UNIVERSIDAD TECNOLOGICA NACIONAL

Facultad Regional Buenos Aires

Departamento UDB Matemática

Probabilidad y Estadística

Guía de Trabajos Prácticos

Buenos Aires, Argentina, 2019

Índice general

Programa Analítico de la Asignatura

Unidad 1: PROBABILIDAD. Experimentos aleatorios. Espacios muestrales, sucesos y operaciones. Frecuencia relativa de un suceso. Probabilidad laplaciana. Definición axiomática de probabilidad y propiedades derivadas. Probabilidad condicional e independencia. Ley del producto. Teoremas de la probabilidad total y de Bayes.

Unidad 2: VARIABLE ALEATORIA. Variables aleatorias discretas y continuas. Función de probabilidad y de densidad de probabilidad. Función de distribución. Función de una variable aleatoria. Esperanza matemática de una variable aleatoria. Varianza. Desviación estándar. Momentos de orden superior. Propiedades. Covarianza y coeficiente de correlación lineal.

Unidad 3: DISTRIBUCIONES ESPECIALES. Binomial, Poisson, Uniforme, Gamma y Normal. Otras distribuciones especiales. Uso de tablas y de programas de computación para obtener los valores de las funciones asociadas. Aplicaciones.

Unidad 4: ESTIMACIÓN. Muestra aleatoria. Estimadores de parámetros de una distribución. Media y varianza muestrales. La estimación de la diferencia de medias. La estimación de la probabilidad de éxito de un ensayo de Bernoulli. El teorema central del límite. La distribución de los estimadores. Error cuadrático medio. Propiedades de los estimadores. Estimación por intervalos: diferentes casos.

Unidad 5: PRUEBA DE HIPÓTESIS. Hipótesis. Errores tipo I y II. Pruebas de hipótesis referentes a una media y a la diferencia de medias cuando se conocen las varianzas. Las pruebas ?t? de Student. La prueba ji-cuadrado para la varianza. Prueba sobre una proporción. El uso del valor p para la toma de decisiones. El concepto de significación estadística.

Unidad 6: REGRESIÓN Y CORRELACIÓN. El modelo de regresión lineal simple. Los estimadores de mínimos cuadrados de los parámetros de la regresión. El estimador de la varianza del error. El coeficiente de determinación. Prueba de significación de la regresión. Estimación del coeficiente de correlación. La predicción mediante el modelo.

Bibliografía de Consulta

- Devore, Jay L. *Probabilidad y Estadística para Ingeniería y Ciencias*. International Thomson Editores.
- Canavos, George C. *Probabilidad y Estadística - Aplicaciones y Métodos*. McGraw-Hill.
- Spiegel, Murray R. *Teoría y problemas de Probabilidad y Estadística*. Ed. McGraw-Hill, Serie Schaum; México.
- Walpole, Ronald E. y Myers, Raymond H. *Probabilidad y Estadística*. McGraw-Hill.
- Ross, Sheldon. *Introducción a la Estadística*. Ed. Reverté; México.
- Mendenhall, William. *Introducción a la Probabilidad y Estadística*. Ed. Cengage Learning; México.
- Maronna Ricardo A. *Probabilidad y Estadística elementales para estudiantes de Ciencias*. Facultad de Ciencias Exactas; Universidad Nacional de La Plata.

Capítulo 1

Probabilidades

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

- Experimentