiva comparamos la probabilidad de padecer la enfermedad con la de padecerla estando vacunado.

$$P(E/V)=\frac{P(V/E)\ \ P(E)}{P(V)}=\frac{4}{5}P(E)$$

Vemos que la probabilidad de contraer la enfermedad estando vacunado es ligeramente inferior a la probabilidad general de contraer la enfermedad.

Además,

$$P(E/\bar{V})=\frac{P(\bar{V}/E)\ \ P(E)}{P(\bar{V})}=\frac{16}{15}P(E)$$

La probabilidad de coger la enfermedad no estando vacunado es ligeramente superior a la probabilidad general de contraer la enfermedad.

Por último,

$$\frac{P(E/\bar{V})}{P(E/V)} = \frac{4}{3}$$

y vemos que la vacuna es efectiva, pero no demasiado.

b) La información es que

$$P(E/V) = \frac{1}{12}$$

y de aquí por a)

$$P(E) = \frac{5}{48}$$

Nos pide

$$P(E/\bar{V}) = \frac{16}{15}\frac{5}{48} = \frac{1}{9}$$

117. Un test detecta la presencia de un cierto tipo T de bacterias en el agua con probabilidad 0,9 en el caso de haberlas. Si no las hay, detecta la ausencia con probabilidad 0,8. Sabiendo que la probabilidad de que una muestra de agua contenga bacterias tipo T es 0,2, calcular la probabilidad:

a) De que haya realmente bacterias T, cuando el test ha dado positivo.

b) Ídem cuando el resultado del test es negativo.

c) De que haya bacterias y el test sea positivo.

d) De que no haya bacterias y el test dé negativo.

Solución:

A = el test da positivo, B = hay bacterias T en el agua

$$P(A/B) = 0{,}9 \qquad P(\bar{A}/\bar{B}) = 0{,}8 \qquad P(B) = 0{,}2$$

a) $$P(B/A)=\frac{P(A/B)P(B)}{P(A)}=\frac{P(A/B)P(B)}{P(A/B)P(B)+P(A/\overline{B})P(\overline{B})}=$$

$$=\frac{0,9\cdot 0,2}{0,9\cdot 0,2+0,2\cdot 0,8}=0.529$$

b) $$P(B/\overline{A})=\frac{P(\overline{A}/B)\ P(B)}{P(\overline{A})}=\frac{0,1\cdot 0,2}{0,66}=0,03$$

c) $P(A\cap B)=0,18$

d) $P(\overline{A}\cap\overline{B})=0,98$

118. Si se seleccionan 3 dígitos de una tabla de dígitos aleatorios. ¿Cuál es la probabilidad de que su suma sea 6?

Solución:

Entendemos que se pueden dar repeticiones, por ejemplo los tres dígitos elegidos podrían ser 6, 6 y 8, o también 3, 3 y 3. Cualquier terna es tan probable como cualquier otra. Contamos el número de casos favorables:

Terna	0, 0, 6	0, 1, 5	0, 2, 4	0, 3, 3	1, 1, 4	1, 2, 3	2, 2, 2
Nº casos	3	6	6	3	3	6	1

Por tanto:

$$P(\text{sumar seis})=\frac{\text{casos favorables}}{\text{casos posibles}}=\frac{28}{\mathrm{VR}_{10}^{3}}=0,028$$

Soluciones 2.

Inferencia estadística

1. Un metalúrgico ha realizado cuatro determinaciones del punto de fusión del manganeso: 1269°, 1265°, 1271° y 1263°. Según un manual el punto de fusión es 1260°. ¿Están de acuerdo los datos con el manual o debe aceptar que el punto de fusión es mayor?

Solución:

Sea X = temperatura de fusión del manganeso, como no tenemos más información que los cuatro datos de la muestra, supondremos que X sigue una distribución normal $N(\mu , \sigma)$.

$$\begin{cases} H_0 : \mu \leq 1260^{\circ} \\ H_1 : \mu > 1260^{\circ} \end{cases}$$

Si H_0 es cierta $X \in N(1260, \sigma)$, por tanto $\overline{X} \in N(1260, \sigma/2)$ y se tiene:

$$\frac{\overline{X}-1260}{S/2} \in t_3 \text{ y como } \begin{cases} \overline{X} = 1267 \\ S = 3,6515 \end{cases} \Rightarrow \frac{\overline{X}-1260}{S/2} = 3,834$$

$$RC_{0,05} = [2.353, +\infty) \quad RC_{0.01} = [4.541, \ +\infty)$$

CONCLUSIÓN: depende del nivel de significación que queramos:

Si nos basta con una confianza del 95%, podemos decir que hay suficiente evidencia para afirmar que el punto de fusión del manganeso es superior a 1260°.

Si queremos una confianza del 99%, debemos concluir que no hay suficiente evidencia para afirmar que el punto de fusión del manganeso no sea 1260º.

2. Las compañías de auditorías generalmente seleccionan una muestra aleatoria de los clientes de un banco y verifican los balances contables reportados por el banco. Si una compañía de este tipo se encuentra interesada en estimar la proporción de cuentas para las cuales existe una discrepancia entre el cliente y el banco, ¿cuántas cuentas deberán seleccionarse de manera tal que con una confiabilidad del 0,995, la proporción muestral se encuentre a no más de 0.02 unidades de la proporción real?

Solución:

Sea X = nº de cuentas de las n estudiadas en que los balances presentan anomalías, $X \in B(n, p)$ siendo $p = P$(una cuenta presenta anomalías); p es el parámetro a estimar y su estimador será:

$$\hat{p} = \frac{\text{n}^{\text{o}} \text{ de cuentas con anomalías}}{\text{n}^{\text{o}} \text{ de cuentas auditadas}}$$

El intervalo de confianza para p viene dado por

$$\mathrm{I}_{p,\alpha} = \left[\hat{p} \pm \frac{\hat{p}(1-\hat{p})}{\sqrt{n}} \cdot Z_{\alpha/2} \right] \quad \text{ó} \quad \mathrm{I}_{p,\alpha} = \left[\hat{p} \pm \frac{1}{2\sqrt{n}} \cdot Z_{\alpha/2} \right]$$

según tengamos o no algún valor para $\hat{p}$.

En este caso tenemos α=0.05 y $\frac{1}{2\sqrt{n}} \cdot Z_{\alpha/2} \leq 0.02$, de donde:

$$\frac{1}{2\sqrt{n}} \cdot 1.96 \leq 0.02 \quad \Rightarrow \quad \frac{1}{2}\left(\frac{1.96}{0.02}\right)^2 \leq n \quad \Rightarrow \quad n \geq 4802$$

3. Al terminar la Segunda Guerra Mundial se intentó dar razones científicas para decidir si el bombardeo de Londres por los proyectiles V2 había sido al azar o se había hecho contra lugares de interés militar. Con tal fin se cuadriculó el plano de Londres en 576 casillas y se contó el nº de casillas en que habían caído 0, 1, 2, etc., bombas V2, obteniéndose:

nº bombas V2	0	1	2	3	4
nº de casillas	229	211	93	35	8

A la vista de estos datos, decidir si las bombas obedecían a un plan de ataque a objetivos de interés militar o no.

Solución:

Sea X = nº de bombas que caen en una casilla. Si el bombardeo de Londres se hizo al azar X será una variable aleatoria, por tanto lo que vamos a hacer es tratar de buscar algún modelo estadístico y con un test de bondad de ajuste ver si se puede aceptar o no que X se distribuye según ese modelo. Dada la forma en que fue recogida la información, parece tener sentido que intentemos ajustar X a un modelo de Poisson. Así tenemos:

$$\begin{cases} H_0 : X \in P(\lambda) \\ H_1 : X \text{ no sigue ese modelo} \end{cases}$$

Lo primero será estimar λ:

$$\hat{\lambda} = \frac{0 \cdot 229 + 1 \cdot 211 + 2 \cdot 93 + 3 \cdot 35 + 4 \cdot 8}{576} = 0.927$$

Ahora debemos calcular las frecuencias esperadas, siendo

$$p_i = \frac{\lambda^i}{i!} e^{-\lambda} \text{ y } E_i = n \cdot p_i$$

$$p_0 = P(X = 0) = \frac{0.927^0}{0!} e^{-0.927} = 0.395 \Rightarrow E_0 = 227.946$$

$$p_1 = P(X = 1) = \frac{0.927^1}{1} e^{-0.927} = 0.367 \Rightarrow E_1 = 211.306$$

$$p_2 = P(X = 2) = \frac{0.927^2}{2!} e^{-0.927} = 0.170 \Rightarrow E_2 = 97.940$$

$$p_3 = P(X = 3) = \frac{0.927^3}{3!} e^{-0.927} = 0.052 \Rightarrow E_3 = 30.362$$

$$p_4 = P(X \geq 4) = 1 - P(X \leq 3) = 0.015 \Rightarrow E_4 = 8.545$$

El estadístico del contraste es

$$\sum \frac{(O_i - E_i)^2}{E_i} = \sum \frac{O_i^2}{E_i} - n = 1.030$$

y de aquí

$$RC_{0,05} = \left[X^2_{3,\ 0,05} \ ,\ +\infty \right) = \left[7,815 \ ,\ +\infty \right)$$

CONCLUSIÓN: la muestra se adapta muy bien al modelo de Poisson. Podemos decir que el bombardeo de Londres fue al azar.

4. Un agricultor quiere saber si hay diferencia entre las producciones de dos variedades de trigo, A y B. La tabla indica la producción por unidad de área de ambas variedades. ¿Puede concluirse que existe diferencia a los niveles de significación a) 0.01; b) 0.05?

trigo A	15.9 15.3 16.4 14.9 15.3 16.0 14.6 15.3 14.5 16.6 16.0
trigo B	16.4 16.8 17.1 16.9 18.0 15.6 18.1 17.2 15.4

Solución:

Sea X = producción de trigo A, Y = producción de trigo B.

Como no nos dicen nada del comportamiento de estas variables, usaremos un test no paramétrico: el test de la suma de rangos de Wilcoxon.

$$\begin{cases} H_0 : M_X = M_Y \\ H_1 : M_X \neq M_Y \end{cases}$$

Para aplicar este test, asignamos rango a cada uno de los 20 datos muestrales, conservando la muestra de la que proceden.

Rango	1	2	3	4	5	6	7	8	9	10
Valor	14.5	14.6	14.9	15.3	15.3	15.3	15.4	15.6	15.9	16.0
Muestra	A	A	A	A	A	A	B	B	A	A

Rango	11	12.5	12.5	14	15	16	17	18	19	20
Valor	16.0	16.4	16.4	16.6	16.8	16.9	17.1	17.2	18.0	18.1
Muestra	A	A	B	A	B	B	B	B	B	B

$$W_1 = 77.5 \quad n_1 = 11 \Rightarrow U_1 = W_1 - \frac{n_1(1+n_1)}{2} = 77{,}5 - 66 = 11{,}5$$

$$W_2 = 132.5 \; n_2 = 9 \quad \Rightarrow U_2 = W_2 - \frac{n_2(1+n_2)}{2} = 132{,}5 - 45 = 87{,}5$$

Tomamos U= min{ U_1, U_2}=11,5. Para buscar la región crítica recordemos que en la tabla entran los tamaños muestrales de menor a mayor, es decir el tamaño más pequeño es denotado por n_1, y el mayor por n_2. Como el test es bilateral no tenemos tabla para α=0,01 sino para α=0,02.

a) $RC_{0,02}$ = [0,18]

b) $RC_{0,05}$ = [0,23]

CONCLUSIÓN: hay evidencia estadística de que las medianas de la dos producciones de trigo son distintas a ambos niveles de significación pedidos.

5. La tabla adjunta muestra datos referentes a horas trabajadas en un taller y nº de unidades producidas:

horas	80	79	83	84	78	60	82	85	79	84	80	62
unidades	300	302	315	330	300	250	300	340	315	330	310	240

a) Estudiar la recta de regresión y decir si el modelo lineal es suficientemente bueno para relacionar las variables.

b) ¿Cuántas unidades se espera producir con 75 horas de trabajo?

c) ¿Se puede admitir, a nivel de significación 0.05, que la pendiente de la recta de regresión es mayor que 2.5?

Solución:

Tenemos un problema de regresión lineal simple.

$$\left.\begin{array}{l}\sum x_i = 936 \\ \sum x_i^2 = 73760\end{array}\right\} S_{xx} = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n} = 752$$

$$\left.\begin{array}{l}\sum y_i = 3632 \\ \sum y_i^2 = 1.109.254\end{array}\right\} S_{yy} = \frac{n\sum y_i^2 \_ \left(\sum y_i\right)^2}{n} = 9.968{,}66$$

$$\sum x_i y_i = 285.908 \} S_{xy} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n} = 2.612$$

a)

$$b = \frac{S_{xy}}{S_{xx}} = 3{,}473$$

$$a = \frac{\sum y_i}{n} - b\frac{\sum x_i}{n} = 31{,}741$$

recta de regresión:

$\overline{y}_x = 31{,}772 + 3{,}473 \cdot x$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = 0{,}9534 \Rightarrow r^2 = b\frac{S_{xx}}{S_{yy}} = 0{,}909 \approx 0.91$$

Podemos decir que el 91% de la variabilidad de *Y* se explica por su dependencia lineal de *X*, y por tanto que el modelo lineal es bueno para explicar la relación entre las dos variables.

b) Con 75 horas de trabajo se espera producir:

$$\overline{y}_{/75} = 31{,}772 \times 3{,}473 \cdot 75 = 272 \text{ unidades}$$

Podemos usar la recta de regresión para predicción porque hemos visto que el modelo lineal es suficientemente bueno y 75 es un valor que se encuentra dentro del campo de las observaciones de la muestra.

c)

$$\begin{cases} H_0 : \beta \leq 2{,}5 \\ H_1 : \beta > 2{,}5 \end{cases} \qquad b \in N(\beta, \frac{\sigma}{\sqrt{S_{xx}}})$$

Como σ es desconocida debemos estimarla por S.

$$S=\sqrt{\frac{SSE}{n-2}}=\sqrt{\frac{S_{yy}-bS_{xy}}{n-2}}=\sqrt{89,71}=9,47$$

$$\frac{b-\beta}{S/\sqrt{S_{xx}}}\in t_{n-2} \text{ y } \frac{b-\beta}{S/\sqrt{S_{xx}}}=\frac{3,473-2,5}{9,47/\sqrt{752}}=2,82$$

$$RC_{0,05}=\left[t_{10,\,0,05},\ +\infty\right)=\left[1,812,\ +\infty\right)$$

CONCLUSIÓN: como nuestro estadístico toma valor en la región crítica, podemos decir que a nivel 0,05 hay evidencia estadística de que la pendiente de la recta de regresión es mayor que 2,5.

6. Un técnico opina que introduciendo un nuevo tipo de maquinaria en cierto proceso de producción se disminuye sustancialmente el tiempo requerido en la misma. A causa del alto costo de mantenimiento, el empresario dice que salvo que se reduzca el tiempo de producción en más de 8 minutos, no vale la pena la inversión. Seis experiencias arrojan una disminución media de dicho tiempo de 8.4 minutos con desviación típica de 0.32. Contrastar la hipótesis de que el proceso merece ser renovado, con niveles de significación a) 0.01 y b) 0.05.

Solución:

Sea X= tiempo requerido para el proceso, $X\in N(\mu\ ,\ \sigma)$

Queremos estudiar si hay evidencia estadística de que el proceso ha reducido su duración en más de 8 minutos, por tanto el contraste adecuado es:

$$\begin{cases} H_0: \ \mu\leq 8 \\ H_1: \mu>8 \end{cases}$$

Si H_0 es cierta, entonces

$$\overline{X}\in N(8,\sigma)\Rightarrow \frac{\overline{X}-8}{S/\sqrt{6}}\in t_{6-1}$$

$$\frac{\overline{X}-8}{S/\sqrt{6}}=\frac{0.4\cdot\sqrt{6}}{0.32}=1.25\cdot\sqrt{6}=3.062$$

$$RC_{0,01}=\left[t_{5,\,0,01},\ +\infty\right)=\left[2,015\ ,\ +\infty\right)$$

$$RC_{0,05} = \left[t_{5,\,0,05}\ ,\ +\infty\right) = \left[3,365,\ +\infty\right)$$

CONCLUSIÓN:

1) A nivel 0,01 el estadístico del contraste no cae en la región crítica, por lo que debemos concluir que no merece la pena cambiar la maquinaria.
2) A nivel 0,05 el estadístico del contraste cae en la región crítica, por lo que debemos concluir que debe cambiar la maquinaria.

7. La tabla adjunta recoge una muestra de 40 notas de un examen de ámbito nacional. Contrastar, a un nivel de significación de 0.05, la hipótesis de que la nota mediana de todos los participantes es a) 66, b) 75.

71	67	55	64	82	66	74	58	79	61
78	46	84	93	72	54	78	86	48	52
67	95	70	43	70	73	57	64	60	83
73	40	78	70	64	86	76	62	95	66

Solución:

a) Como se trata de estudiar el valor de la mediana tenemos el contraste:

$$\begin{cases} H_0 : M_X = 66 \\ H_1 : M_X \neq 66 \end{cases}$$

Comparamos cada uno de los datos con la hipotética mediana, X_i-M_X, contando el nº de diferencias positivas $N^+ = 23$ y negativas $N^- = 15$ (dos de las diferencias son 0 y no las contamos)

Si H_0 fuera cierta

$$N^- \in B(38, 0.5) \approx N(19, 2\sqrt{38})$$

y se tiene

$$P\left(N^- \leq 15\right) = P\left(N^- \leq 15,5\right) = P\left(Z \leq \frac{15,5-19}{2\sqrt{38}}\right) = P\left(Z \leq -0,1838\right) = 0,43$$

A nivel de significación 0,05 se puede aceptar que la mediana es 66.

b) Realizando el mismo proceso sustituyendo 66 por 75, tenemos

$$N^+ = 13 \qquad N^- = 27$$

Si H_0 fuera cierta

$$N^+ \in B(40\ ,\ 0{,}5) \approx N(20\ ,\ \sqrt{10})$$

y por tanto:

$$P(N^+ \leq 13) = P(N^+ \leq 13{,}5) = P(Z \leq \frac{13{,}5 - 20}{\sqrt{10}}) = P(Z \leq -2{,}0554) \approx 0$$

A nivel de significación 0,05 debemos afirmar que la mediana no es 77.

8. En una investigación se midió el largo de las alas y de la lengua de 16 abejas, obteniéndose una correlación muestral r = 0.415. ¿Es significativa esta correlación, a nivel $\alpha = 0.01$?

Solución:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

El estadístico del contraste es

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 1{,}7066$$

que si la hipótesis nula fuera cierta, se comporta como una t_{n-2}.

$$RC_{0,01} = \left[t_{14;0,01}, +\infty\right) = \left[2{,}624, +\infty\right)$$

CONCLUSIÓN: como el estadístico no cae en la región crítica, debemos concluir que con la información de que disponemos y con una confianza del 99% no podemos afirmar que el coeficiente de correlación sea distinto de cero. No es significativa.

9. En el proceso de fabricación de un tipo de piezas se ha llegado a la conclusión de que el porcentaje de defectuosas es 2 ó 4. Se quiere contrastar la hipótesis de que el porcentaje es 2 tomando una muestra de tamaño 100, se rechaza la hipótesis si en la muestra hay más de tres defectuosas. ¿Qué valores tienen las probabilidades de error tipo I y tipo II?

Solución:

Sea p la verdadera proporción de defectuosas.

Sea X= número de defectuosas de las 100 observadas,

$$X \in B(100, p) \cong N\left(100p, 10\sqrt{p(1-p)}\right)$$

$$\begin{cases} H_0 : p = 0.02 \\ H_1 : p = 0.04 \end{cases} \qquad RC = [3.5, +\infty[$$

$$\alpha = P(\text{error tipo I}) = P(X \geq 3.5 / p = 0.02) = P\left(\frac{X-2}{1.4} \geq \frac{3.5-2}{1.4}\right) =$$
$$= P(Z \geq 1.07) = 0.1423$$

$$\beta = P(\text{error tipo II}) = P(X \leq 3.5 / p = 0.04) = P\left(\frac{\text{X-4}}{1.96} < \frac{3.5-4}{1.96}\right) =$$
$$= P(Z < 0.255) = 0.3994$$

10. Se espera que dos trabajadores produzcan en promedio, el mismo nº de unidades terminadas en el mismo tiempo. El nº de unidades terminadas por cada operario en cierta semana fue:

operario 1	12	11	18	16	13
operario 2	14	18	18	17	16

Si se supone que el nº de unidades terminadas diariamente por cada trabajador es una variable aleatoria normal y que ambas tienen la misma varianza, ¿hay alguna diferencia entre los trabajadores a nivel de significación α =0.1?

Solución:

Sea X = nº de unidades diarias del trabajador 1, $X \in N(\mu_1, \sigma)$.

Sea Y = nº de unidades diarias del trabajador 2, $Y \in N(\mu_2, \sigma)$.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

$$\overline{X} - \overline{Y} \in N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \Rightarrow \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = t_{n_1+n_2-2}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 5.65 \Rightarrow S_p = 2.37$$

$$\frac{14 - 16.6}{2.37\sqrt{\frac{1}{5} + \frac{1}{5}}} = -1.73$$

$$RC_{0.01} = \left(-\infty, -t_{8,0.05}\right] \cup \left[t_{8,0.05}, +\infty\right) = (-\infty, -1.86] \cup [1.86, +\infty)$$

CONCLUSIÓN: como el estadístico cae en la región crítica diremos que no tenemos evidencia estadística suficiente para afirmar que el rendimiento medio de los dos operarios es distinto.

11. Para estudiar las diferencias sobre la producción de remolacha de cuatro fertilizantes, se dispuso de cinco fincas, cada una de las cuales se dividió en cuatro parcelas del mismo tamaño y tipo. Los fertilizantes fueron asignados al azar en las parcelas de cada finca y el rendimiento en toneladas fue:

Se desea saber si hay diferencias entre los fertilizantes y entre las fincas prescindiendo del tipo de terreno.

		FERTILIZANTE			
		A	B	C	D
	1	2.1	2.2	1.8	2.1
	2	2.2	2.6	1.9	2.0
FINCA	3	1.8	2.7	2.6	2.2
	4	2.0	2.5	2.0	2.4
	5	1.9	2.8	1.9	2.1

Solución:

Puesto que no nos dicen que la v. a. rendimiento por parcela sea normal intentaremos estudiar si hay diferencias entre los tratamientos con un contrate no paramétrico para k muestras relacionadas: el test de Friedman.

Lo primero es asignar rango a los individuos dentro de cada bloque (las distintas fincas):

	A	B	C	D
1	2,5	4	1	2,5
2	3	4	1	2
3	1	4	3	2
4	1,5	4	1,5	3
5	1,5	4	1,5	3
	$R_1=9{,}5$	$R_2=20$	$R_3=8$	$R_4=12{,}5$

$$\begin{cases} H_0 : M_A = M_B = M_C = M_D \\ H_1 : \text{no ocurre } H_0 \end{cases}$$

El estadístico del contraste es

$$H = \left(\frac{12}{b \cdot k \cdot (k+1)} \sum R_i^2\right) - 3(k+1) \cdot b = 10{,}26$$

Si H_0 es cierta, entonces $H \in X^2_{k-1}$, en nuestro caso b=5, k=4 y tenemos:

$$RC_{0,05} = \left[7{,}815, +\infty\right);\ RC_{0,01} = \left[11{,}345, +\infty\right);\ RC_{0,10} = \left[6{,}251, +\infty\right)$$

CONCLUSIÓN: podemos decir que los abonos tienen efectos medianos diferentes a un nivel de significación de hasta 0,05, pero no a nivel altamente significativo (0,01).

12. Se ha dividido un terreno en 100 parcelas de 1 m^2 y se ha contado el número de semillas de cierta especie que han germinado espontáneamente en cada parcela, con los resultados siguientes:

nº de semillas germinadas	0	1	2	3	4	5
frecuencia	9	31	30	16	10	4

¿Puede admitirse que el nº de semillas germinadas por m^2 sigue una ley de Poisson? (Con nivel de significación $\alpha = 0.05$).

Solución:

Sea X = número de semillas que germinan en una parcela,

$$\begin{cases} H_0 : X \in P(\lambda) \\ H_1 : X \text{ no sigue ese modelo} \end{cases}$$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad E_i = np_i$$

$$\hat{\lambda} = \frac{\sum_i x_i p_i}{n} = \frac{0 \cdot 9 + 1 \cdot 31 + 2 \cdot 30 + 3 \cdot 16 + 4 \cdot 10 + 5 \cdot 4}{100} = 1.99$$

	Número de semillas germinadas					
	0	1	2	3	4	5 ó más
$p_i = P(X=i)$	0.1366	0.2720	0.2706	0.1795	0.0833	0.0517
$E_i = np_i$	13.66	27.20	27.06	17.95	8.93	5.17

$$\sum_{i=1}^{6} \frac{O_i^2}{E_i} - N = 103,88 - 100 = 3,05$$

$$RC_{0,05} = \left[\chi^2_{3;\,0,05}, \ +\infty \right) = \left[7,82, \ +\infty \right)$$

CONCLUSIÓN: como el estadístico del contraste no cae en la región crítica podemos aceptar que el nº de semillas germinadas por parcela es una v. a. que sigue una distribución de Poisson.

13. Una compañía de helados desea saber si debe enfocar la publicidad de un nuevo sabor sólo a un grupo determinado de edades o si la debe dirigir a todas las edades. Para hacer un estudio se establecen tres grupos de edades y se seleccionan aleatoriamente 20 personas de cada grupo. A cada uno se le da a probar el nuevo helado y se le pide que lo clasifique como 1) excelente, 2) normal o 3) malo.

a) ¿Qué procedimiento se debe usar? Describirlo con detalle.

b) ¿Cuál es la región crítica para $\alpha = 0.05$?

c) Si el estadístico del contraste toma el valor 0.5, ¿cuál es la conclusión y por qué?

d) ¿Tenemos alguna información sobre si el nuevo helado se venderá bien?

Solución:

a) Las respuestas deben presentarse en una tabla de contingencia 3x3 del tipo:

	excelente	normal	malo
grupo 1	O_{11}	O_{12}	O_{13}
grupo 2	O_{21}	O_{22}	O_{23}
grupo 3	O_{31}	O_{32}	O_{33}

En cada celda se escribirá la frecuencia de aparición de la respuesta sobre la calidad en el grupo de edad que corresponde al cruce de fila y columna. Los totales parciales de fila son prefijados, los tres iguales a 20. Se debe aplicar el test de homogeneidad de la X^2.

Las hipótesis del contraste son:

$$\begin{cases} H_0 : P(A_i / \text{grupo } j) = P(A_i / \text{grupo } k) \\ \qquad \forall A_i = \text{excelente, normal, malo } \forall j,k = 1,2,3 \\ H_1 : \text{ no ocurre } H_0 \end{cases}$$

Se calculan las frecuencias esperadas E_{ij}, con:

$$E_{ij}=\frac{\text{total fila}\cdot\text{total columna}}{\text{gran total}}=\frac{\text{total columna}}{3}$$

Una vez que se tenga la tabla de las frecuencias esperadas se puede calcular el valor que toma el estadístico del contraste

$$\sum_{i,j}\frac{O_{ij}^2}{E_{ij}}-n$$

donde n es el nº total de observaciones, que en este caso son 60.

b) La región crítica para un nivel de significación α es:

$$RC_\alpha=\left[\mathrm{X}^2_{4;\,\alpha},\ +\infty\right)$$

c) Si el valor del estadístico cae en la región crítica se concluirá que la calificación de los helados depende de la edad.

Para $\alpha = 0{,}05$ la región crítica es $[9{,}488,\ +\infty)$.

Si el estadístico toma el valor 0,5 no caerá en región crítica y se aceptará la hipótesis nula de que la aceptación de los helados es homogénea en los distintos grupos de edades.

d) En cuanto a si el nuevo helado se venderá bien podemos considerar la proporción de gente que lo ha calificado de excelente

$$\hat{p}=\frac{O_{11}+O_{21}+O_{31}}{60}$$

o bien considerar como estimación de p:

$$P(\text{excelente})=\frac{O_{11}}{20}P(\text{grupo 1})+\frac{O_{21}}{20}P(\text{grupo 2})+\frac{O_{31}}{20}P(\text{grupo 3})$$

La muestra no nos da ninguna estimación de las probabilidades de cada uno de los grupos, puesto que los totales parciales son prefijados.

14. En una fábrica de componentes electrónicos se sabe que la proporción usual de piezas defectuosas es del 10% y la máxima proporción aceptable es del 20%. Se establece la siguiente regla de decisión para control de calidad: se toma una muestra de 10 unidades y se cuenta el número de piezas defectuosas; si el nº de defectuosas es 3 o más se decide que la proporción real de defectuosas es mayor que el 10%, en otro caso se admite que la proporción real de tales piezas no sobrepasa el 10%.

a) Decir explícitamente cuáles son las hipótesis del contraste.

b) Calcular el error tipo I.

c) Calcular el error tipo II, cuando la verdadera proporción de defectuosas fuera del 20%.

d) Comentar si es buena esta regla de decisión.

Solución:

Sea X= número de defectuosas de las 10 observadas.

a) Debemos suponer que las realizaciones son independientes, por ejemplo suponiendo que el número de piezas producidas es suficientemente grande. En estas condiciones $X \in B\,(10\;, p)$ y el contraste adecuado será

$$\begin{cases} H_0 : p \leq 0.1 \\ H_1 : p > 0.1 \end{cases} \qquad RC = \{X \geq 3\}$$

b) $\alpha = P(\text{error tipo I}) = P\left(X \geq 3 \;\middle|\; p = 0.1\right) = 1 - \sum_{k=0}^{2} P\left(X = k\right) = 0.1399$

c) $\beta_{0.2} = P(\text{error tipo II}) = P\left(X < 3 \middle| p = 0.2\right) = \sum_{k=0}^{2} P(X = k) = 0.3020$

d) Tanto el error tipo I como el error tipo II son demasiado grandes: el 14% de los lotes de 10 piezas tendrán más de 2 defectuosas, aún cuando la verdadera proporción de defectuosas sea del 10%; además si la proporción de defectuosas fuera realmente del 20%, el 30% de los lotes de 10 piezas tendrían menos de 3 defectuosas, con lo que no se podría detectar que en esos lotes la proporción de defectuosas es inadmisible.

15. Una perfumería lleva a cabo la siguiente prueba: Se le pregunta a 15 clientes, elegidos aleatoriamente, cuál de dos fragancias que se le presentan prefiere. Sin saberlo los entrevistados se trata del mismo perfume, la única diferencia es que uno de los envases tiene una apariencia mucho más cara. El dueño piensa que los compradores tenderán a elegir el perfume del envase con mejor apariencia. Resulta que 10 de los entrevistados eligen ese perfume. ¿Tiene razón el dueño de la perfumería?

Solución:

Sea X = nº de personas de las 15 que prefieren el perfume A. Puesto que los dos envases, que denotamos A y B, contienen el mismo aroma la probabilidad de elegir uno u otro debería ser $1/2$ y sería $X \in B(15\ , p)$

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p \neq 1/2 \end{cases}$$

Como X=10, tenemos $P\ (X \geq 10) = 0.1509$

CONCLUSIÓN: no hay suficiente evidencia de que los clientes prefieren el envase con mejor apariencia, ni a nivel de significación 0,05 ni 0,01.

16. Los siguientes datos representan los tiempos de funcionamiento en horas de tres tipos de calculadoras de bolsillo, antes de que necesiten ser recargadas:

CALCULADORA		
A	B	C
4.9	5.5	6.4
6.1	5.4	6.8
4.3	6.2	5.6
4.6	5.8	6.5
5.3	5.5	6.3
5.2	6.6	4.8

Estudiar la hipótesis de que los tiempos de funcionamiento de las tres calculadoras son iguales, con nivel de significación 0.01.

Solución:

Como el enunciado no nos dice nada de la normalidad del tiempo de duración de las baterías, debemos intentar aplicar un test no paramétrico, en este caso el test de Kruskall-Wallis (más de dos muestras independientes). Para ello hemos asignado su rango a cada uno de los datos de la muestra. El contraste será:

$$\begin{cases} H_0 : M_A = M_B = M_C \\ H_1 : \text{no ocurre } H_0 \end{cases}$$

El estadístico del contraste es:

$$H = \frac{12}{n(n+1)} \sum_i \frac{R_i}{n_i} - 3n = 10,47 \in \mathrm{X}^2_{k-1}$$

siendo $R_A = 25 \qquad R_B = 56 \qquad R_C = 90$ y $n_A = 5; \; n_B = 7; \; n_C = 6$

$$n = 18; \quad k = 3 \qquad RC_{0,01} = \left[\mathrm{X}^2_{2;\,0,01}, \;\; +\infty\right) = \left[9,210, \;\; +\infty\right)$$

CONCLUSIÓN: como el estadístico cae en la región crítica podemos afirmar que con una confianza del 99% la duración de las baterías de las tres marcas de calculadoras, no son iguales.

17. Se tiene una muestra aleatoria de una población normal cuya varianza es 36. A partir de la media muestral, al 95% de confianza, se obtiene que los límites del intervalo de confianza para la media poblacional son 21.08 y 28.92.

a) Calcular la media muestral y el tamaño de la muestra.

b) Calcular el tamaño que debería tener la muestra para que el intervalo de confianza reduzca su amplitud a la mitad sin variar el nivel de confianza.

Solución:

Sea X la variable en estudio, tenemos que $X \in N(\mu\ , 6)$ y nos dan que

$$I_{\mu;0.05} = (21.08, 28.92)$$

a) En general

$$I_{\mu,\alpha} = \left[\overline{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \right]$$

y en nuestro caso tenemos que:

$$\begin{cases} \overline{X} - \dfrac{6}{\sqrt{n}} 1.96 = 21.08 \\ \overline{X} + \dfrac{6}{\sqrt{n}} 1.96 = 28.92 \end{cases} \Rightarrow 2\overline{X} = 50 \Rightarrow \overline{X} = 25$$

y de aquí

$$n = \left(\frac{6 \cdot 1.96}{25 - 21.08} \right)^2 = 13.50 \cong 14$$

b) La amplitud del intervalo de confianza dado es

$$28{,}92 - 21{,}08 = 7{,}84$$

y por tanto la amplitud del nuevo intervalo debe ser 3,92 de donde la semiamplitud del este nuevo intervalo que buscamos será a su vez la mitad de esta cantidad

$$1{,}96 = \frac{6}{\sqrt{\mathrm{n}}} 1.96 \Rightarrow n = 36$$

18. La tabla adjunta da las edades X y las presiones sanguíneas en sístole Y de 12 mujeres.

Edad	56	42	72	36	63	47	55	49	38	42	68	60
Presión	147	125	160	118	149	128	150	145	115	140	152	155

Calcular:

1) El coeficiente de correlación y la recta de regresión de Y sobre X.

2) Decir si es válido el modelo lineal.

3) Intervalo de confianza al 95% de la presión sistólica de las mujeres de 45 años.

Solución:

$$\left.\begin{array}{l}\sum x_i = 628\\ \sum x_i^2 = 394.384\end{array}\right\} \quad S_{xx} = \frac{n\sum x_i^2 - (\sum x_i)^2}{n} = 1.550,\widehat{6}$$

$$\left.\begin{array}{l}\sum y_i = 1.684\\ \sum y_i^2 = 2.835.856\end{array}\right\} \quad S_{yy} = \frac{n\sum y_i^2 - (\sum y_i)^2}{n} = 2.500,\widehat{6}$$

$$\left.\sum x_i y_i = 89.894 \right\} \quad S_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n} = 1.764,\widehat{6}$$

a)

$$\left.\begin{array}{l} b = \dfrac{S_{xy}}{S_{xx}} = 1,138 \\[2ex] a = \dfrac{\sum y_i - b\sum x_i}{n} = 80,777 \end{array}\right\} \Rightarrow \overline{y}_{/x} = 80,778 + 1,138x$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = 0,8961$$

b) Sea el contraste

$$\begin{cases} H_0 : \beta = 0 & \text{(no es válido el modelo lineal)} \\ \mathrm{H}_1 : \beta \neq 0 & \text{(es válido el modelo lineal)} \end{cases}$$

Fuente de variación	Grados libertad	Cantidad de variación	Cuadrados medios
regresión	1	SSR=bS_{xy}=2008,18	*MSR*=SSR/1=2008,18
aleatoria	10	SSE=S_{yy}-SSR=492,48	*MSE*=SSE/(n-2)=49,248
total	11	SST= S_{yy}=2.500,6	

El estadístico del contraste vine dado por:

$$F = \frac{MSR}{MSE} = 40,775$$

Si H_0 es cierta el estadístico debe comportarse como una $F_{1,\,n-2} = F_{1,\,10}$ y tenemos $F_{1,\,10;\,0,05} = 4{,}96$ y $F_{1,\,10;\,0,01} = 10{,}04$. El estadístico cae en las dos regiones críticas y por tanto:

CONCLUSIÓN: el modelo lineal es suficientemente bueno para representar la relación entre las variables edad y presión arterial.

c) El intervalo de confianza, para un valor de la variable, con nivel de significación α es:

$$I_{y_0,\alpha} = \left[\hat{y}_0 \pm t_{n-2;\,\alpha/2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

Y tenemos:

$$\begin{cases} t_{0,025;\,10} = 2,228 \\ \hat{y}_0 = a + bx_0 = 131,988 \\ \bar{x} = 52,33 \\ S = \sqrt{\dfrac{SSE}{n-2}} = 7,02 \end{cases} \Rightarrow I_{y_0,\alpha} = [115,451\ ,\ 148,525]$$

19. Supongamos que se desea contrastar la hipótesis H_0: a = 5, contra la alternativa H_1: a = 8, por medio de un solo valor observado de una variable aleatoria con función de densidad:

$$f(x,\ a) = (1/a)e^{-x/a}$$

Si el tamaño máximo del error tipo I es 0.15, ¿cuál de las siguientes pruebas es la mejor para contrastar nuestras hipótesis?

a) Rechazar H_0 si $x \geq 9$.

b) Rechazar H_0 si $x \geq 10$.

c) Rechazar H_0 si $x \geq 11$.

Solución:

La diferencia entre los tres test propuestos es la región crítica, veamos la probabilidad de error tipo I de cada uno de ellos.

$$\begin{cases} H_0 : a = 5 \\ H_1 : a = 8 \end{cases} \quad \text{siendo } f(x) = \begin{cases} \dfrac{1}{a} e^{-\frac{x}{a}} & x \geq 0 \end{cases}$$

Se realiza el experimento y X toma el valor x_0.

$$P(X < x_0) = \int_0^{x_0} (1/a) e^{-x/a} dx = \left[-e^{-x/a} \right]_0^{x_0} = 1 - e^{-x_0/a} \Rightarrow P(X > x_0) = e^{-x_0/a}$$

a) $\alpha_1 = P(\text{rechazar } H_0 / a = 5) = P(X > 9) = e^{-9/5} = 0{,}165$

b) $\alpha_2 = P(\text{rechazar } H_0 / a = 5) = P(X > 10) = e^{-10/5} = 0{,}135$

c) $\alpha_3 = P(\text{rechazar } H_0 / a = 5) = P(X > 11) = e^{-11/5} = 0{,}118$

Vemos que la prueba a) no cumple el requisito de tener

$$\alpha = P(\text{error tipo I}) < 0{,}15.$$

Estudiemos la probabilidad de error tipo II en las pruebas b) y c).

b) Supuesto que a = 8 y la región crítica es $X \geq 10$.

$$\beta = P(\text{error tipo II}) = P(X < 10 / a = 8) = 1 - e^{-\frac{10}{8}} = 0{,}286 \Rightarrow$$
$$\Rightarrow \text{potencia} = 1 - \beta = 0{,}714$$

c) Supuesto que a = 8 y la región crítica es $X \geq 11$.

$$\beta = P(\text{error tipo II}) = P(X < 11 / a = 8) = 1 - e^{-\frac{11}{8}} = 0{,}253 \Rightarrow$$
$$\Rightarrow \text{potencia} = 1 - \beta = 0{,}747$$

CONCLUSIÓN: la mejor prueba sería la 3) porque la probabilidad de cometer ambos errores es la menor.

20. El nº de horas de sueño ganadas por 7 personas después de la administración de un somnífero fue:

1, 0,5; 1,1; -0,7; 3; 1,2; 1,4

Testar si el nº de horas ganadas se desvía significativamente de cero, usando el test T de Student, el test de rangos con signo de Wilcoxon y el test de los signos. Observar que cada test es más sensible que el siguiente. Decir cuáles son las hipótesis en cada caso y sacar conclusiones.

Solución:

i) Si podemos suponer que

D = horas de sueño "después"- horas de sueño "antes"

es una variable aleatoria con distribución normal $N(\mu_D\ ,\ \sigma_D)$, entonces el contraste adecuado sería

$$\begin{cases} H_0 : \mu_D \leq 0 \\ H_1 : \mu_D > 0 \end{cases}$$

El estadístico del contraste es $\bar{X}_D$, que si H_0 es cierta debe comportarse como una normal

$$N(\mu_D\ ,\ \frac{\sigma_D}{\sqrt{7}})$$

Como σ_D es desconocida tenemos que estimar su valor por S_D y tenemos que:

$$S_D^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} = 1,219 \Rightarrow S_D = 1,104$$

$$\frac{\bar{X}_D - \mu_D}{\frac{S_D}{\sqrt{7}}} = \frac{1,0714}{\frac{1,104}{\sqrt{7}}} = 2,5677$$

debe ser un valor de una t_6.

La región crítica será:

$$RC_{0.05} = \left[t_{6;\ 0.05},\ \ +\infty\right) = \left[1,943,\ \ +\infty\right)$$

CONCLUSIÓN: como el estadístico cae en región crítica para nivel significación 0,05 nos quedamos con la hipótesis alternativa y diríamos que hay evidencia de que es mayor el número de horas de sueño después del tratamiento.

ii) Si no podemos suponer que la variable D sea normal, pero sí que es simétrica en torno a su mediana, podríamos aplicar un test de rangos con signo de Wilcoxon.

$$\begin{cases} H_0 : M_D \leq 0 \ \text{ y } D \text{ es simétrica} \\ H_1 : M_D > 0 \ \text{ o } D \text{ no es simétrica} \end{cases}$$

Asignamos a cada diferencia su rango, puesto que comparamos con la mediana que según nuestra hipótesis nula es cero, y un signo según sea mayor o no que dicha mediana.

d_i	1	0,5	1,1	-0,7	3	1,2	1,4
Rango	+3	+1	+4	-2	+7	+5	+6

$$W^* = 22 \quad W^- = 2 \qquad W = \min\{W^+, W^-\} = 2$$

La región crítica es:

$$RC_{0,05} = \left[0 \ , \ W_{7;\,0,05}\right] = \left[0 \ , \ 4\right]$$

CONCLUSIÓN: el estadístico cae en región crítica, luego aceptamos la hipótesis alternativa: la mediana es mayor que cero o D no es simétrica respecto a la mediana, pero como hemos supuesto simetría podemos decir que hay evidencia, a nivel de significación 0,05, de que el número mediano de horas de sueño aumenta después del tratamiento.

iii) Si no hacemos ningún tipo de suposición acerca del comportamiento de la variable D, la única herramienta disponible será el test de los signos. El contraste será:

$$\begin{cases} H_0 : M_D \leq 0 \\ H_1 : M_D > 0 \end{cases}$$

Comparamos cada diferencia con el hipotético valor de la mediana y contamos el nº de diferencias positivas y negativas, se obtiene:

$$N^+ = 6 \qquad N^- = 1 \qquad N = \min\left\{N^+ \ , \ N^-\right\} = 1$$

Si la mediana es cero:

$$N \in B(7 \ , \ 0,5) \ \text{ y } \ P(N \leq 1) = \binom{7}{1}(0,5)^7 = 0,054 > 0,05$$

CONCLUSIÓN: lo ocurrido no es menos probable que 0,05 luego nos quedamos con la hipótesis nula. No hay evidencia de que el tratamiento aumente la mediana de las horas de sueño.

COMENTARIO: sabemos que cuanto menos sepamos del comportamiento de una variable aleatoria nos vemos obligados a usar test cada vez menos finos y por tanto a ser más amplios en nuestra aceptación de la idea más conservadora. Deberá ocurrir que la muestra sea realmente muy extrema, para que nos convenza de que se cumple algo distinto de lo que hayamos fijado en la hipótesis nula.

21. En 22 parcelas de la misma extensión sembradas con remolacha se han medido las variables X = "nº de plantas, en millares" e Y = "producción de azúcar", en Tm/ha". Los datos obtenidos fueron:

$$\overline{X} = 31,2 \qquad \overline{Y} = 59,3 \qquad S_X = 7,4 \qquad S_Y = 12,6 \qquad r = 0,918$$

a) Respecto a la plantación, ¿se puede aceptar que aumentando 1000 plantas por parcela, el aumento esperado de azúcar será 2 Tm/ha? (α = 0.01).

b) Si en una parcela se han plantado 28 millares de plantas:

i) ¿Entre qué valores se encuentra la producción esperada de azúcar, con probabilidad 0.95?

ii) Hallar el intervalo de confianza para la predicción de la producción de azúcar.

c) ¿Qué diferencia hay entre los resultados a) y b)?

Solución:

$$\left.\begin{array}{l} S_x = \sqrt{\dfrac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}} \\ S_{xx} = \dfrac{n\sum x_i^2 - (\sum x_i)^2}{n} \end{array}\right\} \Rightarrow S_{xx} = (n-1)S_x^2;$$

$$\left.\begin{array}{l} S_y = \sqrt{\dfrac{n\sum y_i^2 - \left(\sum y_i\right)^2}{n(n-1)}} \\ S_{yy} = \dfrac{n\sum y_i^2 - (\sum y_i)^2}{n} \end{array}\right\} \Rightarrow S_{yy} = (n-1)S_y^2$$

$$\left.\begin{array}{l} r = b\sqrt{\dfrac{S_{xx}}{S_{yy}}} \Rightarrow b = r\sqrt{\dfrac{S_{yy}}{S_{xx}}} = r\dfrac{S_y}{S_x} = 0,918 \cdot \dfrac{12,6}{7,4} = 1,563 \\ a = \overline{y} - b\overline{x} = 59,3 - 1,563 \cdot 31,2 = 10,5344 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \overline{y}_{/x} = 10,5344 + 1,563x$$

a) Aumentar un millar de plantas por parcela equivale a que la variable pasa de valer x a valer x+1.

$$\overline{y}_{/x+1} = 10,5344 + 1,563(x+1) = 10,5344 + 1,563x + \underbrace{1,563}_{b}$$

Como el aumento esperado de la producción de azúcar es precisamente b para resolver esta apartado debemos contrastar si β es igual a 2

$$\begin{cases} H_0 : \beta = 2 \\ H_1 : \beta \neq 2 \end{cases} \quad \text{el estadístico del contraste es } \frac{\beta\text{-b}}{S/\sqrt{S_{xx}}} \in t_{n-2}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \Rightarrow S_{xy} = r\sqrt{S_{xx} \cdot S_{yy}} = r(n-1)S_x S_y =$$

$$= 0,918 \cdot 21 \cdot 7,4 \cdot 12,6 = 1.797,481$$

$$S=\sqrt{\frac{S_{yy}-bS_{xy}}{n-2}}=5,121 \Rightarrow \frac{\beta\text{-b}}{S/\sqrt{S_{xx}}}=\frac{2-1,563}{5,121/7,4\cdot\sqrt{21}}=2,894$$

$$RC_{0,01}=\left(-\infty\,,\,-t_{n-2;\,0,0005}\right]\cup\left[t_{n-2;\,0,0005}\,,\,+\infty\right)=$$
$$=\left(-\infty,-2,845\right]\cup\left[2,845\,,\,+\infty\right)$$

Como el estadístico cae en región crítica, rechazamos la hipótesis nula. Hay evidencia estadística de que $\beta\neq2$ y por tanto no es admisible que aumentando un millar de plantas por parcela el incremento de la producción de azúcar sea de 2 Tm por parcela.

b) x_0 = 28 millares de plantas

i) $$\left.\begin{array}{l} \overline{y}_{x_0}=a+bx_0=10,5344+1,563\cdot28=54,2984 \\ S=2,894 \\ t_{n-2;\,\alpha/2}=t_{20;\,0,025}=2,086 \\ S_{xx}=(n-1)\cdot S_x^2 \end{array}\right\}$$

$$I_{\mu_{/x_0}}=\left[\overline{y}_{/x_0}\pm t_{n-2;\,\alpha/2}S\sqrt{\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}}\right]=$$
$$=\left[54,2984\pm1,4075\right]=\left[52,8909\ ,\ 55,7059\right]$$

ii) $$I_{y_{/x_0}}=\left[\overline{y}_{/x_0}\pm t_{n-2;\,\alpha/2}S\sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}}\right]=$$
$$=\left[54,2984\pm6,1987\right]=\left[48,0996\ ,\ 60,4972\right]$$

c) La diferencia entre los dos resultados de los subapartados a) y b) está en que a) representa la predicción para la media de producción de azúcar de muchas parcelas en que se plantasen 28 millares de plantas y b) representa el intervalo donde estaría la producción de un campo en que se plantaran 28 millares de plantas; a) nos aproxima la esperanza matemática de la variable $y_{/28}$ mientras b) nos aproxima al campo de valores de dicha variable.

22. Una v.a. discreta toma los valores 0, 1 y 2 con función de probabilidad:

$P(X=0) = p^2 \quad P(X=1) = 2\,p\,(1-p) \quad P(X=2) = (1-p)^2$ con $0 < p < 1$.

Estimar p, por el método de máxima verosimilitud, a partir de una muestra de tamaño 100 en la que se ha presentado 22 veces el 0, 53 veces el 1 y 25 veces el 2.

Solución:

Sea X_0 = nº de veces que aparece el 0 en las 100 observaciones.
X_1 = nº de veces que aparece el 1 en las 100 observaciones.
X_2 = nº de veces que aparece el 2 en las 100 observaciones.

Sea $X = (X_0, X_1, X_2)$ como las observaciones son independientes, esta variable sigue un comportamiento multinomial $M(p^2, 2p(1-p), (1-p)^2)$.

La función de probabilidad conjunta es, según sabemos:

$$P(X_0 = 22,\ X_1 = 53,\ X_2 = 25) = \frac{100!}{22!\cdot 53!\cdot 25!}\left(p^2\right)^{22}\left(2p(1-p)\right)^{53}\left(\left(1-p\right)^2\right)^{25} =$$

$$= \frac{100!\cdot 2^{53}}{22!\cdot 53!\cdot 25!}\,p^{97}\left(1-p\right)^{103}.$$

Para hallar el estimador máximo verosímil de p debemos maximizar esta expresión y tenemos:

$$\frac{\partial P}{\partial p} = \frac{100!\cdot 2^{53}}{22!\cdot 53!\cdot 25!}\left[97p^{96}\left(1-p\right)^{103} - 103\quad p^{97}\left(1-p\right)^{102}\right] = 0 \Leftrightarrow$$

$$\Leftrightarrow 97\quad\left(1-p\right) - 103p = 0 \Leftrightarrow p = \frac{97}{200}$$

Luego debemos tomar este valor como estimación de p.

23. Por fistulización se obtuvo el pH de seis muestras de bilis hepática, con los siguientes resultados:

7.83 , 8.52, 7.32, 7.79, 7.57, 6.98.

Se desea saber, al nivel de significación 0.05, si la bilis hepática puede considerarse neutra (se supone normalidad).

Si se conociera que la desviación típica σ =0.5, ¿qué decisión tomaríamos? (El pH neutro es 7).

Solución:

Sea X = pH de la bilis hepática, $X \in N(\mu\ , \sigma)$

$$\begin{cases} H_0 : \mu = 7 \\ H_1 : \mu > 7 \end{cases}$$

Como σ desconocida, tenemos $\begin{cases} \overline{X} = 7.668 \\ S = 0.5236 \end{cases}$

Si H_0 es cierta $\overline{X} \in N\left(7,\ \sigma/\sqrt{6}\right)$ y por tanto $\dfrac{\overline{X}-7}{S/\sqrt{6}} = 3.1262 \in t_5$

$$RC_{0.05} = \left[t_{5,\ 0.05},\ +\infty\right) = \left[2.015,\ +\infty\right)$$

CONCLUSIÓN: como el estadístico cae en la región crítica, tenemos evidencia estadística suficiente para afirmar que el pH de la bilis hepática es mayor que 7, o sea que no es neutra, con una confianza del 95%.

Si consideramos que σ = 0,5 tenemos:

$$\overline{X} \in N\left(7, \frac{0.5}{\sqrt{6}}\right) \Rightarrow \frac{\overline{X}-7}{\frac{0.5}{\sqrt{6}}} = 3.2791 \in Z$$

Y ahora la región crítica sería $RC_{0.05} = [Z_{0.05}, +\infty) = [1.64, +\infty)$.

CONCLUSIÓN: estaríamos en el mismo caso anterior, la conclusión es que la bilis hepática tiene pH superior a 7 y por tanto que no es neutra.

24. Se ensayaron dos tratamientos antirreumáticos administrados al azar sobre 10 y 22 pacientes; los resultados, medidos en una escala convencional (a mayor puntuación más eficacia) fueron:

tratamiento 1	12 15 21 17 38 42 10 23 35 28
tratamiento 2	21 18 42 25 14 52 65 40 43 35 18 56 29 32 44 15 15 68 41 37 43 58

Decidir si existe diferencia entre los tratamientos.

Solución:

Sean X = puntuación con tratamiento A.

Y = puntuación con tratamiento B.

Vamos a comparar si las medianas de las dos variables son iguales con el test de la suma de rangos, ya que se trata de v. a. independientes y medidas convencionales.

$$\begin{cases} H_0 : M_X = M_Y \\ H_1 : M_X \neq M_Y \end{cases}$$

Para aplicar este test asignamos rango a cada uno de los valores muestrales y sumamos los rangos correspondientes a cada muestra.

A	$12_{(2)}$ $15_{(5)}$ $21_{(10.5)}$ $17_{(7)}$ $38_{(20)}$ $42_{(23.5)}$ $10_{(1)}$ $23_{(12)}$ $35_{(17.5)}$ $28_{(14)}$
B	$21_{(10.5)}$ $18_{(8)}$ $42_{(23.5)}$ $25_{(13)}$ $14_{(3)}$ $52_{(28)}$ $65_{(31)}$ $40_{(21)}$ $43_{(25)}$ $35_{(17.5)}$ $18_{(9)}$ $56_{(29)}$ $29_{(15)}$ $32_{(16)}$ $44_{(27)}$ $15_{(5)}$ $15_{(5)}$ $68_{(32)}$ $41_{(22)}$ $37_{(19)}$ $43_{(26)}$ $58_{(30)}$.

$$W_1 = \sum_{x_i} Rx_i = 112.5 \qquad W_2 = \sum_{y_i} Ry_i = 415.5 \qquad W_1 + W_2 = 528$$

$$W = 112.5 \qquad U = W_1 - \frac{(1+n_1)n_1}{2} = 112.5 - 55 = 57.5$$

$$RC_{0.05} = \left[0, W_{10,22;\ 0.025}\right] \approx [0,60]$$

CONCLUSIÓN: el estadístico del contraste cae en la región crítica, por tanto aceptamos H_1 y podemos decir con confianza del 95%, que los dos tratamientos no son iguales.

25. Al medir el tiempo de reacción ante cierto estímulo, un psicólogo estima que la desviación típica del mismo es de 0.05 s. ¿Cual será el nº de medidas que deberá realizar para tener a) el 99%, b) el 95% de confianza de que el error en su estimación del tiempo medio de reacción no excederá de 0.01 s?

Solución:

Sea X = tiempo de reacción en segundos, suponemos que $X \in N(\mu\ , 0{,}05)$.

El intervalo de confianza para la verdadera medida, con nivel de significación α, viene dado por

$$I_{\mu,\alpha} = \left[\overline{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right]$$

a) Queremos que la semiamplitud del intervalo de confianza sea menor que 0,01, con confianza del 99% (α=0.01)

$$\frac{\sigma}{\sqrt{n}} Z_{0.005} \le 0.01 \Rightarrow \frac{0.5}{\sqrt{\text{n}}} 2.575 \le 0.01 \Rightarrow n \ge \left(\frac{0.05 \cdot 2.575}{0.01}\right)^2 = 16.576{,}56$$

b) Es lo mismo de antes, ahora con confianza del 95% (α=0.05)

$$\frac{\sigma}{\sqrt{n}} Z_{0.025} \le 0.01 \Rightarrow \frac{0.5}{\sqrt{\text{n}}} 1.96 \le 0.01 \Rightarrow n \ge \left(\frac{0.05 \cdot 1.96}{0.01}\right)^2 = 9.604$$

26. En un experimento sobre percepción extrasensorial (P.E.) un sujeto fue preguntado sobre el color (rojo o azul) de una carta elegida por una persona que se hallaba en otra habitación de entre 50 cartas bien barajadas. Ambos individuos desconocen el nº de cartas de cada color que hay en el mazo de cartas. Si el sujeto identifica correctamente 32 cartas, determinar si este resultado es significativo.

Solución:

Si el sujeto no tiene ningún poder extrasensorial la probabilidad de acertar el color de una determinada carta será de 0.5; además las realizaciones pueden considerarse independientes puesto qué en ningún momento se sabe ni la composición de la baraja, ni que cartas han salido, en realidad es como si se tirara una moneda equilibrada 52 veces y acertara 32 de ellas. Por tanto podemos definir la variable como:

X = nº de veces que acierta de las 52, $X \in B(52, p) \cong N(52p, \sqrt{52p(1-p)})$

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p > 0.5 \end{cases} \qquad \hat{p} \in N(0.5, \frac{1}{2\sqrt{52}}) \Rightarrow \frac{\hat{p}-0.5}{\frac{1}{2\sqrt{52}}} = 1.66$$

$$RC_{0.05} = \left[Z_{0.05},\ +\infty\right) = \left[1,645,\ +\infty\right)$$

$$RC_{0.01} = \left[Z_{0.01},\ +\infty\right) = \left[2,326,\ +\infty\right)$$

CONCLUSIÓN: con una confianza del 95% aceptamos que el sujeto tiene un cierto poder extrasensorial; si quisiéramos un nivel de confianza del 99% tendríamos que concluir que no hay suficiente evidencia estadística de que el sujeto tenga poder de adivinación superior a lo normal.

27. Los resultados de una encuesta, hecha para determinar si la edad de los conductores a partir de 21 años tiene influencia en el nº de accidentes de tráfico, se indican en la tabla adjunta. Estudiar la hipótesis de que el nº de accidentes es independiente de la edad del conductor.

	Edad del conductor					
nº accid.	21 - 30	31 - 40	51 - 50	51 - 60	61 - 70	totales
0	748	821	786	720	672	3747
1	74	60	51	66	50	301
2	31	25	22	16	15	109
3 ó más	9	10	6	5	7	37
totales	862	916	865	807	744	4194

Solución:

El contraste que vamos a realizar es el test de independencia de la X^2 cuyas hipótesis son:

$$\begin{cases} H_0 : p_{ij} = p_{i.} \cdot p_{.j} \quad \forall i = 1,...,4, \forall j = 1,...,5, \\ \qquad \text{(nº de accidentes y edad del conductor son indep.)} \\ \text{H}_1 : \text{no ocurre H}_0 \end{cases}$$

El estadístico del contraste es

$$\sum_i \sum_j \frac{O_{ij}^2}{E_{ij}} - N$$

que si la hipótesis nula es cierta debe comportarse como una X^2 con (r-1)(c-1)=3x4=12 grados de libertad.

En la tabla de los datos añadimos las filas y columnas de los totales parciales y construimos la tabla de las frecuencias esperadas:

$$E_{ij} = \frac{\text{total de filas} \times \text{total de columnas}}{\text{gran total}}$$

	21-30	31-40	41-50	51-60	61-70
0	770.127	818.372	775.488	720.990	660.704
1	61.865	65.740	62.080	57.918	53.396
2	22.403	23.806	22.481	20.973	19.336
3 o más	7.605	8.081	7.631	7.119	6.564

El estadístico toma el valor 15,769 y tenemos:

$$RC_{0.05} = \left[X_{12,0.05}^2, +\infty\right) = \left[21.03, +\infty\right)$$

el valor del estadístico no cae en la región crítica.

CONCLUSIÓN: con la información que tenemos no podemos afirmar que la edad del conductor y el nº de accidentes que tiene sean dos caracteres relacionados.

28. El test de personalidad de Eysenk tiene dos formas: A y B. Se supone que ambas miden igualmente la extroversión. Probar esta afirmación a partir de los siguientes datos datos:

	Sujeto											
	1	2	3	4	5	6	7	8	9	10	11	12
forma A	12	18	21	10	15	27	31	6	15	13	8	10
forma B	10	17	20	5	21	24	29	7	11	13	8	11

Solución:

Se trata de dos variables (la puntuación de un sujeto con el test A o el B) de carácter cualitativo, por ello utilizamos un test no paramétrico. Como además son datos emparejados, trabajaremos como si fuese una sola variable (la diferencia) y, como no sabemos nada sobre la simetría, descartamos el test de rangos con signo, para quedarnos con el test de los signos.

$$D = \text{puntuación test A - puntuación test B}$$

$$\begin{cases} H_0 : M_D = 0 \\ H_1 : M_D \neq 0 \end{cases}$$

SUJETO	1	2	3	4	5	6	7	8	9	10	11	12
D_i	2	1	1	5	-6	3	2	-1	4	0	0	-1

$N^+ = 7$, $N^- = 3$. Si H_0 es cierta tanto N^+ como N^- deberían comportarse como binomiales de parámetros 10 (despreciamos las diferencias nulas) y p=0,5 y tenemos

$$P(N^- \leq 3) = 0,1719$$

CONCLUSIÓN: no hay evidencia estadística de que la mediana de las diferencias sea distinta de 0. Luego los dos test miden igualmente la extroversión.

29. Un contratista encarga un pedido de vigas cuya longitud promedio es de 5 m. Se sabe que la longitud de una viga se distribuye normalmente con desviación estandar de 0.02 m. Después de recibir el pedido se seleccionan 16 vigas al azar y se miden: si la media muestral es menor que la esperada, se devuelve el pedido.

a) Si la probabilidad de rechazar el pedido es de 0.04, ¿cuál debe ser el valor de la media muestral para que el pedido sea rechazado?

b) Si la longitud promedio real es de 4.98 m, ¿cuál es la potencia de la prueba del apartado a)?

Solución:

Sea X = longitud de una viga, por hipótesis $X \in N(\mu\ , 0{,}02)$

$$\begin{cases} H_0 : \mu \geq 5 \\ H_1 : \mu < 5 \end{cases} \qquad \text{Si } H_0 \text{ es cierta, } \overline{X} \in N(5, \frac{0.02}{4}) \Rightarrow \frac{\overline{X}-5}{\frac{0.02}{4}} = Z$$

a) $$0.04 = P(\text{rechazar pedido}) = \text{P}(\overline{X} < A) = P\left(\frac{\overline{X}-5}{0.02/4} < \frac{A-5}{0.02/4} \right) =$$

$$\Rightarrow \frac{A-5}{\frac{0.02}{4}} = -1.75 \qquad \Rightarrow A = 4.99$$

Luego si la media de las longitudes de las 16 vigas que se toman como muestra es menor que 4,99 m, el pedido será rechazado.

b) Si la longitud promedio real es de 4,98 m, entonces:

$$\overline{X} \in N\big(4.98,\ 0.02\big) \text{ y será}$$

$$\beta = P(\text{error tipo II}) = P(\text{Aceptar } H_0/\mu = 4.98) =$$

$$= P(\overline{X} > 4.99/\mu = 4.98) = P\left(\frac{\overline{X}-4.98}{0.02/4} > \frac{4.99-4.98}{0.02/4} \right) =$$

$$= P\big(Z > 2)\big) = 0.0228 \ \Rightarrow \quad \textit{potencia} = 1 - \beta = 0.9772$$

30. Se pide a un laboratorio que compare la durabilidad de pelotas de golf de cuatro marcas. Para ello se seleccionan de forma aleatoria 8 pelotas de cada marca y una máquina las golpea repetidamente, con fuerza constante. Se mide el nº de golpes que cada pelota resiste sin deteriorarse.

MARCA			
A	B	C	D
205	242	237	212
229	253	259	244
238	226	265	229
214	219	229	272
242	251	218	255
225	212	262	233
209	224	242	224
204	247	234	245

a) ¿Existe alguna razón para creer que la durabilidad promedio es diferente para alguna de las cuatro marcas? ($\alpha = 0.05$)

b) ¿Existe alguna razón para dudar de la suposición de que las varianzas de los errores son iguales?

Solución:

Se trata de cuatro v. a. independientes que en principio no tienen por qué ser normales: elegimos el test de Kruskall-Wallis. Para aplicarlo tenemos que asignar a cada uno de los 32 datos su rango.

MARCA

A	B	C	D
205(2)	242(21)	237(18)	212(4,5)
229(14)	253(27)	259(29)	244(23)
238(19)	226(12)	265(31)	229(14)
214(6)	219(8)	229(14)	272(32)
242(21)	251(26)	218(7)	255(28)
225(11)	212(4,5)	262(30)	233(16)
209(3)	224(9,5)	242(21)	224(9,5)
204(1)	247(25)	234(17)	245(24)
R_A=77	R_B=133	R_C=167	R_D=151

a) Sea X_i = nº de golpes que resiste hasta romperse una pelota de la marca i, i=A, B, C, D.

$$\begin{cases} H_0 : M_A = M_B = M_C = M_D \\ H_1 : \text{ las medianas no son todas iguales} \end{cases}$$

El estadístico del contraste es:

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1) = 6,5511$$

Si H_0 es cierta este estadístico debe comportarse como una ji-cuadrado con k-1 = 3 grados de libertad y por tanto:

$$RC_{0.05} = \left[X^2_{3;\,0.05},\ +\infty \right) = \left[7,815,\ +\infty \right)$$

CONCLUSIÓN: el estadístico no cae en la región crítica luego no hay evidencia estadística de que la durabilidad de las pelotas no sea la misma para las cuatro marcas.

b) Si consideramos las variables como normales podemos aplicar un test de comparación de las varianzas, como además los tamaños muestrales son iguales usamos el test de Cochran.

$$\begin{cases} H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2 \\ H_1 : \text{no son todas las varianzas iguales} \end{cases}$$

El estadístico del contraste es:

$$G = \frac{\text{máx}\left\{S_i^2\right\}}{\sum S_i^2} = 0,01686$$

$$RC_{0.05} = \left[G_{n,k;0.05}, +\infty\right) = \left[0,3720, +\infty\right)$$

$$RC_{0,01} = \left[G_{n,k;0.01}, +\infty\right) = \left[0,4757, +\infty\right) \text{ con } n=32 \text{ y } k=4$$

CONCLUSIÓN: el estadístico no cae en las regiones críticas, por tanto no podemos decir que las varianzas no sean iguales.

Observación: con la suposición del inicio de este apartado de que las variables son normales cabe aplicar un test anova I, ya que son independientes y podemos considerar que se cumple la hipótesis de homocedasticidad.

31. En el censo de población de una ciudad se han obtenido los siguientes datos referidos al nº de miembros por unidad familiar:

nº de miembros	1 a 2	3 a 4	5 a 6	7 ó más
nº de familias	9200	12400	6300	4100

Escogida una muestra de 600 familias se obtuvo la siguiente distribución:

nº de miembros	1 a 2	3 a 4	5 a 6	7 ó más
nº de familias	190	260	100	50

¿Podemos considerar que esta es una muestra aleatoria a un nivel de significación del 10%?

Solución:

Se trata de un problema de bondad de ajuste: si la muestra es aleatoria debe mantener entre los distintos tipos de familias las mismas proporciones aproximadamente, que la población total.

Las probabilidades de los distintos valores de la variable son:

$$P(1 \text{ a } 2 \text{ hijos}) = \frac{9200}{32000} \qquad P(5 \text{ a } 6 \text{ hijos}) = \frac{6300}{32000}$$

$$P(3 \text{ a } 4 \text{ hijos}) = \frac{12400}{32000} \qquad P(7 \text{ o más hijos}) = \frac{4100}{32000}$$

Las frecuencias esperadas en la muestra serán $n.p_i$

nº miembros	1 a 2	3 a 4	5 a 6	7 ó más
frecuencias esperadas	172.5	232.5	118.125	76.875

El estadístico del contraste es:

$$\sum_i \frac{O_i^2}{E_i} - N = 617,20 - 600 = 17,20$$

que si la hipótesis nula es cierta debería comportarse como una ji-cuadrado con k-1 =3 grados de libertad, con lo que la región crítica será:

Para significación del 10% tenemos $RC_{0.1} = \left[X^2_{3;0.1}, +\infty\right) = [6.251, +\infty)$

Para significación del 1% tenemos $RC_{0.01} = \left[X^2_{3;0.01}, +\infty\right) = [11.345, +\infty)$.

Vemos que el estadístico cae en región crítica aun cuando tomamos nivel de significación 0.01.

CONCLUSIÓN: hay evidencia estadística con confianza del 99% de que la muestra no es aleatoria.

32. Un investigador desea evaluar el porcentaje de habitantes de una población que están vacunados contra la polio. Para ello planifica un muestreo de la población, deseando obtener resultados correctos dentro del ±3% con una confianza del 99%. De su experiencia previa, el investigador cree que el porcentaje de la población que está vacunada contra la polio está entre el 70% y el 80%. Si puede obtener una muestra aleatoria de individuos, ¿cuántos debe incluir en la muestra?

Solución:

La expresión general del intervalo de confianza para una proporción es:

$$\mathrm{I}_{p,\alpha} = \left[\hat{p} \pm \sqrt{\frac{\hat{p}\cdot\left(1-\hat{p}\right)}{n}} Z_{\alpha/2}\right]$$

En nuestro caso queremos que la semiamplitud del intervalo sea de 0,03 y que el nivel de significación sea $\alpha = 0{,}01$. Como además tenemos el dato de que la verdadera proporción se encuentra entre 0,7 y 0,8 entonces la verdadera desviación típica de $\hat{p}$ que viene dada por la expresión:

$$\sqrt{\frac{p\cdot\left(1-p\right)}{n}} \quad \text{estará entre} \quad \sqrt{\frac{0.8\cdot 0.2}{n}} = \frac{0{,}4}{\sqrt{n}} \leq \sigma_p \leq \sqrt{\frac{0.7\cdot 0.3}{n}} = \frac{0{,}46}{\sqrt{n}}$$

En cualquier caso será:

$$\sigma_p \leq \frac{0.46}{\sqrt{n}}$$

de donde:

$$\frac{0.46}{\sqrt{n}}\cdot 2.575 \leq 0.03 \;\Rightarrow\; \left(\frac{0.46\cdot 2.575}{0.03}\right)^2 = 1558.9 \leq n$$

33. Para probar las diferencias respecto a la conducción eléctrica de tres tipos de alambres A, B y C, se obtuvo muestras y se midió la resistencia en ohmios, con los resultados que se acompañan. Estudiar si existen diferencias significativas al nivel 0.05. Se considera que las varianzas de las resistencias son iguales para los tres tipos de alambres.

A	0.135	0.138	0.144	0.142	0.139
B	0.138	0.143	0.140	0.137	0.141
C	0.141	0.143	0.145	0.142	0.138

Solución:

Sea X_i = resistencia en ohmios del alambre i-ésimo, i=A, B, C. Consideramos que las variables son normales, claramente independientes, y de igual varianza, con lo que podemos aplicar un test anova I.

Fuente de variación	Estadísticos	Grados Libertad	Cuadrados medios
Tratamientos	$SST = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} =$ $= 0,00004455$	k-1 =2	$MST = \frac{SST}{k-1} =$ $= 0,000022275$
Aleatoria	$SSE = S_{yy} - SST =$ $= 0,00013145$	N-k =12	$MSE = \frac{SSE}{N-k} =$ $= 1,095 \cdot 10^{-5}$
Total	$S_{yy} = \sum_{i,j} y_{ij}^2 - \frac{T_{..}^2}{N} =$ $0,000176 =$	N-1 =14	

El estadístico del contraste es:

$$F = \frac{MST}{MSE} = 2,03347$$

que si la hipótesis nula es cierta debe comportarse como una F de Snedecor de 2 y 12 grados de libertad. La región crítica será:

$$RC_{0.05} = \left[F_{2,\,12;\,0.05},\ +\infty\right) = \left[3.89,\ +\infty\right)$$

CONCLUSIÓN: como el estadístico no cae en la región crítica, al menos a nuestro nivel de significación, debemos quedarnos con la hipótesis nula y concluir que con la información de que disponemos debemos afirmar que las resistencias medias de los tres tipos de cables son iguales.

34. La tabla adjunta muestra las respectivas alturas X e Y de una muestra de 12 padres y sus hijos varones mayores.

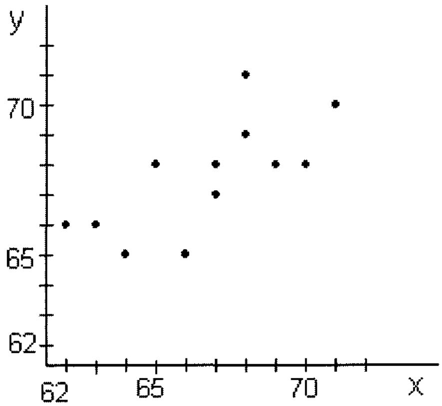
X = estatura del padre, en pulgadas
Y = estatura del hijo, en pulgadas

X	65	63	67	64	68	62	70	66	68	67	69	71
Y	68	66	68	65	69	66	68	65	71	67	68	70

a) Construir un diagrama de dispersión.
b) Hallar la recta de regresión de Y sobre X.
c) Calcular la variación total y la variación explicada por el modelo lineal.
d) Calcular el coeficiente de corrrelación y el coeficiente de determinación, y explicar si podemos usar la recta de regresión para predicción de las alturas de los hijos a partir de la altura de los padres.

Solución:

a)



b) $\sum_i x_i = 800;\quad \sum_i x_i^2 = 53418;\quad S_{xx} = \dfrac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n} = 84,66$

$\sum_i y_i = 811;\quad \sum_i y_i^2 = 54849;\quad S_{yy} = \dfrac{n\sum y_i^2 - \left(\sum x_i\right)^2}{n} = 38,92$

$$\sum_i x_i y_i = 54.107 \qquad S_{xy} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n} = 40,33$$

Los coeficientes de la recta de regresión vienen dados por:

$$\begin{cases} b = \dfrac{S_{xy}}{S_{xx}} = 0.4763 \\ a = \dfrac{\sum y_i - b\sum x_i}{n} = 35.82 \end{cases} \Rightarrow \text{recta}: y = 35.82 + 0.4763x$$

c) Entendemos que la variación total viene dada por S_{yy}=38,92 y la variación debida a la regresión por $SSR = bS_{xy}$ =19,21, el resto se considera que es variación debida a la aleatoriedad.

d) $r = b\sqrt{\dfrac{S_{xx}}{S_{yy}}} = 0,7025; \qquad r^2 = 0,4935$

Se entiende que sólo algo menos del 50% de la variabilidad de la altura de los hijos se puede explicar por linealidad con la altura de los padres. Este porcentaje no es suficiente y no debemos por tanto predecir la altura de los hijos a partir de la de los padres.

35. Dos personas A y B juegan a cara y cruz con una moneda. A elige cara y B cruz. Al cabo de 100 partidas A ha ganado 62 veces. Tras este resultado, B afirma que la moneda está trucada, y que la probabilidad de salir cara es mayor que 2/3. A mantiene que la moneda es correcta. ¿Quien tiene razón? Elegir un nivel de significación del 5% y calcular la potencia del test.

Solución:

Sea X = nº de caras en las 100 tiradas, $X \in N(100\ , p)$.

Para ver si A tiene razón planteamos el contraste unilateral derecho, ya que una vez realizado el experimento no tiene sentido el contraste bilateral.

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p > 0.5 \end{cases}$$

Si H_0 es cierta $X \in N(100\ ,\ 0{,}5) \cong N\ (50\ ,\ 5)$ y el estadístico vale:

$$\frac{X-50}{5} = \frac{62-50}{5} = 2.4$$

Pero $P(Z>2{,}4)=0{,}0082<0{,}01$. Luego hay evidencia estadística de que la moneda no está bien equilibrada y la probabilidad de cara es mayor que 1/2: A no tiene razón.

Para estudiar la afirmación de B tomamos el contraste:

$$\begin{cases} H_0 : p \leq 2/3 \\ H_1 : p > 2/3 \end{cases}$$

Si H_0 es cierta $X \in N(100\ ,\ 2/3) \cong N\ (66.\widehat{6}\ ,\ 4.71)$ y por tanto

$$\frac{X-66.\widehat{6}}{4.71} = \frac{62-66.\widehat{6}}{4.71} = -0.99$$

$$P(Z > -0.99) = 1 - P(Z \leq -0{,}99) = 1 - 0.1611 = 0{,}84 > 0.05$$

luego no hay evidencia estadística de lo que afirma B y concluimos que la moneda no está equilibrada, pero tampoco tiene razón B.

Además la potencia del primer contraste será:

$$\text{potencia} = 1\text{-}\beta = 1\text{-}P\left(\frac{X\text{ - }50}{5} < 1.645\right) = 1 - P(X < 50 + 1.645 \cdot 5) =$$

$$= 1 - P\left(\frac{X-66.\widehat{6}}{4.71} < \frac{50+1.645\cdot 5 - 66.\widehat{6}}{4.71}\right) =$$

$$= 1 - P\left(\frac{X-66.\widehat{6}}{4.71} < -1.79\right) = 1 - 0.0367 = 0.9633$$

Por tanto diremos que ni A ni B tienen razón y que la verdadera probabilidad de cara será algún nº entre 1/2 y 2/3.

36. La tabla muestra el número de soldados muertos en el ejército prusiano por coces de caballos, en diferentes unidades militares de caballería y es debida a Bortkiewicz. ¿Puede decirse que la variable sigue un modelo de Poisson?

Nº de muertos	0	1	2	3	4
Unidades con dichos muertos	109	65	22	3	1

Solución:

Sea X=nº de muertos en una unidad del ejército prusiano.

Si esta variable sigue una distribución de Poisson la estimación de λ será:

$$\hat{\lambda} = \frac{0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1}{200} = \frac{122}{200} = 0.61$$

$$\begin{cases} H_0 : X \in P(0.61) \\ H_1 : X \text{ no sigue ese modelo} \end{cases}$$

Calculamos las frecuencias esperadas mediante la expresión

$$E_i = 200 \cdot p_i \quad \text{siendo } p_i = P(X = i) = \frac{0.61^i}{i!} \cdot e^{-0.61}$$

y obtenemos:

Nº de muertos	0	1	2	3	4 ó más
frec. esperadas	108.670	66.288	20.218	4.111	0.712

El estadístico del contraste es

$$\sum \frac{O_i^2}{E_i} - n = 0.599$$

que si la hipótesis nula es cierta debería comportarse como una X^2 con 3 grados de libertad, con lo que la región crítica, para un nivel de confianza del 99%, será:

$$RC_{0.01} = \left[X^2_{3;\, 0.01},\ +\infty \right) = \left[11.32,\ +\infty \right)$$

CONCLUSIÓN: la variable número de muertos por unidad en el ejército prusiano sigue una distribución de Poisson.

37. En un estudio para determinar el efecto de una droga sobre la agresividad, se formaron dos grupos: el A, al que se administró la droga, y el B que recibió un placebo. Después de la administración se realizó una prueba para medir la agresividad. Las puntuaciones obtenidas, a mayor puntuación mayor agresividad, fueron:

grupo A	10	8	12	16	5	9	7	11	6
grupo B	12	15	20	18	13	14	9	16	

Expresar el estudio en términos estadísticos formales y establecer la conclusión.

Solución:

Está claro que para comparar las dos poblaciones debemos usar un contraste no paramétrico puesto que los resultados son en realidad cualitativos. Usaremos el test de la suma de rangos.

valor	5	6	7	8	9	9	10	11	12
rango	1	2	3	4	5.5	5.5	7	8	9.5
muestra	A	A	A	A	A	B	A	A	A

valor	12	13	14	15	16	16	18	20
rango	9.5	11	12	13	14.5	14.5	16	17
muestra	B	B	B	B	B	A	B	B

$$\begin{cases} H_0 : M_A = M_B \\ H_1 : M_A < M_B \end{cases}$$

$$\begin{cases} W_1 = 54.5 \quad U_1 = W_1 - \dfrac{(1+n_1)\cdot n_1}{2} = 54.5 - 45 = 9.5 \\ W_2 = 98.5 \quad U_2 = W_2 - \dfrac{(1+n_2)\cdot n_2}{2} = 98.5 - 36 = 62.5 \end{cases} \quad U = 9.5$$

$$RC_{0.05} = \left[0, U_{n_2,n_1;0.05}\right] = [0,18] \quad RC_{0.01} = \left[0, U_{n_2,n_1;0.01}\right] = [0,11]$$

CONCLUSIÓN: como el valor del estadístico cae en la región crítica concluimos que tenemos evidencia estadística suficiente con confianza del 99% de que los resultados medianos de los dos tratamientos son diferentes. La droga A tiene efecto sobre la agresividad, la disminuye.

38. La tabla adjunta representa el rendimiento de la cosecha de un cereal en 10 pares de parcelas, con un abono fosfórico y la correspondiente sin él, siendo las demás condiciones las mismas. Admitiendo que la distribución de rendimientos es normal, ¿se puede concluir la eficacia del abono fosforado?

parcela 1	6.5	5.6	6.6	6.1	5.8	6.0	6.1	6.3	6.1	6.6
parcela 2	5.4	5.8	5.4	5.8	5.7	5.4	5.7	6.0	5.3	6.0

Solución:

Sea X = rendimiento de una parcela con abono fosfórico; $X \in N(\mu_1, \sigma_1)$.
Sea Y = rendimiento de una parcela sin abono fosfórico; $Y \in N(\mu_2, \sigma_2)$.
Tenemos una situación clara de datos emparejados

$$\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$$

Sea $D=X-Y$, tenemos

$$D \in N(\mu_1 - \mu_2, \sigma_D) \Rightarrow \overline{D} \in N(\mu_1 - \mu_2, \frac{\sigma_D}{\sqrt{n}}) \Rightarrow \frac{\overline{D} - (\mu_1 - \mu_2)}{\frac{S_D}{\sqrt{n}}} \in t_{n-1}$$

$d_i = x_i - y_i$	1,1	-0,2	1,2	0,3	0,1	0,6	0,4	0,3	0,8	0,6

$$\begin{cases} \bar{D} = \dfrac{\sum d_i}{10} = 0,52 \\ S_D = \sqrt{\dfrac{n\sum d_i^2 - (\sum d_i)^2}{n(n-1)}} = 0,4341 \end{cases} \Rightarrow \frac{\bar{D} - (\mu_1 - \mu_2)}{\frac{S_D}{\sqrt{n}}} = 3,78$$

$$RC_{0,05} = \left[t_{9;0,05}\ ,\ +\infty\right) = \left[1,833\ ,\ +\infty\right); \qquad RC_{0,01} = \left[t_{9;0,01}\ ,\ +\infty\right) = \left[2,821\ ,\ +\infty\right)$$

CONCLUSIÓN: como el estadístico cae en ambas regiones críticas, hay evidencia estadística de que el rendimiento por parcela con abono fosfórico es superior, a nivel de significación de hasta el 0,01.

39. Contrastar si la muestra siguiente de duración de vida puede suponerse exponencial de parámetro 0,089:

16 8 10 12 6 10 20 7 2 24

Solución:

Sea X = duración de vida de un sujeto

$$\begin{cases} H_0 : X \in Exp(0,089) \\ H_1 : X \quad \text{no sigue ese modelo} \end{cases}$$

Como la muestra es pequeña usaremos el test de bondad de ajuste de Kolmogorov-Smirnov:

$$F(x) = P(X \le x) = \int_0^x 0,089e^{-0,089x}dx = \left[-e^{-0,089x}\right]_0^x = 1 - e^{-0,089x}$$

$$G(2) = 0.1 \qquad F(2) = 0,1630 \qquad \left|F(2) - G(2)\right| = 0,0630$$

$$G(6) = 0.2 \qquad F(6) = 0,4137 \qquad \left|F(6) - G(6)\right| = 0,2137$$

$$G(7) = 0.3 \qquad F(7) = 0,4636 \qquad \left|F(7) - G(7)\right| = 0,1636$$

$$G(8)=0.4 \qquad F(8)=0,5093 \qquad |F(8)-G(8)|=0,1093$$

$$G(10)=0.6 \qquad F(10)=0,5893 \qquad |F(10)-G(10)|=0,0106$$

$$G(12)=0.7 \qquad F(12)=0,6563 \qquad |F(12)-G(12)|=0,0437$$

$$G(16)=0.8 \qquad F(16)=0,7592 \qquad |F(16)-G(16)|=0,0407$$

$$G(20)=0.9 \qquad F(20)=0,8313 \qquad |F(20)-G(20)|=0,0683$$

$$G(24)=1 \qquad F(24)=0,8812 \qquad |F(24)-G(24)|=0,1181$$

$$Máx\left\{|F(x)-G(x)|\right\}=0,2137$$

$$RC_\alpha=\left[D_{10;\alpha}\ ,+\infty\right]=\begin{cases}\left[0,669\ ,+\infty\right] & \text{para } \alpha=0,01\\ \left[0,410\ ,+\infty\right] & \text{para } \alpha=0,05\\ \left[0,368\ ,+\infty\right] & \text{para } \alpha=0,10\end{cases}$$

CONCLUSIÓN: el estadístico no cae en la región crítica, luego podemos decir que nuestra información no nos permite dudar fundadamente de que la variable sea una exponencial de parámetro 0,089.

40. El producto nacional bruto (PNB) *per capita* es reconocido como un estimador del nivel de vida de una nación. A su vez, se sabe que el consumo de energía eléctrica *per capita* es un buen indicador del PNB *per capita*. A continuación se da una lista de algunas naciones con sus correspondientes datos del PNB, en dólares, y de consumo de energía en kWs/h *per capita*.

PAÍS	PNB	KWS/H
India	110	81
Ecuador	240	399
Colombia	290	503
Perú	330	330
Chile	510	370
Costa Rica	510	754
Méjico	580	461

Panamá	660	1233
España	820	1013
Venezuela	1000	715
Argentina	1060	2487
URSS	1200	2007
Japón	1430	2377
Reino Unido	1890	3796
Alemania Federal	2190	3205
Francia	2460	2287
Estados Unidos	4240	6612

Hacer un estudio de la regresión lineal entre el consumo de energía y el PIB:

a) dar la recta de regresión

b) estudiar la validez del modelo lineal

c) sabiendo que el consumo de energía de Canadá es 2310 Kws/h. predecir su PNB; decir un intervalo de confianza al 95% para el verdadero valor del PNB per cápita de Canadá.

Solución:

Tenemos un problema de regresión lineal simple.

Sean X = consumo de energía *per capita*, (Kw/h)

Y = producto nacional bruto *per capita*, PNB

$$\left.\begin{array}{l}\sum x_i = 28.630 \\ \sum x_i^2 = 93.997.192\end{array}\right\} S_{xx} = \frac{n\sum x_i^2 - (\sum x_i)^2}{n} = 45.780904$$

$$\left.\begin{array}{l}\sum y_i = 19.520 \\ \sum y_i^2 = 40.233.200\end{array}\right\} S_{yy} = \frac{n\sum y_i^2 - (\sum y_i)^2}{n} = 17.819647$$

$$\sum x_i y_i = 59.857.520\} \quad S_{yy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n} = 26.983544$$

a) $$\left.\begin{array}{l} b = \dfrac{S_{xy}}{S_{xx}} = 0,5894 \\ a = \dfrac{\sum y_i - b\sum x_i}{n} = 155,606 \end{array}\right\} \quad \hat{y}_{/x} = 155,606 + 0,5894x$$

b) $$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = 0,9447 \quad \Rightarrow \quad r^2 = 0,8925$$

Por ser el coeficiente de determinación, r^2, mayor que 0,75 aceptamos el modelo lineal, ya que podemos decir que más del 89% de la variabilidad de las y_i respecto a su media es explicable por la relación lineal con los x_i.

c) Si el consumo de energía en Canadá son 2310 kWs./h. el PNB *per capita* estimado será:

$$\overline{y}_{2310} = 155,606 + 0,5894 \cdot 2310 = 1517,1202 \quad \$$$

Y el intervalo de confianza para el PNB *per capita* de Canadá será:

$$I_{\mu_{Y/x}} = \left[\overline{y} \pm S\sqrt{\frac{1}{n} + \frac{(x - x_o)^2}{S_{xx}}} \cdot t_{n-2;\, \alpha/2} \right] =$$

$$[1517,1202 \pm 184,6970] = [1332,4233 \;,\; 1701,8171]$$

Donde

$$\begin{cases} S = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{S_{yy} - bS_{xy}}{n-2}} = 357,3556 \\ \overline{X} = \dfrac{\sum x_i}{n} = 1684,1176 \end{cases}$$

41. Los ingenieros consideran válido suponer que la dureza de un aluminio tratado térmicamente está normalmente distribuida con una desviación estandar igual a 0.5. Deseamos probar la hipótesis H_0: m = 13.7 contra H_1: m≠ 13.7. La regla de decisión adoptada es: a) tomar una muestra de 30 piezas de aluminio; b) calcular la media muestral; c) si la media muestral es menor que 13.60 o mayor que 13.80 rechazamos H_0. En este caso calcular α y luego calcular β suponiendo que la verdadera dureza media del metal es 13.85. Comentar esta regla de decisión.

Solución:

$$\begin{cases} H_0 : m = 13.7 \\ H_1 : m \neq 13.7 \end{cases}$$

La regla de decisión nos dice que la región crítica es:

$$RC = (-\infty, 13.6] \cup [13.8, +\infty) \text{ para } \overline{X}$$

Si H_0es cierta

$$\overline{X} \in N\left(13.7, \frac{0.5}{\sqrt{30}}\right)$$

entonces

$$\alpha = P(\text{rechazar}H_0 \,/\, H_0\text{cierta}) = P(13.6 < \overline{X}) + P(\overline{X} > 13.8) =$$
$$= P\left(\frac{13.6 - 13.7}{\frac{0.5}{\sqrt{30}}} < Z\right) + P\left(Z > \frac{13.8 - 13.7}{\frac{0.5}{\sqrt{30}}}\right) =$$
$$= 2 \cdot P(Z < -0.2\sqrt{30}) = 2 \cdot P(Z < -1.09) = 0.2758$$

Si el verdadero valor de m es 13,85 entonces

$$\overline{X} \in N\left(13.85, \ \frac{0.5}{\sqrt{30}}\right)$$

$$\beta = P(\text{Aceptar } H_0 / H_0 \text{ falsa}) = P(13.6 \leq \overline{X} \leq 13.8) =$$
$$= P\left(\frac{13.6-13.85}{0.5/\sqrt{30}} \leq Z \leq \frac{13.8-13.85}{0.5/\sqrt{30}}\right) = P(-0.5\sqrt{30} \leq Z \leq 0.1\sqrt{30}) =$$
$$= P(-2.738 \leq Z \leq 0.547) = 0.707 - 0.0028 = 0.704$$

Vemos que la regla de decisión no es buena porque ambas probabilidades de error son altas.

42. Para una red de alto voltaje se necesitan cables uniformes de gran resistencia a la tensión. Tenemos 4 tipos de cables y los resultados de medir la resistencia a la tensión de doce cables de cada tipo se indican en la tabla. Como la resistencia a la tensión de cada cable es de 340 Kg. aproximadamente, para simplificar los cálculos hemos restado de todas las observaciones 340. Interesa contrastar la uniformidad de los cables.

cable 1	cable 2	cable 3	cable 4
5	-11	0	-12
-13	-13	-10	4
-5	-8	-15	2
-2	8	-12	10
-10	-3	-2	-5
-6	-12	-8	-8
-5	-12	-5	-12
0	-10	0	0
-3	5	-4	-5
2	-6	-1	-3
-7	-12	-5	-3
-5	-10	-11	0

Solución:

Se trata de 4 poblaciones independientes. Sea X_i= resistencia a la tensión del cable i, i=1, 2, 3, 4, y de cada una de ellas tenemos una muestra de tamaño 12. Como no se dice que las variables sean normales utilizaremos un test no paramétrico: el de Kruskall-Wallis.

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1 : \text{no} \quad \text{ocurre } H_0 \end{cases}$$

Asignamos rango a cada uno de los valores muestrales:

cable nº 1	cable nº 2	cable nº 3	cable nº 4
45,5	10,5	39	6,5
2,5	2,5	13,5	44
25	17	1	42,5
34,5	47	6,5	48
13,5	31,5	34,5	25
20,5	6,5	17	17
25	6,5	25	6,5
39	13,5	39	39
31,5	45,5	29	25
42,5	20,5	36	31,5
19	6,5	25	31,5
25	13,5	10,5	39

$$R_1 = 323,5 \qquad R_2 = 221 \qquad R_3 = 276 \qquad R_4 = 355,5$$

$$\sum R_i = \frac{(1+48)\cdot 48}{2} = 1176$$

$$H = \frac{12}{n(n+1)}\sum \frac{R_i^2}{n_i} - 3(n+1) = 4,3816$$

Si la hipótesis nula es cierta el estadístico H debe comportarse como una X^2 con k-1= 3 grados de libertad.

$$RC_{0,05} = \left[X^2_{3;\,0,05}\ ,\ +\infty\right) = \left[7,82\ ,\ +\infty\right)$$

CONCLUSIÓN: el estadístico no cae en región crítica. No hay evidencia estadística de que haya diferencia entre las resistencias medias de los cuatro cables.

43. La tabla muestra las producciones por acre de cuatro semillas sembradas en campos tratados con tres fertilizantes distintos. Con nivel de significación 0.01 determinar si hay diferencias en producción por acre: a) debida a los fertilizantes y b) debida a las semillas.

	Finca 1	**Finca 2**	**Finca 3**	
A	**12**	**15**	**14**	$T_{\cdot 1}=41$
B	**15**	**19**	**18**	$T_{\cdot 2}=52$
C	**10**	**12**	**15**	$T_{\cdot 3}=37$
D	**14**	**11**	**12**	$T_{\cdot 4}=37$
	$T_{1\cdot}=51$	$T_{2\cdot}=57$	$T_{3\cdot}=59$	$T_{\cdot\cdot}=167$

Solución:

Aplicamos una ANOVA II debido a que nos piden los efectos debidos a los tratamientos (fertilizantes) y debidos a los bloques (semillas). Para poder aplicarlo correctamente suponemos que las variables X_A, X_B, X_C y X_D que nos dan las producciones con los distintos fertilizantes son todas normales, independientes y con la misma varianza σ^2.

Calculamos, en la tabla, los totales parciales de filas $T_{i\cdot}$ y columna $T_{\cdot j}$ y el gran total $T_{\cdot\cdot}$

$$S_{yy} = \sum_{i,j} y_{ij}^2 - \frac{T_{\cdot\cdot}^2}{bk} \qquad SST = \sum_i \frac{T_{i\cdot}^2}{b} - \frac{T_{\cdot\cdot}^2}{bk}$$

$$SSB = \sum_j \frac{T_{\cdot j}^2}{k} - \frac{T_{\cdot\cdot}^2}{bk} \qquad SSE = S_{yy} - SST - SSB$$

Fuente de variabilidad	*Grados libertad*	*Estadístico*	*Cuadrados medios*
Tratamientos	k-1=2	SST=8,67	$MST = \frac{SST}{2} = 4,335$
Bloques	b-1=3	SSB=50,25	$MSB = \frac{SSB}{3} = 16,75$
Aleatoria	(b-1)(k-1)=6	SSE = 22	$MSE = \frac{SSE}{6} = 0,386$
Total	bk-1	SS_{yy}=80,92	

a)

$$\begin{cases} H_0 : \mu_{F_1} = \mu_{F_2} = \mu_{F_3} & \text{(los efectos medios de los tratamientos son iguales)} \\ H_1 : \text{ no ocurre } H_0 \end{cases}$$

$$RC_{0.01} = \left[F_{2,\,6;0.01},\ +\infty\right) = \left[10,92\ ,\ +\infty\right)$$

CONCLUSIÓN: como el estadístico de este contraste, MST, no cae en región crítica, no hay evidencia de que haya diferencias entre los fertilizantes, con confianza del 99%.

b)

$$\begin{cases} H_0' : \mu_A = \mu_B = \mu_C & \text{(no hay diferencia entre las semillas)} \\ \text{H}_1' : \text{ no ocurre H}_0' \end{cases}$$

$$RC_{0,01} = \left[F_{3,6;0,01}, +\infty \right) = \left[9,78 \ , +\infty \right)$$

CONCLUSIÓN: como el estadístico de este contraste, MSB, no cae en región crítica, no hay evidencia de que haya diferencias entre los tipos de semillas, con confianza del 99%.

44. En un experimento diseñado para saber si hay diferencia en la eficacia de 5 máquinas A, B, C, D y E, cinco operarios **expertos** trabajaron con cada una de las máquinas, todos el mismo tiempo con cada una. Los resultados se recogen en la tabla, en nº de unidades producidas. Contrastar la hipótesis de que no hay diferencia entre ellas.

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Solución:

Sea X_i = nº de unidades producidas con la máquina *i*-ésima, *i*=1, 2, 3, 4.

Como se dice que los trabajadores son igualmente expertos, las variables se pueden considerar independientes y las posibles diferencias en número de unidades producidas depende sólo de las máquinas, por tanto tenemos una situación de cuatro v. a. independientes, como no sabemos que sean normales, usamos un test no paramétrico: el test de Kruskall-Wallis.

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\ H_1 : \text{ las medias no son todas iguales} \end{cases}$$

Asignamos rango a cada uno de los datos de la muestra:

A	17,5	21	24	1	6,5	R_1=70
B	21	6,5	12	6,5	2,5	R_2=48,5
C	10	25	14	23	21	R_3=93
D	2,5	11	9	14	4	R_4=40,5
E	14	16	19	17,5	6,5	R_5=73

El estadístico del contraste es:

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1) = 6,4449$$

Si la hipótesis nula es cierta, H debe comportarse como una ji-cuadrada con k-1= 4 grados de libertad y será

$$RC_{0,05} = \left[X^2_{4;0,05} \ , +\infty \right) = \left[9,488 \ , +\infty \right)$$
$$RC_{0,01} = \left[X^2_{4;0,01} \ , +\infty \right) = \left[13,227 \ , +\infty \right)$$

CONCLUSIÓN: no hay diferencias significativas entre las máquinas en cuanto al nº medio de unidades producidas, al menos no son detectables con este experimento.

45. En dos provincias españolas se realizó una encuesta a 70 y 60 personas respectivamente, sobre la distribución de gastos semanales en alimentación (en miles de pts.), obteniéndose:

	0-5	5-10	10-15	15-20	20-25	total
prov. A	9	25	21	11	4	70
prov. B	5	20	15	11	9	60

a) Estudiar la homogeneidad del gasto en alimentación en las dos provincias.

b) Describir detalladamente todos los pasos del contraste.

Solución:

Se trata de estudiar si es aceptable que el gasto en alimentación es homogéneo en las dos provincias, para ello aplicamos el test de homogeneidad de la X^2.

$$\begin{cases} H_0 : P(\text{gasto} \quad k / \text{provincia } A) = P(\text{gasto } k / \text{provincia } B) \ \ \forall k \\ H_1 : \text{no ocurre } H_0 \end{cases}$$

Construimos la tabla de las frecuencias esperadas E_{ij}, supuesto cierta H_0, donde cada una de dichas frecuencias esperadas se obtiene:

$$E_{ij} = \frac{\text{total fila } i \text{ x total columna } j}{\text{gran total}}$$

	gasto alimentación (miles de pts.)					
	0-5	5-10	10-15	15-20	20-25	total
prov.A	7,54	24,38	19,38	11,85	7	70
prov.B	6,46	20,77	16,62	10,15	6	60

El estadístico del contraste es

$$\sum \frac{O_{ij}^2}{E_{ij}} - N$$

que debe comportarse como una X^2 con (5-1)(2-1) = 4 grados de libertad, si la hipótesis nula es cierta. Tenemos:

$$\sum \frac{O_{ij}^2}{E_{ij}} - N = 133{,}8768 - 130 = 3{,}8768$$

$$RC = \left[X^2_{4;0.05}, \ +\infty\right) = [9{,}488, \ +\infty)$$

CONCLUSIÓN: el gasto es alimentación es homogéneo en las dos provincias.

46. Con objeto de analizar si existe relación entre el consumo de energía eléctrica Y_i (en kW/h) y el volumen de producción X_i (en millones de pts.) de una empresa, se ha obtenido la siguiente información:

X_i millones de pts.	2.1	3.2	5.7	8.5	8.9	10.2	10.3	12.3	13.2
Y_i kW/h	5.5	7.2	7.5	9.2	9.0	9.3	8.7	10.2	11.1

a) Determinar la ecuación de la recta que mejor explique el consumo de electricidad en función del volumen de producción.

b) Si la empresa tuviese un volumen de producción de 9.2 millones de pts. ¿entre qué valores estaría el gasto en electricidad con confianza del 90%?

Solución:

Tenemos un problema de regresión lineal simple, en el que consideramos el volumen de producción como variable independiente.

$$\left.\begin{array}{l}\sum x_i = 74,4 \\ \sum x_i^2 = 734,26\end{array}\right\} \quad S_{xx} = \frac{n\sum x_i^2 - (\sum x_i)^2}{n} = 119,22$$

$$\left.\begin{array}{l}\sum y_i = 77,7 \\ \sum y_i^2 = 693,41\end{array}\right\} \quad S_{yy} = \frac{n\sum y_i^2 - (\sum y_i)^2}{n} = 22,6$$

$$\sum x_i y_i = 692,09 \;\} \quad S_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n} = 49,77$$

a)
$$\left.\begin{array}{l} b = \dfrac{S_{xy}}{S_{xx}} = 0,4174 \\[2ex] a = \dfrac{\sum y_i - b\sum x_i}{n} = 5,1823 \end{array}\right\} \Rightarrow \overline{y}_{/x} = 5,1823 + 0,4174x$$

b) $$\left.\begin{array}{l} S=\sqrt{\dfrac{S_{yy}-bS_{xy}}{n-2}}=0,5103 \\ t_{n-2;\ \alpha/2}=t_{7;\,0,05}=1,895 \\ \overline{x}=\dfrac{74,4}{9}=8,2667 \end{array}\right\}$$

$$I_{y/9,2\,;0,10}=\left[\overline{y}/_{9,2}\pm t_{9-2;\,0,05}\cdot S\cdot\sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}}\right]=$$

$$=[9,023\pm 1,0226]=[8,0003\ ,\ 10,0456]$$

47. Supongamos que una caja contiene circuitos de tres colores: rojo, marrón y azul y que se desea contrastar la hipótesis nula de que las proporciones de circuitos de cada uno de los tres colores son iguales, frente a la alternativa de que esas proporciones no son todas iguales. Supongamos que el método de contraste elegido consiste en seleccionar al azar tres circuitos de la caja y rechazar la hipótesis nula si, y solo si, al menos dos de los circuitos de la muestra son del mismo color.

a) Calcular el nivel de significación del contraste.

b) Calcular la potencia del contraste si $1/7$ de los circuitos son rojos, $2/7$ son marrones y los restantes azules.

Solución:

Antes que nada suponemos que en la caja hay un número suficientemente grande de circuitos de modo que las extracciones pueden ser consideradas independientes. El experimento consiste en extraer tres circuitos y mirar su color, lo que nos permite definir las variables aleatorias:

X_1 = nº de circuitos rojos de los 3 extraídos.

X_2 = nº de circuitos marrones de los 3 extraídos.

X_3 = nº de circuitos azules de los 3 extraídos.

$$\begin{cases} H_0: p_1=p_2=p_3=1/3 \\ H_1: \text{ no ocurre } H_0 \end{cases}$$

siendo $p_1 = P(\text{rojo}); p_2 = P(\text{marrón}); p_3 = P(\text{azul})$ y se acepta H_0 si y sólo si ocurre que $X_1 = 1$ y $X_2 = 1$ y $X_3 = 1$.

a) $$\alpha = P(\text{Rechazar } H_0 / H_0 \ \text{ cierta}) =$$
$$= P(\textit{no} \text{ son de distinto color}/p_1 = p_2 = p_3 = 1/3) =$$
$$= 1 - P(X_1 = 1,\ X_2 = 1,\ X_3 = 1) = 1 - \frac{3!}{1!\cdot 1!\cdot 1!}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3} = \frac{7}{9} = 0.77$$

b) $$\text{Potencia} = 1 - \beta = 1 - P(\text{aceptar } H_0 / p_1 = 1/7, p_2 = 2/7,\ p_3 = 4/7) =$$
$$= 1 - P(\mathrm{X}_1 = 1,\ X_2 = 1,\ X_3 = 1) = 1 - \frac{3!}{1!\cdot 1!\cdot 1!}\cdot\frac{1}{7}\cdot\frac{2}{7}\cdot\frac{4}{7} = \frac{99}{147} = 0.67$$

Comentario: a simple vista se ve que no es un buen test. Si fuera cierta que las proporciones de los tres tipos de componentes fueran iguales, la probabilidad de que sacando tres fueran distintos es demasiado baja, casi nunca vamos a aceptar la hipótesis nula.

48. Se ha aplicado un cierto test de memoria a un gran nº de estudiantes de bachillerato, que dio como desviaciones típicas para las alumnas y alumnos 33,5 y 38,5, respectivamente. Aplicando el test a una muestra aleatoria simple de 18 alumnos dio un promedio de 183,7 y a una muestra aleatoria simple de 25 alumnas dio una media de 165,4. Suponiendo que las poblaciones son independientes y normales, ¿son estos resultados significativos de una diferencia de memoria entre ambos sexos?

Solución:

Sea X = puntuación de una alumna, $X \in N(\mu_1\ , 33{,}5)$

Sea Y = puntuación de un alumno, $Y \in N(\mu_2\ , 38{,}5)$

De donde $\bar{X} \in N(\mu_1\ , 33{,}5/5)$ y $\bar{Y} \in N(\mu_2\ , 38{,}5/\sqrt{18})$

El contraste es:

$$\begin{cases} H_0 : \mu_2 - \mu_1 \leq 0 \\ H_1 : \mu_2 - \mu_1 > 0 \end{cases}$$

Si H_0 es cierta

$$\bar{Y} - \bar{X} \in N(\mu_2 - \mu_1 \,,\, \sqrt{\frac{(38,5)^2}{18} + \frac{(33,5)^2}{25}})$$

y por tanto

$$\frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{\sqrt{\frac{(38,5)^2}{18} + \frac{(33,5)^2}{25}}} \approx Z$$

que toma el valor 1,6223 y $P(Z > 1,6223) = 0,054$.

CONCLUSIÓN: a niveles de significación hasta 0,05 concluimos que no hay evidencia estadística de una diferencia de memoria significativa entre los dos sexos.

49. Para contrastar la hipótesis de que una moneda está equilibrada mediante lanzamientos sucesivos, se imponen las siguientes restricciones:

a) La prob. de error tipo I debe ser como mucho 0,05.

b) La prob. de error tipo II cuando p difiera de 0,5 en 0,1 o más, debe ser a lo sumo 0,05.

Determinar el tamaño mínimo de la muestra necesario y enunciar la regla de decisión.

Solución:

Sea X = nº de caras en los n lanzamientos, $X \in N(n, p)$

$$\begin{cases} H_0 : p = 0,5 \\ H_1 : p \neq 0,5 \end{cases}$$

Sabemos que

$$\hat{p} = \frac{\text{n}^\text{o}\ \text{caras}}{\text{n}^\text{o}\ \text{lanzamientos}} \in N\left(p \,,\, \sqrt{\frac{p(1-p)}{n}} \right)$$

porque suponemos que n será mayor que 30. Supongamos que se acepta que la moneda está equilibrada si la proporción muestral de caras está entre A y B.

$$\alpha = P(\text{error tipo I})=\text{P}(\text{rechazar } H_0/\, H_0 \text{ cierta}) =$$

$$= P\left(\hat{p} > B/p = 0{,}5\right) + P\left(\hat{p} < A/p = 0{,}5\right) = 2P\left(Z < \frac{A\text{ - }0{,}5}{\sqrt{0.25/n}}\right) = 0{,}05$$

$$\Rightarrow \frac{A\text{ - }0{,}5}{\sqrt{0.25/n}} = -1{,}96 \quad (1)$$

$$\beta = P(\text{error tipo } II) = P(\text{aceptar } H_0 \,/\, H_0 \text{ falsa}) =$$
$$= P(A < \hat{p} < B \,/\, p = 0{,}5 \pm 0{,}1) = 0.05$$

Si p = 0.5+0.1= 0,6 entonces $\hat{p} \in N\left(0.6, \sqrt{0.24/n}\right)$ y tenemos

$$\beta_{0.6} = P(A < \hat{p} < B \,/\, p = 0.6) = P\left(\frac{A - 0{,}6}{\sqrt{0{,}24/n}} < Z < \frac{B - 0{,}6}{\sqrt{0{,}24/n}}\right) = 0{,}05$$

$$\Rightarrow P\left(Z < \frac{A - 0{,}6}{\sqrt{0{,}24/n}}\right) = \frac{0.95}{2} = 0.475 \Rightarrow \frac{A - 0{,}6}{\sqrt{0{,}24/n}} = -0.06 \quad (2)$$

Uniendo las dos ecuaciones (1) y (2) en A y en n, y resolviendo el sistema:

$$\begin{cases} \dfrac{A - 0{,}5}{\sqrt{0.25/n}} = -1{,}96 \\ \dfrac{A - 0{,}6}{\sqrt{0{,}24/n}} = -0.06 \end{cases} \Rightarrow \begin{cases} A = 0.5 - \dfrac{0.98}{\sqrt{n}} \\ A = 0.6 - \dfrac{0.0294}{\sqrt{n}} \end{cases}$$

de donde:

$$\begin{cases} n = 100 \\ A = 0{,}402 \\ B = 0.598 \end{cases}$$

CONCLUSIÓN: el nº de lanzamientos necesarios para conseguir las condiciones impuestas son 100 o más. La regla de decisión es no aceptar la hipótesis nula si la proporción muestral de caras es mayor que 0,598 o menor que 0,402. O, lo que es lo mismo, se aceptará que la moneda está equilibrada si en 100 lanzamientos el número de caras está entre 40 y 59, ambas inclusive.

50. Se desea probar la eficacia de un fármaco sobre la angina de pecho. Se toma una muestra de 10 enfermos y se mide el tiempo que son capaces de realizar un determinado esfuerzo hasta sentir dolor. Los datos en segundos son:

Enfermo	1	2	3	4	5	6	7	8	9	10
Antes	720	600	160	245	125	185	80	136	124	226
Después	780	635	225	310	75	302	80	184	122	288

Se pide:

a) Estudiar si hay diferencias significativas, suponiendo normalidad.

b) Ídem sin suponer normalidad.

Solución:

Como se trata claramente de datos emparejados, calculamos las diferencias $d_j = y_j - x_j$.

d_j	60	35	65	65	-50	117	0	48	-2	62

a) Si suponemos normalidad tenemos:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases}$$

(Tomamos el contraste unilateral porque a simple vista parece que los datos "después" son mayores que los datos "antes" y precisamente lo que queremos ver es si hay evidencia estadística de eso).

Tenemos que

$$D \in N(\mu_D\ , \sigma_D) \Rightarrow \overline{D} \in N(\mu_D\ , \frac{\sigma_D}{\sqrt{9}})$$

y por tanto el estadístico del contraste será

$$\frac{\overline{D}}{S_D/3} = \frac{40}{46.68/3} = 2.709$$

que debe ser una t_8. La región crítica será:

$$RC_{0.05} = \left[t_{8;\ 0.05},\ +\infty\right) = \left[1{,}860,\ +\infty\right)$$

CONCLUSIÓN: como el estadístico cae en región crítica la conclusión es que hay evidencia de que el fármaco es efectivo, en cuanto que alarga la capacidad de realizar esfuerzo antes de sentir dolor.

b) Si consideramos que las poblaciones quizás no son normales tendremos que aplicar un test no paramétrico, y como no sabemos nada de la simetría de la población en torno a su mediana aplicaremos el test de los signos.

$$\begin{cases} H_0 : M_D = 0 \\ H_1 : M_D > 0 \end{cases} \text{ y se tiene } \begin{cases} N^+ = 7 \\ N^- = 2 \end{cases} \quad P(N^- \leq 2) = 0{,}090 \text{ ya que}$$

$$N^- \in B(9\ ,\ 0{,}5)$$

CONCLUSIÓN: con confianza del 95%, no tenemos evidencia para afirmar que el fármaco alarga la duración del esfuerzo antes de sentir dolor.

51. A los datos obtenidos al observar las roturas de un telar medidas cada 10 000 pasadas, se ajustó una distribución de Poisson. Comprobar la bondad del ajuste.

nº de roturas	0	1	2	3	4	5	6
frecuencias	3	39	48	16	5	1	1

Solución:

Sea X = nº de roturas del hilo cada 10 000 pasadas

$$\begin{cases} H_0 : X \in P(\lambda) \\ H_1 : X \notin P(\lambda) \end{cases}$$

estimamos el valor del parámetro

$$\hat{\lambda} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{214}{113} = 1{,}89$$

Para calcular las frecuencias esperadas tenemos:

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ para } k=0, 1,...,4 \qquad p_5 = P(X \geq 5) = 1 - \sum_{k=1}^{4} p_k$$

nº roturas	0	1	2	3	4	5	6 o más
p_i	0,15	0,28	0,27	0,17	0,08	0,03	0,02
E_i=n. p_i	17,00	32,21	30,49	19,25	9,14	3,45	1,47

Como hay dos grupos (28%) con frecuencia esperada menor que 5 unimos los dos últimos grupos y tenemos

Nº roturas	0	1	2	3	4	5 o más
O_i	3	39	48	16	5	2
E_i	17,00	32,21	30,49	19,25	9,14	4,92

El estadístico es:

$$\sum_{i=0}^{5} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=0}^{5} \frac{O_i^2}{E_i} - n = 140{,}562 - 113 = 27{,}15$$

y la región crítica es

$$RC_{0.05} = \left[\mathrm{X}^2_{4;\,0.05},\ +\infty\right) = \left[9{,}49,\ +\infty\right)$$

(son 4 grados de libertad porque hay 6 grupos y hemos estimado el parámetro λ).

CONCLUSIÓN: el estadístico cae en la región crítica, debemos concluir por tanto que nos quedamos con la hipótesis alternativa. A nivel de significación de 0,05 las roturas del hilo no se ajustan a una distribución de Poisson.

52. Un proceso de producción emplea 5 máquinas en sus 3 procesos de desplazamiento. Se clasificó una muestra aleatoria de 164 fallas de acuerdo con la máquina y el tipo de desplazamiento en que ocurrió la falla obteniéndose los datos de la tabla. De acuerdo con esta información, ¿existe fundamento para dudar de la independencia entre la operación de desplazamiento y la falla de la máquina? (α = 0,01).

	Máquina				
Desplazamiento	A	B	C	D	E
tipo 1	10	12	8	14	8
tipo 2	15	8	13	8	11
tipo 3	12	9	14	12	10

Solución:

Se trata de una tabla de frecuencias, o sea una tabla de contingencia 3x4. Queremos estudiar la independencia.

$$\begin{cases} H_0 : p_{ij} = p_{i.} \cdot p_{.j} \quad \forall i = 1,2,3 \quad \forall j = 1,2,3,4,5 \\ H_1 : \text{no ocurre } \mathrm{H_o} \end{cases}$$

El estadístico del contraste es

$$\sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_i} = \sum_{i,j} \frac{O_{ij}^2}{E_{ij}} - N \in \mathrm{X}^2_{(r-1)(k-1)}$$

Construimos la tabla de los

$$E_{ij} = \frac{\text{total fila i} \cdot \text{ total columna j}}{\text{gran total}}$$

Máquina

despl.	A	B	C	D	E	total
tipo 1	11,7317	9,1951	11,0975	10,7805	9,1951	52
tipo 2	12,4085	9,7256	11,7372	11,4024	9,7256	55
tipo 3	12,8600	10,0793	12,1646	11,8170	10,0793	57
total	37	29	35	34	29	164

Una vez hechas las operaciones el estadístico toma el valor 5,7122 y la región crítica es:

$$RC_{0.01} = \left[\mathrm{X}^2_{10;\,0.01}, +\infty\right) = \left[23{,}19, \ \ +\infty\right)$$

CONCLUSIÓN: el estadístico no cae en región crítica, por tanto no podemos rechazar la hipótesis nula. Los fallos de desplazamiento son independientes de la máquina con que se esté trabajando.

53. Se selecciona una muestra de 200 votantes y se encontró que 114 apoyan determinada propuesta. Obtener un intervalo de confianza del 96% para la proporción de la población votante que está a favor de la propuesta.

¿De qué tamaño debe tomarse la muestra si se quiere una confianza del 96% de que la proporción muestral estará dentro de 0,02 de la proporción verdadera de la población de votantes?

Solución:

El intervalo de confianza para p = proporción de votantes que apoyan la propuesta, es

$$I_{p;\,\alpha} = \left[\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot Z_{\alpha/2} \right]$$

El tamaño del error máximo viene dado por

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot Z_{\alpha/2} \leq \varepsilon \quad \text{de donde} \quad \hat{p}(1-\hat{p})\left(\frac{Z_{\alpha/2}}{\varepsilon}\right)^2 \leq n$$

tomamos para nuestro caso $\varepsilon = 0{,}02$,

$$n \geq \left(\frac{Z_{0.02}}{0.02}\right)^2 \cdot 0.57 \cdot 0.43 = 2575.08 \quad \text{o sea} \quad n \approx 2576$$

54. Los siguientes datos representan los tiempos de duración de películas producidas por dos compañías. Calcular un intervalo del 95% de confianza para σ_1^2 / σ_2^2.

Compañía A	103	94	110	87	98		
Compañía B	97	82	123	92	175	88	118

Solución:

Está claro que debemos suponer que ambas poblaciones son normales porque sólo bajo esta hipótesis se cumple que:

$$\frac{(n-1)\cdot S^2}{\sigma^2} \in \mathrm{X}^2_{n-1}$$

condición necesaria para poder escribir:

$$I_{\sigma_1^2/\sigma_2^2;\,\alpha} = \left[\frac{S_1^2}{S_2^2}\cdot F_{(n_2-1),(n_1-1);\,1-\alpha/2} \quad , \quad \frac{S_1^2}{S_2^2}\cdot F_{(n_2-1),(n_1-1);\,\alpha/2}\right]$$

donde:

$$S_1^2 = \frac{1}{n_1-1}\sum_i (x_i-\overline{X})^2 = 76{,}3 \qquad F_{6,4;0.95} = \frac{1}{F_{4,6;0.05}} = \frac{1}{4{,}53} = 0{,}22$$

y

$$S_2^2 = \frac{1}{n_2-1}\sum_i (y_i-\overline{Y})^2 = 1035{,}9 \qquad F_{6,4;0.05} = 6{,}16$$

De aquí

$$\Rightarrow I_{\sigma_1^2/\sigma_2^2;\,0{,}1} = [0{,}0162,\ 0{,}4537]$$

Este intervalo no contiene al 1, además está próximo al cero, lo que indica que la varianza de la primera muestra es claramente menor que la de la segunda.

55. Una empresa de automóviles quiere averiguar si el sexo de sus clientes tiene o no relación con el modelo elegido. Se toma una muestra de 2000 clientes y se tiene:

		modelo		
		A	B	C
sexo	mujer	340	400	260
	hombre	350	270	380

Solución:

Aunque los totales parciales de varones y mujeres son, en ambos casos, 1000, el enunciado no dice que éstos sean prefijados, se trata de un problema de independencia. La tabla es claramente una tabla de frecuencias.

$$\begin{cases} H_0 : p_{ij} = p_{i.} \cdot p_{.j} \quad \forall i = H, M \quad \forall j = A, B, C \\ H_1 : \text{no ocurre } H_0 \end{cases}$$

El estadístico del contraste es:

$$\sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i,j} \frac{O_{ij}^2}{E_{ij}} - N \in \mathrm{X}^2_{(r-1)(k-1)}$$

Donde E_{ij} son las frecuencias esperadas suponiendo que H_0 es cierta y que viene dado por:

$$E_{ij} = \frac{\text{total fila i} \cdot \text{total columna j}}{\text{gran total}}$$

que están en la tabla siguiente:

	A	B	C	totales
mujer	345	335	320	1000
varón	345	335	320	1000
totales	690	670	640	2000 = gran total

El estadístico toma el valor 47,87 y las regiones críticas son:

$$RC_{0,01} = \left[\mathrm{X}^2_{2;0,01} \text{ , } +\infty\right] = [9,22 \text{ , } +\infty]$$

$$RC_{0,05} = \left[\mathrm{X}^2_{2;0,05} \text{ , } +\infty\right] = [5,99 \text{ , } +\infty]$$

CONCLUSIÓN: el valor del estadístico cae dentro de la región crítica para ambos niveles de significación. Hay evidencia estadística de que el modelo de coche elegido no es independiente del sexo.

56. Un investigador afirma que el promedio de vida de los ratones puede ser ampliado hasta en 8 meses cuando se reducen las calorías en su comida aproximadamente un 40% del tiempo en que se les da de mamar. Las dietas restringidas son enriquecidas hasta niveles normales con proteínas y vitaminas. Una muestra aleatoria de 10 ratones es alimentada con una dieta normal y vive un promedio de 32,1 meses con desviación estándar de 3,2 meses, mientras que una muestra aleatoria de 15 ratones es alimentada con la dieta restringida y vive un promedio de 37,6 meses, con una desviación estándar de 2,8 meses. Con un nivel de significación de 0,05, probar la hipótesis de que el promedio de vida de los ratones con la dieta restringida, se incrementa 8 meses, en contra de la alternativa de que el incremento es menor de 8 meses. Suponer que las distribuciones de la duración de vida con las dietas regular y restringida tienen varianzas iguales y son normales.

Solución:

Sea X = vida de un ratón sin dieta $X \in N(\mu_1, \sigma)$.
Y = vida de un ratón con dieta $Y \in N(\mu_2, \sigma)$.

$$\begin{cases} H_0 : \mu_2 - \mu_1 = 8 \\ H_1 : \mu_2 - \mu_1 < 8 \end{cases}$$

El estadístico del contraste será:

$$\frac{(\bar{Y} - \bar{X}) - 8}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \in t_{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} = 8,78 \Rightarrow S_p = 2,96$$

Una vez hechas las operaciones el estadístico toma el valor -2,0688.

$$RC_{0,05} = \left(-\infty\ ,\ t_{23,\,0,95}\right] = \left(-\infty\ ,\ -1,714\right]$$

CONCLUSIÓN: el estadístico cae en la región crítica, luego rechazamos H_0. Con la información de que disponemos podemos afirmar, con nivel de significación 0.05, que la vida de los ratones sometidos a esta dieta se alarga menos de 8 meses.

57. Una zapatería es abastecida por cuatro fabricantes. Cada zapato es inspeccionado antes de ponerlo en venta. Hay tres defectos diferentes que causarían la devolución al fabricante. En cierta muestra se encontraron los siguientes defectos.

		Defecto		
		I	II	III
Fabricante	A	17	10	13
	B	10	10	10
	C	18	15	17
	D	15	5	10

¿Puede decirse que el tipo de defecto es independiente del fabricante a un nivel $\alpha = 0.01$?

Solución:

Tenemos una tabla de contingencia de 4x3 y se trata de aplicar un test de independencia. El contraste es:

$$\begin{cases} H_0 : p_{ij} = p_{i\cdot} p_{\cdot j} & \forall i = \text{A, B, C, D} \quad \forall j = \text{I, II, III} \\ H_1 : \text{ no ocurre } H_0 \end{cases}$$

El estadístico del contraste será:

$$\sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \sum_{i,j} \frac{O_{ij}^2}{E_{ij}} - N \in \mathrm{X}^2_{(r-1)(k-1)}$$

Construimos la tabla de las frecuencias esperadas:

$$E_{ij} = \frac{\text{total fila } i \times \text{total columna } j}{\text{gran total}}$$

	I	II	III	total
A	16	10.66	13.34	40
B	12	8	10	30
C	20	13.34	16.66	50
D	12	8	10	30
total	60	40	50	150

Al hacer operaciones el estadístico toma el valor 3.234 y la región crítica viene dada por:

$$RC_{0.01} = \left[X^2_{6;\,0.01},\ +\infty\right) = \left[16.81,\ +\infty\right)$$

CONCLUSIÓN: el estadístico no cae en la región crítica, por tanto no podemos rechazar la hipótesis nula. Los defectos son independientes del fabricante o al menos no podemos afirmar otra cosa con la información disponible y con significación 0.01.

58. Tenemos el contraste: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$

Calcular su potencia para dos m. a. s. de tamaños $n_1 = 10$ y $n_2 = 5$, sobre poblaciones normales, con $\alpha = 0,1$ y supuesto que $\sigma_1^2 = 2\sigma_2^2$.

Solución:

El contraste será:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

o lo que es lo mismo:

$$\begin{cases} H_0 : \sigma_1^2 / \sigma_2^2 = 1 \\ H_1 : \sigma_1^2 / \sigma_2^2 \neq 1 \end{cases}$$

y se tiene,

$$\frac{\sigma_1^2}{\sigma_2^2}\cdot\frac{S_2^2}{S_1^2}\in F_{n_2-1,n_1-1}$$

El estadístico del contraste es:

$$\frac{S_2^2}{S_1^2}$$

que será una F_{n_2-1,n_1-1} si la hipótesis nula es cierta.

Como el contraste es bilateral la región de aceptación será:

$$RA_{0,1}=\left(F_{n_2-1,n_1-1;\,0,95}\,,F_{n_2-1,n_1-1;\,0,05}\right)=\left(\frac{1}{F_{9,\,4\,;\,0,05}}\;,\;F_{4,9;\,0,05}\right)=\left(\;0{,}254\;,\;2{,}64\right)$$

Si la verdadera relación es $\sigma_1^2=2\sigma_2^2$ entonces:

$$\frac{\sigma_1^2}{\sigma_2^2}\frac{S_2^2}{S_1^2}=\frac{2S_2^2}{S_1^2}\in F_{n_2-1,\,n_1-1}$$

y se tiene

$$\beta=P(\text{error}\quad\text{tipo}\quad\text{II})=P(\text{aceptar}\quad H_0/\frac{\sigma_1^2}{\sigma_2^2}=2)=P(0,254<\frac{S_2^2}{S_1^2}<2,64)=$$

$$=P(2\cdot 0,254<2\cdot\frac{S_2^2}{S_1^2}<2\cdot 2,64)=P(0,580<f_{4,9}<5,28)$$

Calculamos la cantidad de probabilidad interpolando a ojo:

Si $P(f_{4,9}>6{,}4)=0{,}01$ y $P(f_{4,9}>3{,}63)=0{,}05$

$$P(f_{4,\,9}<0,2538)=P(f_{9,4}>\frac{1}{0,2538}=3,94)=0,10\text{, se tiene}$$

$$P(0,2538<f_{4,\,9}<6,4)=0,89$$

Con lo que la probabilidad buscada podría ser $\cong 0{,}75$ y en ese caso la potencia sería: potencia = 1- error tipo II = 0,25. El test tiene una potencia muy baja.

59. De una población se ha extraído una muestra y se midió la variable X, obteniéndose la siguiente tabla de frecuencias:

valores de X	1	2	3	4	5
frecuencia	20	40	85	53	13

Estudiar si se puede aceptar que la distribución de la variable X es uniforme. ¿Qué hipótesis habrá que hacer para llevar a cabo el estudio?

Solución:

El contraste que debemos aplicar es:

$$\begin{cases} H_0 : p_i = p_j \ \forall i, j = 1, \dots, 5 \\ H_1 : \text{no todas las } p_1 \text{ son iguales} \end{cases} \qquad \text{siendo } p_i = P(X = \text{i}).$$

Si la hipótesis nula es cierta las frecuencias esperadas serán todas iguales

$$f_i = \frac{\text{n°casos}}{5} = \frac{211}{5} = 42,2$$

El estadístico del contraste es:

$$\sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{5} \frac{O_i^2}{E_i} - N = 289,17 - 211 = 78,17$$

que si la hipótesis nula es cierta debe comportarse como una ji-cuadrada con k-1 =4 grados de libertad.

La región crítica es:

$$RC_{0.05} = \left[\mathrm{X}^2_{4;\ 0.05},\ \ +\infty \right) = \left[9.488,\ \ +\infty \right)$$

CONCLUSIÓN: el estadístico cae en la región crítica; por tanto concluimos que hay evidencia estadística de que la variable X no sigue una distribución uniforme, cosa que era bastante evidente desde el principio sin más que observar las frecuencias muestrales.

60. Una prueba clínica se basa en la asignación al azar de 40 pacientes a grupos de tratamiento y control de forma que se obtienen 20 individuos en cada grupo. Entre los 20 pacientes control se obtienen 12 aciertos. ¿Cuántos aciertos deben obtenerse entre los 20 pacientes sometidos a tratamiento para poder concluir, con un nivel del 5%, que el índice de aciertos es significativamente mayor en el grupo de tratamiento que en el de control?

Solución:

Vamos a resolverlo por tanteo. Puesto que los pacientes del grupo de control con los que se obtiene acierto son 12, para que haya diferencia significativa los pacientes con acierto entre los sometidos a tratamiento debe ser cercano a 20.

Sea X = nº de pacientes del grupo de control con acierto.

Y = nº de pacientes del grupo de tratamiento con acierto.

El contraste que debemos plantear es:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases} \quad \text{o bien} \quad \begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 < 0 \end{cases}$$

$$\begin{cases} \hat{p}_1 = \dfrac{1}{n} X \approx N(p_1, \sqrt{\dfrac{p_1(1-p_1)}{n}}) \\ \hat{p}_2 = \dfrac{1}{n} Y \approx N(p_2, \sqrt{\dfrac{p_2(1-p_2)}{n}}) \end{cases} (*) \Rightarrow$$

$$\Rightarrow \hat{p}_1 - \hat{p}_2 \in N(p_1 - p_2 \ , \sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{n}} \ )$$

(*) La convergencia de la binomial a la normal se considera buena a partir de n = 30 y en nuestro caso n = 20, pero como no tenemos otra herramienta para intentar responder usaremos esta aproximación porque el estadístico del contraste es $\hat{p}_1 - \hat{p}_2$ y no sabemos cómo se comporta a menos que consideremos que X e Y se parecen suficientemente a una normal, con lo que serán "casi normales" $\hat{p}_1$ y $\hat{p}_2$, y en consecuencia también nuestro estadístico de interés, el ya mencionado $\hat{p}_1 - \hat{p}_2$.

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \in N(0,1)$$

Y por tanto:

$$RC_{0,05} = \left(-\infty\ ,\ -Z_{0,05}\right] = \left(-\infty\ , -1,645\right]$$

Para aceptar la hipótesis alternativa el valor muestral del estadístico del contraste debe ser menor que -1,645.

$$n_1 = n_2 = 20 \qquad \hat{p}_1 = \frac{12}{20} \qquad \hat{p}_2 = \frac{k}{20}$$

$$k = 18 \Rightarrow \frac{\frac{12}{20} - \frac{18}{20}}{\sqrt{\frac{\frac{12}{20}\frac{8}{20}}{20} + \frac{\frac{18}{20}\frac{2}{20}}{20}}} = \frac{\frac{-6}{20}}{\sqrt{\frac{12\cdot 8+18\cdot 2}{20^3}}} = \frac{-6}{\sqrt{\frac{132}{20}}} = -2,3355 \in \text{RC}$$

$$k = 17 \Rightarrow \frac{\frac{12}{20} - \frac{17}{20}}{\sqrt{\frac{\frac{12}{20}\frac{8}{20}}{20} + \frac{\frac{17}{20}\frac{3}{20}}{20}}} = \frac{\frac{-5}{20}}{\sqrt{\frac{12\cdot 8+17\cdot 3}{20^3}}} = \frac{-5}{\sqrt{\frac{147}{20}}} = -0,41 \notin \text{RC}$$

Con 18 aciertos entre los pacientes tratados habría evidencia de que la proporción de aciertos entre los sometidos sólo a control es menor que la proporción de aciertos entre los pacientes con tratamiento, a nivel de significación 0,05.

61. Se ha observado el número de hijos varones en 1000 familias que tuvieran 5 hijos en total, obteniéndose los siguientes resultados:

nº de hijos varones	0	1	2	3	4	5
nº de familias	31	168	319	308	150	24

Estudiar la bondad de ajuste a una distribución binomial, a un nivel de significación de 0.05.

Solución:

Sea X = nº de hijos varones de una familia con 5 hijos,

$$\begin{cases} H_0 : X \in B(5\ ,\ 0{,}5) \\ H_1 : X \text{ no sigue ese modelo} \end{cases}$$

el estadístico del contraste es:

$$\sum_i \frac{(O_i - E_i)^2}{E_i} = \sum_i \frac{O_i^2}{E_i} - N$$

donde,

$$E_i = n.p_i \text{ y } p_i = P(X = i) = \binom{5}{i} \cdot (0{,}5)^5$$

supuesto que la hipótesis nula es cierta

X	O_i	p_i	$E_i = n.p_i$
0	31	0,03125	31,25
1	168	0,15625	156,25
2	319	0,3125	312,5
3	308	0,3125	312,5
4	150	0,15625	156,25
5	24	0,03125	31,25

El estadístico toma el valor 3,0176.

$$RC_{0,05} = \left[\mathrm{X}^2_{5,\,0,05}\ ,\ +\infty\right) = \left[9{,}49\ ,\ +\infty\right)$$

CONCLUSIÓN: el estadístico no cae en la región crítica. Aceptamos la hipótesis nula. El nº de varones de una familia de 5 hijos se distribuye según una Binomial y la probabilidad de varón es ½.

62. En un estudio acerca de la precipitación pluvial y la cantidad de contaminación de aire eliminada, se obtuvieron los siguientes datos:

Calcular:

a) Recta de regresión de Y sobre X y el coeficiente de correlación.

b) ¿Es válido el modelo lineal? Justificar la respuesta.

c) Intervalo de predicción, al 95% de las partículas eliminadas cuando la cantidad de lluvia diaria asciende a 4.8. cm^3/cm^2.

Lluvia diaria, X (cm^3/cm^2)	4,3	4,5	5,9	5,6	6,1	5,2	3,8	2,1	7,5
Partículas eliminadas, Y (mg/m)	126	121	116	118	114	118	132	141	108

Solución:

$$\sum x_i = 45 \quad ; \quad \sum x_i^2 = 244.26 \quad \Rightarrow S_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 19,26$$

$$\sum y_i = 1094 \quad ; \quad \sum y_i^2 = 133786 \quad \Rightarrow S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 804,22$$

$$\sum x_i y_i = 5348,2 \quad \Rightarrow S_{xy} = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} = -121,8$$

a) Para calcular la recta de regresión $y = a + b\,x$, tenemos:

$$\left.\begin{array}{l} b = \dfrac{S_{xy}}{S_{xx}} = -6,324 \\ a = \dfrac{\sum y_i - b\sum x_i}{n} = 153,175 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \hat{\mu}_{Y/x} = 153,175 - 6,324x \\ r = b\sqrt{\dfrac{S_{xx}}{S_{yy}}} = -0,9962 \end{array}\right.$$

b) Dado que el coeficiente de correlación r, es tan próximo a -1 se ve que hay casi correlación negativa perfecta. De todas formas el coeficiente de determinación es $r^2 = 0,9924$, lo que nos permite decir que más del 99% de la varibilidad de la concentración de la polución es debida a la dependencia lineal de ésta con respecto a la cantidad de lluvia. El modelo lineal es adecuado.

c) Queremos un intervalo de predicción para el valor de la *variable* y, cuando la lluvia caída sea de 4,8 cm^3/cm^2; éste viene dado por:

$$I_{y_0;\alpha} = \left[y_0 \pm t_{n-2;\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})}{S_{xx}}} \right]$$

$$\begin{cases} x_0 = 4{,}8 \\ \hat{y}_0 = 153{,}175 - 6{,}324 \cdot x_0 = 123{,}395 \\ s = \sqrt{\dfrac{S_{yy} - bS_{xy}}{n-2}} = 2{,}202 \\ t_{7;0{,}025} = 2{,}365 \\ \overline{x} = \dfrac{\sum x_i}{n} = \dfrac{45}{9} = 5 \end{cases}$$

$$\Rightarrow I_{y_0,0{,}05} = [123{,}395 \pm 4{,}5485] = [118{,}846\ ,\ 127{,}943]$$

63. Se ensaya un tipo de fertilizante sobre la producción de trigo de dos maneras: con riego y sin riego. Para ello se eligen 20 parcelas y se agrupan en diez pares, siendo cada par de la misma extensión, pero pares distintos no tienen por qué tener la misma extensión. Las producciones obtenidas (en Qm) después de la cosecha fueron:

Parcelas	1	2	3	4	5	6	7	8	9	10
con riego	46	110	70	54	60	120	82	76	37	28
sin riego	42	87	75	50	48	108	80	67	40	25

Con nivel de significación α=0,06, ¿puede aceptarse que el fertilizante aplicado con riego, supera al aplicado sin riego?

Solución:

Las variables en estudio son:

X = producción de trigo (Qm) en una parcela con riego.

Y = producción de trigo (Qm) en una parcela sin riego.

Tenemos las parcelas emparejadas por el tamaño, considerando que no hay diferencias por el tipo de suelo. Son dos v. a. continuas con datos emparejados.

No nos dice nada de la normalidad de las distribuciones, por tanto en principio debemos aplicar un test no paramétrico, que será el test de los signos aplicado sobre las diferencias $d_i = x_i - y_i$.

Parcela	1	2	3	4	5	6	7	8	9	10
d_i	4	23	-5	4	12	12	2	9	-3	3

Aplicaremos un test unilateral derecho porque por los datos parece que las parcelas con riego producen más y queremos ver si hay evidencia estadística.

$$\begin{cases} H_0 : M_{X-Y} \leq 0 \\ H_1 : M_{X-Y} > 0 \end{cases}$$

Tomamos como estadístico del contraste

$$N^- = \text{nº de diferencias negativas} = 2$$

y si la hipótesis nula es cierta sería $N^- \in B(10, 0{,}5)$, pero entonces:

$$P(N^- \leq 2) = 0{,}0547 < 0{,}06$$

CONCLUSIÓN: ha ocurrido algo que es menos probable que nuestro α. Hay evidencia estadística de que la mediana de la variable diferencia es mayor que 0 y por tanto las parcelas con riego producen más que las sin riego.

64. En lanzamientos sucesivos de un dado, se han dado los siguientes resultados:

puntuación	1	2	3	4	5	6
frecuencia	95	92	80	123	140	70

¿Se puede decir que el dado es regular? (nivel de confianza 95%).

Solución:

Comprobar si el dado es regular es lo mismo que comprobar que la variable X = nº de puntos que salen al tirar el dado, es una variable discreta uniforme con $P(X=k) = 1/6$ para $k=1,...,6$. Por tanto se trata en realidad de un problema de bondad de ajuste con 600 observaciones, donde cada una de las frecuencias esperadas es 100.

Formalmente sería:

$$\begin{cases} H_0 : X \in U(1/6) \\ H_1 : X \text{ no sigue esa distribución} \end{cases}$$

$$E_i = n \cdot p_i = 600 \cdot \frac{1}{6} = 100, \forall i = 1,...,6$$

El estadístico del contraste es:

$$\sum \frac{(O_i - E_i)^2}{E_i} = \sum \frac{O_i^2}{E_i} - N = 635,18 - 600 = 35,18$$

que si la hipótesis nula es cierta será una ji-cuadrado con n-1(5,en este caso) grados de libertad.

$$RC_{0.05} = \left[X^2_{5;\,0.05},\ +\infty\right) = \left[11.070,\ +\infty\right)$$

CONCLUSIÓN: el valor del estadístico muestral cae en la región crítica, por tanto rechazamos la hipótesis nula. Hay evidencia estadística, a nivel de significación 0,05 de que el dado no es regular.

65. Se realizó un estudio para saber si el incremento en la concentración de sustrato tiene un efecto apreciable en la velocidad de una reacción química. La reacción se realizó 15 veces con una concentración de sustrato de 1.5 moles por litro, con una velocidad promedio de 7.5 micromoles por 30 minutos y una desviación estándar de 1.5. Con una concentración de sustrato de 2.0 moles por litro, se corrieron 12 pruebas, resultando una velocidad promedio de 8.8 micromoles por 30 minutos con una desviación estándar de 1.2. ¿Hay suficiente razón para creer que este incremento en la concentración de sustrato ocasiona un aumento en la velocidad promedio por más de 0.5 micromoles por 30 minutos? Utilice un nivel de significación de 0.02 y suponga que las poblaciones tienen una distribución aproximadamente normal.

Solución:

Sea X = velocidad de precipitación (μmoles/30min.) cuando la concentración inicial es 1,5 moles/l.

Sea Y = velocidad de precipitación (μmoles/30min.) cuando la concentración inicial es 2,0 moles/l.

$X \in N(\mu_1, \sigma_1)$	$Y \in N(\mu_2, \sigma_2)$
$n_1 = 15$	$n_2 = 12$
$\bar{X} = 7,5$	$\bar{Y} = 8,8$
$S_1 = 1,5$	$S_2 = 1,2$

$$\bar{Y} - \bar{X} \in N(\mu_2 - \mu_1 \ , \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

Planteamos el contraste

$$\begin{cases} H_0 : \mu_2 - \mu_1 \leq 0.5 \\ H_1 : \mu_2 - \mu_1 > 0.5 \end{cases}$$

Si H_0 es cierta, el estadístico

$$\frac{\left(\bar{Y} - \bar{X}\right) - 0.5}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = 2,96$$

sigue una t de Student con f grados de libertad, donde f viene dado por:

$$f = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}} = 24,99 \ \ (*)$$

$$\Rightarrow RC_{0,02} = \left[t_{25;0,02} \ , \ +\infty\right) = \left[2,240 \ , \ +\infty\right)$$

CONCLUSIÓN: el valor del estadístico cae en la región crítica; hay evidencia estadística de que la mayor concentración de sustrato aumenta la velocidad de precipitación al menos en 0,5 μmoles/30 min. (a nivel incluso 0,01el valor del estadístico sigue cayendo en región crítica).

(*) hemos calculado el punto crítico por interpolación, ya que $t_{0.02}$ no viene en la tabla.

66. En un experimento con guisantes, Gregor Mendel observó que 315 eran redondos y amarillos, 108 redondos y verdes, 101 rugosos y amarillos y 32 rugosos y verdes. De acuerdo con su teoría de la herencia, esos números debían estar en la proporción 9:3:3:1. ¿Hay alguna evidencia para dudar de su teoría al nivel de significación 0.05?

Solución:

Estadísticamente tenemos la variable tetradimensional $X = (X_1, X_2, X_3, X_4)$ siendo:

X_1= nº de guisantes redondos y amarillos de los 556 observados.

X_2= nº de guisantes redondos y verdes de los 556 observados.

X_3= nº de guisantes rugosos y amarillos de los 556 observados.

X_4= nº de guisantes rugosos y verdes de los 556 observados.

Queremos comprobar si hay evidencia de que esta variable no sigue una distribución multinomial

$$M\left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$$

El contraste será:

$$\begin{cases} H_0 : X = (X_1, X_2, X_3, X_4) \in M(p_1, p_2, p_3, p_4) \\ \qquad \text{con } p_1 = \frac{9}{16},\ p_2 = \frac{3}{16},\ p_3 = \frac{3}{16},\ p_4 = \frac{1}{16} \\ H_1 : X \text{ no sigue esa distribución} \end{cases}$$

Aplicaremos un contraste de bondad de ajuste de la ji-cuadrado, para lo que construimos la siguiente tabla de frecuencias:

	RA	RV	RgA	RgV	Total
O_i	315	108	101	32	556
$E_i = np_i$	312,75	104,25	104,25	34,25	556

El estadístico del contraste es:

$$\sum \frac{(O_i - E_1)^2}{E_i} = \sum \frac{O_i^2}{E_i} - N = 556,47 - 556 = 0,47$$

que si la hipótesis nula es cierta debe comportarse como una ji-cuadrado con k-1, 4 en nuestro caso, g. de libertad con lo que la región crítica es:

$$RC_{0,05} = \left[\mathrm{X}^2_{4;\,0,05}\,,\, +\infty \right) = [9,488, +\infty)$$

CONCLUSIÓN: incluso sin calcular la región crítica es evidente que el ajuste al modelo propuesto es muy bueno ya que el valor del estadístico es muy pequeño. No hay ninguna evidencia para dudar de la teoría de Mendel.

67. Tenemos una muestra de tamaño ***n*** de una población con distribución de Poisson, cuya función de densidad viene dada (como saben) por :

$$f(x) = (e^{-\lambda} \cdot \lambda^{x}) \,/\, x! \qquad x = 0, 1, 2, ...$$

Calcular la expresión del estimador de máxima verosimilitud del parámetro $\boldsymbol{\lambda}$ de dicha distribución.

Solución:

Suponemos que los valores muestrales son $k_1, k_2, ..., k_n$. Como las realizaciones muestrales se suponen independientes:

$$P(X_1 = k_1, X_2 = k_2, ..., X_n = k_n) = \prod_i P(X_i = k_i) =$$

$$= \frac{\lambda^{k_1}}{k_1!} e^{-\lambda} \cdot ...... \cdot \frac{\lambda^{k_n}}{k_n!} e^{-\lambda} = \frac{\lambda^{\sum k_i} \cdot e^{-n\lambda}}{k_1! \cdot ..... \cdot k_n!}$$

Para obtener ahora la expresión del estimador máximo verosímil para λ, maximizamos la función de probabilidad conjunta de la muestra, basándonos en la filosofía de que si esta muestra es la que ha ocurrido debe ser precisamente porque es la más probable.

$$\frac{\partial P}{\partial \lambda}=\frac{\left(\sum k_i\right)\lambda^{\sum k_i-1}e^{-n\lambda}+\lambda^{\sum k_i}\left(-n\right)e^{-n\lambda}}{k_1!\cdot\ldots\ldots\cdot k_n!}=\frac{\lambda^{\sum k_i-1}e^{-n\lambda}(\sum k_i-n\lambda)}{k_1!\cdot\ldots\ldots\cdot k_n!}$$

$$\frac{\partial P}{\partial \lambda}=0 \Leftrightarrow \sum k_i-n\lambda=0 \Leftrightarrow \lambda=\frac{\sum k_i}{n}$$

Se comprueba fácilmente que este valor de λ corresponde precisamente a un máximo y observamos que no es otra cosa que la media muestral.

68. Existe una rara especie de planta, cuya característica principal es su altura. Se sabe que la altura de este tipo de planta se distribuye normalmente con una desviación típica de 0.2 m. Un jardinero encarga para un parque un gran pedido de plantas de esta especie de 5 m de alto. Después de recibir el pedido el jardinero selecciona 16 plantas al azar y mide su altura: si la media muestral es menor que la esperada, se devuelve el pedido al vivero hasta que crezcan más.

a) ¿Cuál debe ser la altura promedio de las plantas seleccionadas para que se devuelva el pedido al vivero, si sabemos que la probabilidad de rechazar el pedido es de 0.04?

b) ¿Cuál es la potencia de la prueba descrita en el apartado anterior, si la altura promedio real de las plantas es de 4.9 m?

Solución:

Sea X = altura de una planta, $X \in N(\mu, 0{,}2)$.

a)

$$\begin{cases} H_0 : \mu \geq 5 \\ H_1 : \mu < 5 \end{cases}$$ Si la hipótesis nula es cierta se tiene:

$$\bar{X} \in N(5\ ,\ \frac{0{,}02}{\sqrt{16}}) = N(5, 0.05)$$

Estamos buscando un valor A tal que:

$$P(\overline{X} < A) = 0,04$$

$$0,04 = P(\overline{X} < A) = P\left(\frac{\overline{X}-5}{0,05} < \frac{A-5}{0,05}\right) =$$

$$= P\left(Z < \frac{A-5}{0,05}\right) \Rightarrow \frac{A-5}{0,05} = -1,75 \Rightarrow A = 4,9125$$

Si la media muestral de las alturas de las 16 plantas que comprueba el jardinero es menor que 4,9125 m rechazará el pedido.

b) $\beta = P(A\text{ceptar el lote}/\mu{=}4{,}9) = P(\overline{X} > 4,9125) =$

$$= P\left(\frac{\overline{X}-4,9}{0,05} > \frac{4,9125-4,9}{0,05}\right) = P(Z > 0,25) = 0,5987 \quad \Rightarrow$$

$$\Rightarrow \quad potencia = 1-\beta = 0,4013$$

69. Si el coeficiente de correlación entre X e Y es 0.50. ¿Qué porcentaje de la variación total de Y queda inexplicado por la ecuación de regresión?

Solución:

El coeficiente de determinación, que no es otro que $r^2 = 0.25$, es el que nos permite responder a esta pregunta. Tan solo aproximadamente el 25% de la variación total de la variable Y, puede ser explicado por la dependencia lineal de Y respecto a X. El 75% de la variabilidad de Y queda inexplicado con la recta de regresión.

70. Durante un curso, un estudiante obtuvo las calificaciones que figuran en la tabla en algunos de sus exámenes. Determinar si hay diferencia significativa entre las calificaciones de las asignaturas que cursaba, a un nivel de significación de 0.05.

Matemáticas	72 80 83 75
Ciencias	81 74 77
Inglés	88 82 90 87 80
Economía	74 71 77 70

Solución:

Está claro que tenemos cuatro variables; no nos dicen nada de su normalidad, lo que nos lleva a tratar el problema con un test no paramétrico; no son datos emparejados porque no se dice nada que nos permita suponer la existencia de una variable de sesgo (son notas todas de un mismo alumno...): Usaremos el test de Kruskal-Wallis.

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1: & \text{no todas las medias son iguales} \end{cases}$$

Asignamos rango a cada uno de los valores de las 4 muestras.

Matemáticas	72(3) 80(10,5) 73(4) 75(7)	$R_1 = 24,5$
Ciencias	81(12) 74(5,5) 77(8,5)	$R_2 = 26$
Inglés	88(15) 82(13) 90(16) 87(14) 80(10,5)	$R_3 = 68,5$
Economía	74(5,5) 71(2) 77(8,5) 70(1)	$R_4 = 17$

El estadístico del contraste será:

$$H = \frac{12}{N(N+1)}\left[\sum_{i=1}^{4}\frac{R_i^2}{n_i}\right] - 3(N+1) = 54,70 - 51 = 3,70$$

Y si la hipótesis nula es cierta debería comportarse como una ji-cuadrado con 3 grados de libertad, por tanto la región crítica es:

$$RC_{0,05} = \left[\mathrm{X}^2_{3;\,0,05}\ ,\ +\infty\right) = \left[7,82\ ,\ +\infty\right)$$

CONCLUSIÓN: puesto que el valor muestral del estadístico no cae en región crítica nos quedamos con la hipótesis nula: Con la información de que disponemos y a un nivel de significación de 0,05 no tenemos evidencia de que las notas medias en las cuatro asignaturas sean diferentes.

71. Las calificaciones de un grupo de estudiantes en su reporte de medio año (X) y en los exámenes finales (Y) fueron las siguientes:

X	77	50	71	72	81	94	96	99	67
Y	82	66	78	34	47	85	99	99	68

a) Estime la recta de regresión lineal de Y sobre X.

b) Intervalo de confianza al 95% para la calificación del examen final de un estudiante que obtuvo una calificación de 85 en el reporte de medio año.

Solución:

$$S_{xx} = \frac{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}{n} = 2018,222$$

$$S_{yy} = \frac{n\left(\sum y_i^2\right) - \left(\sum y_i\right)^2}{n} = 3872,888$$

$$S_{xy} = \frac{n\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n} = 1568,4$$

a)

$$\left.\begin{array}{l} b = \dfrac{S_{xy}}{S_{xx}} = 0,777 \\ a = \dfrac{\sum y_i - b\sum x_i}{n} = 12,073 \end{array}\right\} \text{recta regresión } \hat{\mu}_{Y/x} = 12,073 + 0,777x$$

b) El intervalo de predicción para el valor de la variable dependiente, a nivel de significación 0,05 viene dado por:

$$I_{y_0, 0,05} = \left[\hat{y}_0 \pm t_{n-2, \alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

donde:

$$x_0 = 85$$

$$\hat{y}_0 = a + bx_0 = 78,118$$

$$s=\sqrt{\frac{S_{yy}-bS_{xy}}{n-2}}=19,472$$

$$\overline{x}=\frac{\sum x_i}{n}=78,555$$

$$t_{n-2,\alpha/2}=t_{7;\,0,025}=2,365$$

$$n=9$$

$$\Rightarrow I_{y_0;0,05}=[78,118\pm 52,1321]=[25,986\ ,\ 130,250]$$

72. La altura promedio de las mujeres en el grupo de primer año de una institución de enseñanza superior, se considera que se distribuye normalmente con una media de 162.5 cm y con una desviación típica de 6.9 cm. ¿Hay alguna razón, al nivel de significación de 0.05, para creer que existe un cambio en la altura promedio, si una muestra aleatoria de 50 mujeres del grupo actual tiene una altura promedio de 165.2 cm?

Solución:

Sea X = altura de una mujer de ese grupo, $X\in N(162.5\ ,\ 6.9)$

$$\begin{cases} H_0: \ \mu \le 162.5 \\ H_1: \ \mu > 162.5 \end{cases}$$

Si H_0 es cierta, entonces

$$\overline{X}\in N(162.5\ ,\frac{6.9}{\sqrt{50}})=N(162.5\ ,0.9758)$$

entonces

$$\frac{\overline{X}-162.5}{0.9758}=\frac{165.2-162.5}{0.9758}=2,7769$$

debe ser un valor de una normal estándar

$$RC_{0,05}=\left[Z_{0,05}\ ,\ +\infty\right)=\left[1,645\ ,\ +\infty\right)$$

CONCLUSIÓN: el estadístico del contraste cae en la región crítica: hay evidencia estadística, a nivel de significación 0,05, para afirmar que la altura promedio de las mujeres de ese grupo de población ha aumentado.

73. Supongamos dos muestras aleatorias de tamaños $n_1 = 10$ y $n_2 = 5$ de dos poblaciones normales de igual varianza de valor 9 y medias μ_1 y μ_2. Si la diferencia entre las medias poblacionales fuera a) 0,1, b) 3 y c) 128, y el nivel de significación $\alpha_1 = 0{,}05$ y $\alpha_2 = 0{,}01$. Calcular la potencia del siguiente contraste:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Solución:

El contraste que nos plantea el enunciado es equivalente a:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases} \quad \text{donde:}$$

$$\overline{X} - \overline{Y} \in N(\mu_1 - \mu_2\ ,\ \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}) = N(\mu_1 - \mu_2\ ,\ 3\sqrt{0,3})$$

Las regiones críticas para esos niveles de significación son:

a) si $\mu_1 - \mu_2 = 0{,}1$

$$\beta = P(\text{error tipo II}) = P(\text{aceptar } \mathrm{H}_0 / \mu_1 - \mu_2 = 0{,}1) =$$
$$= P\left(-1{,}96 \leq \frac{\overline{X} - \overline{Y}}{3\sqrt{0{,}3}} \leq 1{,}96 \right) = P\left(-1{,}96 \cdot 3\sqrt{0{,}3} \leq \overline{X} - \overline{Y} \leq 1{,}96 \cdot 3\sqrt{0{,}3}\right)$$
$$= P\left(-1{,}96 \cdot 3\sqrt{0{,}3} - 0{,}1 \leq \overline{X} - \overline{Y} - 0{,}1 \leq 1{,}96 \cdot 3\sqrt{0{,}3} - 0{,}1\right) =$$
$$= P\left(\frac{-1{,}96 \cdot 3\sqrt{0{,}3} - 0{,}1}{3\sqrt{0{,}3}} \leq \frac{\overline{X} - \overline{Y} - 0{,}1}{3\sqrt{0{,}3}} \leq \frac{1{,}96 \cdot 3\sqrt{0{,}3} - 0{,}1}{3\sqrt{0{,}3}} \right) =$$
$$= P\left(-1{,}96 - \frac{0{,}1}{3\sqrt{0{,}3}} \leq Z \leq 1{,}96 - \frac{0{,}1}{3\sqrt{0{,}3}} \right) = P(-2{,}12 \leq Z \leq 1{,}9) =$$
$$= 0{,}9713 - 0{,}0170 = 0{,}9543 \Rightarrow \textit{potencia} = 1 - \beta = 0{,}0457$$

b) si μ_1 - μ_2 = 3 tendremos:

$$\beta = P(\text{error tipo II}) = P(\text{aceptar H}_0 / \mu_1 - \mu_2 = 3) =$$
$$= P\left(-1{,}96 \le \frac{\overline{X}-\overline{Y}}{3\sqrt{0{,}3}} \le 1{,}96\right) = P\left(-1{,}96 - \frac{3}{3\sqrt{0{,}3}} \le Z \le 1{,}96 - \frac{3}{3\sqrt{0{,}3}}\right) =$$
$$= P(-1{,}96 - 1{,}83 \le Z \le 1{,}96 - 1{,}83) = P(-3{,}79 \le Z \le 0{,}13) =$$
$$= 0{,}5517 - 0 = 0{,}5517 \Rightarrow \textit{potencia} = 1 - \beta = 0{,}4483$$

c) si μ_1 - μ_2 = 128 tendremos:

$$\beta = P(\text{error tipo II}) = P(\text{aceptar H}_0 / \mu_1 - \mu_2 = 128) =$$
$$= P\left(-1{,}96 \le \frac{\overline{X}-\overline{Y}}{3\sqrt{0{,}3}} \le 1{,}96\right) = P\left(-1{,}96 - \frac{128}{3\sqrt{0{,}3}} \le Z \le 1{,}96 - \frac{128}{3\sqrt{0{,}3}}\right) =$$
$$= P(-1{,}96 - 77{,}9 \le Z \le 1{,}96 - 77{,}9) = P(-79{,}86 \le Z \le -75{,}94) = 0$$
$$\Rightarrow \text{potencia} = 1 - \beta = 1$$

Ahora debemos repetir los tres apartados considerando la otra región de aceptación, es decir, (-2,57 , 2,57) .

Se hará exactamente igual, basta poner -2,57 y 2,57 donde antes ponía -1,96 y 1,96, respectivamente. Lo haremos sólo para μ_1 - μ_2= 3.

b) si μ_1 - μ_2 = 3 tendremos:

$$\beta = P(\text{error tipo II}) = P(\text{aceptar H}_0 / \mu_1 - \mu_2 = 3) =$$
$$= P\left(-2{,}57 \le \frac{\overline{X}-\overline{Y}}{3\sqrt{0{,}3}} \le 2{,}57\right) = P\left(-2{,}57 - \frac{3}{3\sqrt{0{,}3}} \le Z \le 2{,}57 - \frac{3}{3\sqrt{0{,}3}}\right) =$$
$$= P(-2{,}57 - 1{,}83 \le Z \le 2{,}57 - 1{,}83) = P(-4{,}40 \le Z \le 0{,}74) = 0{,}7704 - 0$$
$$\Rightarrow \textit{potencia} = 1 - \beta = 0{,}3296$$

Observación: como debe ocurrir, un contraste que está diseñado para detectar si la diferencia de las dos medias es distinta de cero, prácticamente no va a detectarlo nunca si la diferencia es tan pequeña como 0,1, un porcentaje de veces más alto lo detectará si la diferencia es de 3 unidades y prácticamente siempre cuando la diferencia de medias que, no lo olvidemos, suponíamos que era cero, es tan grande como 128 unidades. Podemos también destacar que al aumentar el nivel de significación, con lo que disminuimos la probabilidad de error tipo 1, disminuye la potencia, cosa que también sabíamos.

74. Se determina la pérdida de actividad de un preparado hormonal en el curso del tiempo y se obtiene el siguiente resultado.

Y = tiempo (meses)	1	2	3	4	5
X = % de actividad restante	90	75	42	30	21

Se desea calcular:

a) La expresión de la relación lineal entre ambas variables.

b) ¿Cuándo será del 25% la actividad del preparado hormonal?

c) A un nivel de significación de 0,05, ¿es aceptable el modelo lineal considerado?

d) Calcular el intervalo de confianza al 90% para el tiempo promedio transcurrido para que un preparado conserve el 85% de su actividad.

Solución:

$$\begin{cases} \sum x_i = 258 & \sum x_i^2 = 16.830 & (\sum x_i)^2 = 66564 \\ \sum y_i = 15 & \sum y_i^2 = 55 & (\sum y_i)^2 = 225 \end{cases} \qquad \sum x_i y_i = 591$$

$$S_{xx} = \frac{n(\sum x_i^2) - (\sum x_i)^2}{n} = 3517,2$$

$$S_{yy} = \frac{n(\sum y_i^2) - (\sum y_i)^2}{n} = 10$$

$$S_{xy} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n} = -183$$

a) La recta de regresión viene dada por

$$\hat{\mu}_{Y/x} = a + bx$$

con:

$$b = \frac{S_{xy}}{S_{xx}} = -0,05$$

$$a = \frac{\sum y_i - b\sum x_i}{n} = 5,68$$ de donde $\hat{\mu}_{Y/x} = 5.58 - 0.05x$

b) Y = 5,58 – 0,05. 25 = 4,43 meses

c) Para estudiar si el modelo lineal es aceptable planteamos el contraste:

$$\begin{cases} H_0: \ \rho = 0 \\ H_1: \ \rho \neq 0 \end{cases}$$

El estadístico del contraste es

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -7.72$$

que, si la hipótesis nula es cierta, debe comportarse como una t_3, y por tanto la región crítica será:

$$RC_{0.05} = \left(-\infty, \ -t_{3;\,0.05}\right] \cup \left[t_{3;\,0.05}, \ +\infty\right) = \left(-\infty, \ -3.182\right] \cup \left[3.182, \ +\infty\right)$$

(Hemos tomado $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = -0,9758$ y $n = 5$)

CONCLUSIÓN: el estadístico cae en la región crítica, luego hay evidencia estadística de que el coeficiente de correlación es distinto de cero y, por tanto, el modelo lineal es bueno

d) El intervalo de confianza del tiempo medio transcurrido cuando el preparado tiene el 85% de actividad con una confianza del 90% viene dado por:

$$I_{\mu_0;\,0,1} = \left(\hat{y}_0 \pm t_{n-2;\,\alpha/2}\, s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

$$x_0 = 85$$

$$\hat{y}_0 = 5,58 - 0,05 \cdot 85 = 1,33$$

$$t_{n-2;\ \alpha/2} = t_{3;\ 0,05} = 2,353$$

$$s = \sqrt{\frac{S_{yy} - bS_{xy}}{n-2}} = 0,5323$$

$$\Rightarrow \ I_{\mu_{85};\ 0,10} = \left(1,33 \pm 0,8054\right) = \left(0,5246\ ,\ \ 2,1354\right)$$

75. Los educadores creen que la habilidad de los alumnos de 2º curso puede ser mejorada por medio de un nuevo método con ayuda de computadoras, en lugar del método convencional. Los educadores saben que la medida de habilidad de los estudiantes sigue una distribución aproximadamente normal y que el promedio de habilidad para los alumnos que han aprendido con el método convencional se ha establecido en 78, siendo la desviación típica para ambos métodos de 7. Para probar esta hipótesis se calcula el promedio de habilidad de una muestra de 200 estudiantes. Si el promedio de la habilidad de esta muestra resulta ser mayor o igual que 79 se aceptará que el método nuevo mejora la habilidad, de otro modo se sigue aceptando como válido el método convencional.

1) Calcular el error tipo I.
2) Calcular el error tipo II, suponiendo que el verdadero promedio de habilidad con el nuevo método sea 78,5.

Solución:

Sea X = puntuación de habilidad con el nuevo método, $X \in N(\mu, 7)$.

El contraste de hipótesis que plantea si la media de habilidad de los alumnos mejora con el nuevo método es:

$$\begin{cases} H_0 : \mu \leq 78 \\ H_1 : \mu > 78 \end{cases}$$

Puesto que los educadores admiten que el nuevo método es mejor si la media muestral de la puntuación de los 200 alumnos es 79 o superior, esto significa que están tomando como región crítica $RC = [79, +\infty[$.

El estadístico del contraste será la media muestral y, si la hipótesis nula es cierta debe ser

$$\bar{X} \in N(78, \frac{7}{\sqrt{200}})$$

Como la variable en realidad es discreta, para calcular probabilidades haremos un ajuste de continuidad

a) $\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) =$

$= P(\bar{X} \geq 78.5 / \mu = 78) = \; = P\left(\frac{\bar{X} - 78}{7/\sqrt{200}} > \frac{78.5 - 78}{7/\sqrt{200}} \right) =$

$= P(Z > 1.01) = 0{,}1562$

b) $\beta = P(\text{error tipo II}) = P(\text{aceptar } H_0 / H_0 \text{ falsa}) =$

$= P(\bar{X} < 78.5 / \mu = 78{,}5) = 0.5$

Observación: vemos que la regla de decisión no es buena. Las probabilidades de cometer los dos tipos de errores son demasiado altas.

76. El rendimiento de la cosecha de un cereal se considera muy bueno si la producción es superior a 25 kg por área de cultivo, buena si es superior a 15 kg. y mala si no llega a 15. Se hacen 30 determinaciones del rendimiento en otras tantas parcelas donde se ha sembrado cereal de tipo A y 30 determinaciones en parcelas donde se sembró un cereal de tipo B. Los resultados fueron los siguientes:

	Tipo de cereal	
Rendimiento	A	B
Muy Bueno	10	12
Bueno	14	10
Malo	6	8

¿Son igualmente efectivos para el cultivo los dos tipos de cereales A y B? Tómese α = 0,05.

Solución:

Lo primero que debemos observar es que la información que tenemos es una serie de *frecuencias* de ocurrencias de cierto sucesos. Se trata por tanto de estudiar si dos caracteres, el tipo de cereal y la cantidad de producción son independientes. Como los totales parciales están prefijados (hay 30 parcelas sembradas con cada tipo de cereal) se trata de un test de homogeneidad.

$$\begin{cases} H_0 : P(C_i / \text{cereal} \quad A) = P(C_i / \text{cereal} \quad B) \quad \forall i = 1, \ 2, \ 3 \\ H_1 : \text{ no ocurre } H_0 \end{cases}$$

donde hemos llamado C_1 = cosecha muy buena
C_2 = cosecha buena
C_3 = cosecha mala

Hacemos la tabla de frecuencias esperadas donde en la casilla i,j ponemos:

$$E_{ij} = \frac{\text{total} \quad \text{fila} \quad i \cdot \text{total} \quad \text{columna} \quad j}{\text{gran} \quad \text{total}}$$

Tabla de frecuencias esperadas			
	Tipo de cereal		
Rendimiento	A	B	totales
Muy bueno C_1	11	11	22
Bueno C_2	12	12	24
Malo C_3	7	7	14
totales	30	30	60

El estadístico del contraste es:

$$\sum_{i,\,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \sum_{i,\,j} \frac{O_{ij}^2}{E_{ij}} - N = 61{,}134 - 60 = 1{,}134$$

Si la hipótesis nula es cierta esto debe ser un valor de una Chi-cuadrado con (r-1)(k-1) = 2 grados de libertad.

Tenemos:

$$RC_{0,05} = \left[\mathrm{X}^2_{2;\,0,05},\ +\infty\right) = \left[5,99,\ +\infty\right)$$

CONCLUSIÓN: el estadístico no toma valor en la región crítica. Por tanto con la información de que disponemos no hay motivo para afirmar que los dos tipos de cereal dan rendimientos diferentes. Podemos considerarlos iguales en cuanto a su rendimiento.

77. Un fabricante de unidades de pantallas de vídeo prueba dos diseños de microcircuitos para determinar, con un nivel de significación de 0,05, si producen flujos de corriente equivalentes. La ingeniería de desarrollo ha obtenido los siguientes datos:

Diseño 1	$n_1 = 15$	$\bar{X}_1 = 24,2$	$S_1^2 = 10$
Diseño 2	$n_2 = 10$	$\bar{X}_2 = 23,9$	$S_2^2 = 20$

Suponiendo que ambas poblaciones son normales, aunque sin estar dispuestos a considerar que las varianzas sean iguales, ¿podemos afirmar que tiene razón el fabricante?

Solución:

Sea X = flujo de un circuito tipo 1, $X \in N(\mu_1, \sigma_1)$.

Y = flujo de un circuito tipo 2, $Y \in N(\mu_2, \sigma_2)$.

Se trata de estudiar si las diferencias entre los flujos medios de corriente de los dos tipos de microcircuitos son significativas. El contraste será:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad \text{o lo que es equivalente} \quad \begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

El estadístico del contraste es:

$$\bar{X}-\bar{Y}\in N(\mu_1-\mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})\Rightarrow\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\in t_f$$

$$\text{donde } f=\frac{\left(S_1^2/n_1+S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1}+\frac{\left(S_2^2/n_2\right)^2}{n_2-1}}=\frac{7,1111}{0,4762}=14,931\cong 15$$

Después de hacer operaciones el estadístico, para esta muestra concreta, toma el valor 0,1837. La región crítica es:

$$RC_{0.05}=\left(-\infty,\ t_{15;\ 0.025}\right]\cup\left[t_{15;\ 0.975},\ +\infty\right)=\left(-\infty,\ -2.131\right]\cup\left[2.131,\ +\infty\right)$$

CONCLUSIÓN: el estadístico no toma valor en la región crítica, luego no hay evidencia estadística de que haya diferencia en el flujo medio de los dos microcircuitos.

78. Un científico de computadoras ha desarrollado un algoritmo para generar enteros aleatorios en el intervalo 0 – 9. Codifica el algoritmo y genera 1000 dígitos aleatorios. Los datos obtenidos fueron:

0	1	2	3	4	5	6	7	8	9
94	93	112	101	104	95	100	99	108	94

¿Funciona bien el programa del científico? (α = 0,05)

Solución:

El programa funcionará bien si los números por él generados presentan unas frecuencias de aparición que estadísticamente puedan ser consideradas aleatorias. Se trata por tanto de un ajuste a una uniforme discreta.

$$\begin{cases} H_0: p_i=0,1 \quad \forall i=0,\ 1,...,9 \\ H_1: \text{no ocurre } H_0 \end{cases} \text{donde } p_i=P(\text{salga el número i})$$

Cada una de las frecuencias esperadas es 100 y el estadístico del contraste

$$\sum_i \frac{(O_i - E_i)^2}{E_i} = \sum_i \frac{O_i^2}{E_i} - N = 1.003,72 - 1.000 = 3,72$$

Si la hipótesis nula es cierta, este debe ser un valor de una Chi- cuadrado con i-1=9 grados de libertad.

$$RC_{0,05} = \left[\mathrm{X}^2_{9;\,0,05},\ \ +\infty\right) = \left[16,93,\ \ +\infty\right)$$

CONCLUSIÓN: el estadístico no cae en la región crítica, luego nos quedamos con la hipótesis nula y concluimos que el algoritmo funciona bien.

79. Sobre una muestra de tamaño 5 de una población se midieron dos variables X e Y y se obtuvo que el coeficiente de correlación de Pearson para esa muestra concreta resultó ser –0,082. Estamos interesados en probar:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

a) ¿Conseguiremos probar con este contraste si las dos variables son independientes?, ¿por qué?

b) ¿Cuál es el estadístico adecuado para realizar esta prueba? ¿Qué valor toma el estadístico en este caso?

c) Suponiendo que el valor del estadístico del contraste fuera 1,2 y tomando α = 0,1, ¿cuál de entre las siguientes sería la conclusión?

i) Rechazar H_0 y decir que la correlación es significativa.

ii) Rechazar H_0 y decir que la correlación no es significativa.

iii) Aceptar H_0 y decir que la correlación es significativa.

iv) Aceptar H_0 y decir que la correlación no es significativa.

v) Suponer que se ha cometido un error de cálculo, ya que el valor del estadístico tiene que estar entre 0 y 1.

(Señalar la respuesta adecuada y explicar por qué es esa la elegida).

Solución:

a) Si al realizar el contraste la conclusión es aceptar la hipótesis alternativa, habrá evidencia estadística de que el coeficiente de correlación es distinto de cero y podremos afirmar que las variables no son independientes; si tuviésemos que aceptar la hipótesis nula, que el coeficiente de correlación es cero, no llegaríamos a ninguna conclusión sobre la independencia de las variables.

b) Dado que los únicos datos que tenemos son el coeficiente de correlación y el tamaño de la muestra, el estadístico será:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -0,1425 \in t_{n-2}$$

c) La región crítica sería

$$RC_{0.1} = \left(-\infty,\ t_{3;\ 0.05}\right] \cup \left[t_{3;\ 0.95},\ +\infty\right) = \left(-\infty,\ -2.353\right] \cup \left[2.353,\ +\infty\right)$$

Si el valor estadístico hubiera sido 1,2 no caería en región crítica, luego nos quedaríamos con la hipótesis nula, la respuesta correcta sería la iv) aceptamos H_0 y la correlación hemos de aceptar que por lo que sabemos es cero, y por tanto no hay una correlación significativa.

80. Se clasificó la progenie de cierta pareja en tres grupos, atendiendo a cierto atributo, obteniéndose las frecuencias 10, 53 y 46. Según un modelo genético las frecuencias deberían estar en la proporción

$$p^2 : p(1 - p) : (1 - p)^2$$

¿Concuerdan los datos con el modelo?

Solución:

Se trata de un problema de bondad de ajuste. Denotemos A, B, C los tres resultados del carácter en estudio.

Sean X = nº de sujetos tipo A cuando se observan 109 sujetos.

Y = nº de sujetos tipo B cuando se observan 109 sujetos.

Z = nº de sujetos tipo C cuando se observan 109 sujetos.

(X, Y, Z) la variable discreta tridimensional asociada.

Lo que queremos estudiar es el contraste:

$$\begin{cases} H_0 : (X,\ Y,\ Z) \in M\left(p^2,\ 2p(1-p),\ (1-p)^2\right) \\ H_1 : (X,\ Y,\ Z) \quad \text{no sigue ese modelo} \end{cases}$$

Las frecuencias observadas las tenemos, pero tenemos que calcular las frecuencias esperadas. Para ello lo primero será hacer una estimación del parámetro p. Lo haremos por el método de máxima verosimilitud.

$$P(X=10,\ Y=53,\ Z=46)=$$

$$=\frac{109!}{10!\ 53!\ 46!}(p^2)^{10}(2p(1-p))^{53}((1-p)^2)^{46}=$$

$$=\frac{109!}{10!\ 53!\ 46!}2^{53}p^{73}(1-p)^{145}$$

Calculamos el valor de p que hace máxima esa probabilidad

$$\frac{\partial P}{\partial p}=\frac{109!\cdot 2^{53}}{10!\ 53!\ 46!}\left[73p^{72}(1-p)^{145}-145p^{73}(1-p)^{144}\right]=0 \Leftrightarrow$$

$$\Leftrightarrow p^{72}(1-p)^{144}\left[73(1-p)-145p\right]=0 \Leftrightarrow 73-73p-145p=0 \Leftrightarrow$$

$$\Leftrightarrow p=\frac{73}{218}=0,3348$$

Calculamos ahora las frecuencias esperadas

$$\begin{cases} E_1 = N\cdot \hat{p}^2 = 12,222 \\ E_2 = N\cdot 2\hat{p}(1-\hat{p}) = 48,555 \\ E_3 = N\cdot (1-\hat{p})^2 = 48,223 \end{cases}$$

Ahora podemos ya calcular el valor del estadístico del contraste, que es:

$$\sum_i \frac{(O_i-E_i)^2}{E_i}=\sum_i \frac{O_i^2}{E_i}-N=109,91-109=0,91$$

Si la hipótesis nula es cierta éste debe comportarse como una ji-cuadrado con k-1=2 grados de libertad y restamos otro grado de libertad porque hemos tenido que estimar un parámetro, con lo que nos queda un solo grado de libertad.

$$RC_{0.05} = \left[\mathrm{X}^2_{1;\ 0.05},\ \ +\infty\right) = \left[3.841,\ \ +\infty\right)$$

CONCLUSIÓN: vemos que el estadístico no cae en región crítica para ningún nivel de significación. Debemos aceptar que la muestra sigue el modelo propuesto.

81. El gerente de una empresa de neumáticos tiene la sospecha de que en una de las dos fábricas de la compañía existe un fallo en el control de calidad. Para contrastar esta sospecha se eligió 8 neumáticos de la primera fábrica y tras someterlos a un proceso acelerado de desgaste se obtuvo una duración media de 14300 km, con una desviación típica de 4350 km. En la segunda fábrica se eligió una muestra de 21 neumáticos, que sometidos al mismo proceso de desgaste dieron una media de 13800 km, con desviación típica de 3830 km. Estudiar si los desgastes medios son iguales.

Solución:

Sea X = duración de las ruedas A (km) $X \in N\ (\mu_1, \sigma_1)$.

Y = duración de las ruedas B (km) $Y \in N\ (\mu_2, \sigma_2)$.

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{cases}$$

El estadístico del contraste es:

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = 0,2856 \in t_{27}$$

La región crítica es

$$RC_{0,05} = \left(-\infty,\quad -1.703\right] \cup \left[\ 1.703,\ \ \infty\right)$$

CONCLUSIÓN: el estadístico no cae en región crítica, luego no podemos rechazar la hipótesis nula. Con la información que tenemos y a nivel de significación de un 5% no hay evidencia estadística de que los desgastes medios sean distintos.

82. De una población normal $N(\mu, 1)$ se observa una muestra de tamaño 5. Se considera el contraste de hipótesis simples

$$\begin{cases} H_0: & \mu = 1 \\ H_1: & \mu = 3 \end{cases}$$

y la región crítica dada por $RC = \{(X_1, X_2, X_3, X_4, X_5)/\overline{X} > 2{,}1\}$. Calcular la probabilidad de los dos tipos de error.

Solución:

$\overline{X} \in N\left(\mu,\ 1/\sqrt{5}\right)$ y por tanto:

$$P(\text{error tipo I}) = P\left(\overline{X} > 2{,}1 \quad / \mu = 1\ \right) = P\left(\frac{\overline{X}-1}{1/\sqrt{5}} > \frac{2{,}1-1}{1/\sqrt{5}}\right) =$$

$$= P(Z > 2{,}46) = 0{,}0069$$

$$P(\text{error tipo II}) = P\left(\overline{X} \le 2{,}1 \quad / \mu = 3\right) = P\left(\frac{\overline{X}-3}{1/\sqrt{5}} \le \frac{2{,}1-3}{1/\sqrt{5}}\right) =$$

$$= P(Z \le -2{,}01) = 0{,}222$$

83. Realizar un estudio lo más completo posible de la dependencia lineal de la concentración de glucosa en sangre y la de adrenalina en función de los siguientes datos, obtenidos para 8 pacientes.

Glucosa mg/c.c.	1.2	1.8	3.1	4.9	3.7	7.1	8.6	9.8
Adrenalina, mg/c.c.	4.5	5.9	7.0	7.8	7.2	6.8	4.5	2.7

Solución:

$$\sum_i x_i = 40,2 \qquad \overline{X} = 5,025$$

$$\sum_i y_i = 46,4 \qquad \overline{Y} = 5,8$$

$$\sum_i x_i^2 = 272,4 \qquad \Rightarrow \qquad \begin{array}{l} S_{xx} = 70,395 \\ S_{yy} = 21,4 \\ S_{xy} = -17,14 \end{array}$$

$$\sum_i y_i^2 = 290,52$$

$$\sum_i x_i y_i = 216,02$$

Calculamos los coeficientes de la recta de regresión

$$b = \frac{S_{xy}}{S_{xx}} = -\frac{17,14}{70,395} = -0,243$$

$$a = \overline{y} - b\overline{x} = 7,02$$

Y podemos escribir:

$$\mu_{Y/x} = 7,02 - 0,243x$$

Estudiamos el coeficiente de correlación

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = -0,44$$

Para estudiar si el modelo es adecuado y puesto que la correlación no es demasiado alta planteamos un contraste a ver si es significativamente distinta de cero.

$$\begin{cases} H_0 : \ \rho = 0 \\ H_1 \ : \ \rho \neq 0 \end{cases}$$

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -1,2 \in t_6 \qquad RC_{0,05} = \left(-\infty, \ -1,943\right] \cup \left[1,943, \ \ \infty\right)$$

CONCLUSIÓN: el estadístico no cae en región crítica, luego no podemos afirmar que el coeficiente de correlación sea distinto de cero (al menos para nivel de significación de 0,05) y por tanto rechazamos el modelo lineal y no lo usaremos para predicción.

84. En una clínica se quiere estudiar si existe diferencia significativa entre tres tipos de calmantes. Con este fin se suministran los calmantes a pacientes comparables. La variable de respuesta fue el tiempo en segundos desde la inyección hasta la desaparición del dolor. Los resultados obtenidos fueron:

Droga A	35 30 60 61 42
Droga B	41 32 70 56 80
Droga C	62 57 35 42 78 43

¿Producen las tres drogas el mismo efecto, en promedio? (Suponer normalidad)

Solución:

X_1 = tiempo que tarde en hacer efecto el calmante A, $X_1 \in N(\mu_1, \sigma_1)$.
X_2 = tiempo que tarde en hacer efecto el calmante B, $X_2 \in N(\mu_2, \sigma_2)$.
X_3 = tiempo que tarde en hacer efecto el calmante C, $X_3 \in N(\mu_3, \sigma_3)$.

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \mu_3 \\ H_1: & \text{las tres medias no son iguales} \end{cases}$$

Para aplicar el método del análisis de la varianza, supondremos que las tres varianzas son iguales $\sigma_1 = \sigma_2 = \sigma_3$.

Haciendo operaciones tenemos:

Variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F
Tratamientos Error	SSA=277.16 SSE=3652.83	k-1=2 N-k=13	SSA/(k-1) SSE/(N-k)	0,4932
Total	SST= 3930	N-1=15		

Para significación de 0,05 obtenemos $F_{2,13;0,95} = 3,81$.

CONCLUSIÓN: no hay evidencia estadística de que los efectos de los calmantes sean distintos en promedio.

85. Las edades al morir de una muestra de 19 individuos fallecidos de tuberculosis dan una media de 50 años con una desviación típica muestral de 6 años.

a) Hallar un intervalo de confianza al 95% para la media, suponiendo normalidad.

b) Qué tamaño muestral debe tomarse para tener una confianza del 95% de que la edad media estimada y la verdadera edad media no difieren en más de dos años.

Solución:

Sea X = edad de un individuo que muere de tuberculosis, $X \in N(\mu, \sigma)$.

a) $$I_{\mu,\,1-\alpha} = \left[\overline{X} \pm t_{n-1,\,\alpha/2}\frac{S}{\sqrt{n}}\right] = \left[50 \pm 1{,}743\frac{6}{\sqrt{19}}\right] = [47.548 \quad , \quad 52.425]$$

b) Como no conocemos el tamaño no sabemos los grados de libertad del estadístico T más próximo, por tanto aproximamos por la normal, suponiendo que el tamaño muestral pedido será suficientemente grande. De aquí:

$$Z_{\alpha/2}\frac{S}{\sqrt{n}} \leq 2 \Rightarrow 1.645\frac{6}{\sqrt{n}} \leq 2 \Rightarrow n \cong 25.$$

86. Tenemos un dado del que nos han dicho que cada una de las caras pares tiene el doble de probabilidad de ocurrir que cada una de las caras impares. Realizamos 80 lanzamientos con este dado y obtenemos:

cara	1	2	3	4	5	6
frecuencia	6	22	14	15	5	18

¿Podemos afirmar que el dado no es del modelo descrito?

Solución:

Es un problema de bondad de ajuste: nos piden ver si la probabilidad de cada cara se ajusta al modelo descrito.

Calculamos primero la probabilidad de cada cara de un dado del modelo propuesto

$$\begin{cases} P(1)=P(3)=P(5)=p \\ P(2)=P(4)=P(6)=2p \end{cases} \Rightarrow 9p=1 \Rightarrow p=1/9$$

El contraste viene dado por:

$$\begin{cases} H_0: & P(2)=P(4)=P(6)=2/9; \quad P(1)=P(3)=P(5)=1/9 \\ H_1: & \text{el dado no es de ese modelo} \end{cases}$$

Calculamos las frecuencias esperadas:

cara	1	2	3	4	5	6
observadas	6	22	14	15	5	18
esperadas	8,88	17,77	8,88	17,77	8,88	17,77

El estadístico del contraste es:

$$\sum_{i=1}^{6}\frac{(O_i-E_i)^2}{E_i}=11,89$$

Y el punto crítico es:

$$\chi^2_{5,\ 0,95}=11,070$$

CONCLUSIÓN: para nivel de significación 0,05, el estadístico cae en región crítica. Rechazamos la hipótesis nula, el dado no es del modelo propuesto.

87. Una máquina embotelladora de gaseosas se considera fuera de control si la varianza del contenido excede el valor de 1,1 decilitros. En una muestra de 30 botellas se ha observado que la varianza es 1,5 decilitros. ¿Puede decirse que la máquina está fuera de control?

Solución:

Estudiaremos con un contraste de hipótesis si la varianza muestral de 1,5 es achacable a que la máquina está fuera de control o al azar.

$$\begin{cases} H_0 : \sigma^2 = 1,1 \\ H_1 : \sigma^2 > 1,1 \end{cases}$$

El estadístico del contraste será:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{29 \cdot 1,5}{1,1} = 39,54 \in X_{29}^2$$

Y la región crítica, para nivel de significación de 0,05 es [42,56, +∞].

CONCLUSIÓN: como el valor del estadístico del contraste no cae en la región crítica, no hay evidencia, a este nivel de significación, de que la máquina se halle fuera de control.

88. Se está estudiando el absentismo laboral femenino de una empresa. Se han elegido al azar diez obreras de una determinada sección de la misma y se anota el número de días de falta al trabajo por diversos motivos durante los últimos cuatro meses. Estos fueron los resultados.

Empleada	1	2	3	4	5	6	7	8	9	10
Absentismo (días)	5	4	6	8	7	4	2	7	6	1

Calcular un intervalo de confianza para el promedio de días de absentismo de todas las empleadas en los últimos cuatro meses.

Solución:

Sea X = número de días que faltó al trabajo una obrera en los últimos 4 meses. Consideraremos la variable como continua y normal, $X \in N(\mu, \sigma)$.

De la muestra se tiene $\overline{X} = 5$, $S = 2,26$.

Y el intervalo de confianza para la media viene dado por:

$$I_{\mu;\,0,05} = \left[\overline{X} \pm t_{9;\,0,025} \frac{S}{\sqrt{n-1}} \right] = \left[5 \pm 2,262 \cdot \frac{2,26}{3} \right] =$$
$$= [5 \pm 1,70] = [3,29, \quad 6,70]$$

89. De acuerdo con un estudio, las viudas viven más tiempo que los viudos. Con los siguientes datos de supervivencia, obtenidos para 100 viudas y 100 viudos después de la muerte del cónyuge, ¿se puede concluir que las proporciones de viudas y viudos son iguales respecto a la diferencia de periodos de tiempo que sobreviven después de la muerte de su compañero?

Años vividos	Viuda	Viudo
Menos de 5	25	39
Entre 5 y 10	42	40
Más de 10	33	41

Solución:

Sean A_1 = una persona sobrevive menos de 5 años a su cónyuge

A_2 = una persona sobrevive entre 5 y 10 años a su cónyuge

A_3 = una persona sobrevive más de 5 años a su cónyuge

H = hombre

M = mujer

$$\begin{cases} H_0 : P(A_i / H) = P(A_i / M), \text{ para } i = 1,2,3 \\ H_1 : \text{no ocurre } H_0 \end{cases}$$

Las frecuencias esperadas son:

Años vividos	Viuda	Viudo	total
Menos de 5	25 / 29,09	39 / 34,91	64
Entre 5 y 10	42 / 37,28	40 / 44,73	82
Más de 10	33 / 33,63	41 / 40,36	74
total	100	120	220

El estadístico del contraste es:

$$\sum_{1}^{3}\frac{O_i^2}{E_i}-N=222,174-220=2,174\in \mathrm{X}_2^2$$

Pero,

$$\mathrm{X}^2_{2;\ 0,05}=5,99$$

CONCLUSIÓN: a nivel de significación del 0,05 nos quedamos con que la proporción de viudos que mueren antes de los 5 años de la muerte de su cónyuge es igual que la correspondiente proporción de viudas, o sea que la supervivencia a la muerte del cónyuge es independiente del sexo.

90. Calcule e interprete el coeficiente de correlación muestral para las siguientes calificaciones de 6 alumnos elegidos al azar. Estudie si puede considerarse el coeficiente de correlación poblacional significativamente distinto de cero:

Matemáticas	70	92	80	74	65	83
Inglés	74	84	63	87	78	90

Solución:

$$\begin{array}{lll} \sum X_i=464, & \sum X_i^2=36354 & \qquad S_{xx}=471,33 \\ \sum Y_i=476, & \sum Y_i^2=38254 & \Rightarrow S_{yy}=491,33 \\ \sum X_iY_i=36926 & & \qquad S_{xy}=115,33 \end{array}$$

Y de aquí:

$$r=\frac{S_{xy}}{\sqrt{S_{xx}\ \ S_{yy}}}=0,24$$

Estudiamos el valor del coeficiente de correlación poblacional mediante el contraste:

$$\begin{cases} H_0: & \rho = 0 \\ H_1: & \rho \neq 0 \end{cases}$$

El estadístico del contraste es:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 0,466 \in t_{n-2} \text{ y } t_{4;\,0,90} = 1,533$$

CONCLUSIÓN: como el estadístico del contraste no cae en región crítica nos quedamos con la hipótesis nula y afirmamos con las puntuaciones en Matemáticas e Inglés son incorreladas, con confianza del 95%.

91. Se elige una muestra de tamaño 2 de una población de Poisson para contrastar las hipótesis $H_0 : \lambda = 1$ frente a $H_1 : \lambda = 2$. Se considera como región crítica $\{\overline{X} \geq 1{,}5\}$. Hallar el nivel de significación y la potencia del contraste.

Solución:

$$\overline{X} = \frac{X_1 + X_2}{2}$$

Por tanto:

$$P\left(\overline{X} \geq 1,5\right) = P\left(X_1 + X_2 \geq 3\right)$$

Además como X_1 y X_2 son independientes y sabemos que la distribución de Poisson es aditiva tendremos que, bajo la hipótesis nula, será $X_1 + X_2 \in P(2)$ y por tanto,

$$\begin{aligned} \alpha &= P\left(\text{Aceptar } H_0 / \lambda = 1\right) = P\left(\overline{X} \geq 1,5 / \ \lambda = 1\right) = P\left(X_1 + X_2 \geq 3 / \lambda = 1\right) \\ &= 1 - P\left(X_1 + X_2 < 3\right) = 1 - P\left(X_1 + X_2 \leq 2\right) = 1 - 0.6767 = 0.3233 \end{aligned}$$

Para calcular la potencia cuando $\lambda = 2$ supondremos ese valor del parámetro y tendremos:

$$\text{pot}_{\lambda=2} = P\left(\text{rechazar } H_0 / \lambda = 2\right) = P\left(\bar{X} \geq 1{,}5 / \lambda = 2\right) =$$
$$= P\left(X_1 + X_2 \geq 3 / X_1 + X_2 \in P(4)\right) =$$
$$= 1 - P\left(X_1 + X_2 \leq 2\right) = 1 - 0.2381 = 0.7619$$

92. En una determinada marca de cigarrillos se efectúa un experimento para estudiar el contenido de alquitrán: se estudian 20 cigarrillos elegidos al azar de diferentes lotes. Los datos muestrales para el contenido de alquitrán son:

$$\bar{X} = 22 \text{ mg} \qquad S = 4 \text{ mg.}$$

Encontrar un intervalo de confianza al 90% para el contenido medio de alquitrán en un cigarrillo de la citada marca.

Solución:

$$I_{\mu;;\,0{,}10} = \left[\bar{X} \pm t_{19;\,0{,}05} \frac{S}{\sqrt{n}}\right] = \left[22 \pm 1{,}729 \frac{4}{2\sqrt{5}}\right] =$$
$$= \left[22 \pm 1{,}5464\right] = \left[20{,}4536 \ ,\ \ 23{,}5464\right]$$

93. Una empresa farmacéutica ha desarrollado un medicamento nuevo que requiere el control estricto de la dosificación del ingrediente activo de cada cápsula. El fármaco puede ser peligroso si se excede de la dosis, e inútil si la dosificación es menor que la especificada. El error de estimación máximo para la media es igual a 0,2 mg. Si se sabe que la desviación típica es igual a 0,7 mg, ¿cuántas cápsulas del medicamento deben estudiarse para tener una confianza del 99% de que es correcta la cantidad media de ingrediente por cápsula?

Solución:

Sea X = cantidad de medicamento en una cápsula (mg), supondremos

$$X \in N(\mu,\ 0{,}7)$$

$$I_{\mu;\,0.99} = \left[\bar{X} \pm Z_{\alpha/2} \frac{0{,}7}{\sqrt{n}}\right] \Rightarrow 2.58 \frac{0{,}7}{\sqrt{n}} \leq 0.2 \Rightarrow \left(\frac{2.58 \times 0.7}{0.2}\right)^2 \leq n$$
$$n \geq 8154$$

94. En una urna hay cientos de bolas de tres colores: blancas, rojas y negras. Nos han dicho que hay el triple de rojas que de blancas y el doble de negras que de rojas. Extraemos 20 bolas al azar y salen 4 blancas, 7 rojas y el resto negras. ¿Podemos afirmar que en la urna las bolas no estaban en la proporción indicada?

Solución:

Si la proporción de bolas blancas la llamamos p, la de bolas rojas será $3p$, y la de bolas negras $6p$. Por tanto p=0,1 y así:

$$P(B) = 0{,}1 \qquad P(R) = 0{,}3 \qquad P(N) = 0{,}6$$

Además como la urna tiene *cientos* de bolas consideraremos las extracciones independientes.

$$\begin{cases} H_0: P(B)=0{,}1, \quad P(R)=0{,}3 \quad P(N)=0{,}6 \\ H_1: \text{ las bolas estan en otra proporción} \end{cases}$$

	B	R	N
observadas	4	7	9
esperadas	2	6	12

El estadístico del contraste viene dado por:

$$\sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = 2 + \frac{1}{6} + \frac{3}{4} = \frac{35}{12} = 2{,}9 \in \mathrm{X}^2_{3-1}$$

El punto crítico para nivel de significación 0,05 es 5,991.

CONCLUSIÓN: no hay evidencia estadística de que las bolas de la urna no estén en la proporción indicada.

95. Se quiere hacer un estudio para predecir la calificación en Sistemas Operativos de un alumno de 2º de ITIS de la ULPGC a partir de las calificaciones obtenidas por el mismo alumno en la asignatura de Lenguajes de Programación de primer curso. Las calificaciones de una muestra aleatoria de alumnos son las de la tabla:

LP	2,5	7	5	4	5	6	5	8,3	5
SO	2	6	8	8	5,5	9	7	9	6

LP	6,5	5,5	3,5	5	7	8	3	4,5	6,2
SO	7	5,5	6	6,3	6,5	9	5	3	7,2

a) Hacer un estudio lo más completo posible de la regresión lineal.

b) ¿Puede admitirse que la pendiente de la recta de regresión es mayor que 1?

Solución:

$$\sum X_i = 96,5 \qquad \sum X_i^2 = 566,58 \qquad S_{xx} = 43,8579$$

$$\sum Y_i = 116 \qquad \sum Y_i^2 = 809,28 \qquad S_{yy} = 61,7239$$

$$\sum X_i Y_i = 659,09 \qquad S_{xy} = 33,9789$$

a) Los coeficientes de la recta de regresión vienen dados por

$$\begin{cases} b = \dfrac{S_{xy}}{S_{xx}} = 0,7747 \\ a = \overline{y} - b\overline{x} = 2,265 \end{cases} \Rightarrow \quad \hat{\mu}_{Y/x} = 2,265 + 0,7747x$$

Para estudiar la bondad del ajuste usamos el método de análisis de la varianza.

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

El estadístico del contraste será:

$$\frac{bS_{xy}}{S^2} \in F_{1,(18-2)} \text{ donde } S^2 = \frac{S_{yy} - bS_{xy}}{n-2} = 2,2125$$

Con lo que:

$$\frac{bS_{xy}}{S^2} = 11,897$$

El punto crítico es:

$$F_{1,16;\,0,95} = 4,49$$

CONCLUSIÓN: hay evidencia estadística, a nivel de significación 0,05, de que $\beta \neq 0$ y por tanto el modelo de regresión es aceptable para predicción.

En cuanto al grado de dependencia entre las variables tenemos el coeficiente de correlación muestral:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0,65306$$

que como vemos indica una correlación lineal alta.

b) Si planteamos el contraste

$$\begin{cases} H_0 : \beta \leq 1 \\ H_1 : \beta > 1 \end{cases}$$

dado que el valor muestral que estima β es el valor de b = 0,7747 no hace falta aplicar ninguna técnica para concluir que nuestra información muestral no presenta evidencia alguna de que β >1 y nos quedaremos con la hipótesis de que $\beta \leq 1$.

96. En un estudio sobre la importancia de la lactancia materna en la producción de anticuerpos en los bebés, se realiza un experimento que consiste en medir la proporción de niños que presentan una concentración de anticuerpos menor que un nivel de referencia prefijado. Se van a estudiar dos muestras de igual tamaño de entre los niños con lactancia materna y sin ella. Cuántos niños debemos elegir de cada población para aproximar la verdadera diferencia de proporciones con un error menor del 2%, a nivel de significación del 98%.

Solución:

Sea X= nº de niños con lactancia materna y nivel de defensas por debajo de lo aceptable, de los n que se van a observar, $X\in B(n, p_1)$.

Sea Y= nº de niños sin lactancia materna y nivel de defensas por debajo de lo aceptable, de los n que se van a observar, $Y\in B(n, p_2)$.

$$\hat{p}_1 = X/n,\ \hat{p}_2 = Y/n \text{ sabemos}$$

$$I_{p_1-p_2;\,1-\alpha} = \left[\left(\hat{p}_1 - \hat{p}_2\right) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{n}} \right]$$

En nuestro caso $\alpha/2 = 0{,}01$ y sabemos $p(1-p) \leq 1/4$ de aquí:

$$2{,}33\sqrt{0.5/n} \leq 0{,}02 \Rightarrow \left(\frac{2{,}33}{0{,}02}\right)^2 \cdot 0.5 \leq n \Rightarrow 6786{,}125 \leq n$$

97. Queremos hacer un estudio de la proporción de personas de tienen un nivel de audición menor que el 60%. Sabemos que dicha proporción es menor del 20% ¿Cuál es el tamaño mínimo de la muestra que debemos tomar para tener una seguridad del 95% de que la proporción verdadera no difiere de la proporción muestral en más de 0,02?

Solución:

Sea X= número de personas con nivel de audición menor que el 60% de los n observados: $X\in B(n, p)$.

$$I_{p;\,0{,}95} = \left[\hat{p} \pm Z_{0{,}025}\sqrt{\frac{p(1-p)}{n}} \right] \Rightarrow Z_{0{,}025}\sqrt{\frac{p(1-p)}{n}} \leq 0{,}02 \Rightarrow$$

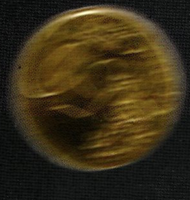
$$\Rightarrow \left(\frac{Z_{0{,}025}}{0{,}02}\right)^2 p(1-p) < n$$

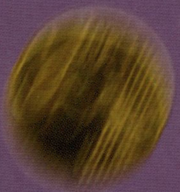
Como $p < 0{,}2$, se tiene

$$p(1-p) < 0{,}2\times 0{,}8 = 0{,}16 \Rightarrow \left(\frac{-1{,}96}{0{,}02}\right)^2 0{,}16 = 1536{,}64 < n$$

Muchos alumnos se quejan: "cuando veo este problema en un libro –es decir dentro de un contexto: una lección, un apartado...– lo sé resolver, pero luego, fuera de cualquier contexto, no sé qué herramientas debo usar". La mayoría de los libros de problemas de Probabilidad y/o Estadística adolecen de una, a mi juicio grave, desventaja: generalmente sus contenidos están tan estrictamente clasificados por objetivos, que el alumno sabe de antemano qué técnica o concepto se va poner en práctica con el ejercicio que pretende resolver, lo deduce por la ubicación, no por sus conocimientos. Así, estos libros ayudan a adquirir técnicas, pero no tanto a usar conceptos.

Los problemas que presento están bastante desordenados, tienen un único criterio de clasificación: son de Probabilidad o son de Estadística. Su desorden es parte de su valor. Pero el mayor valor y el mayor esfuerzo se ha puesto sin duda en la resolución paso a paso de cada ejercicio, lo que convierte el aprendizaje en un proceso que se retroalimenta: se aprende la práctica a partir de la teoría, y la teoría se acaba comprendiendo y fijando con la práctica.
Espero que pueda serle útil a mucha gente.





ISBN: 84-96502-40-6

La Caja de CANARIAS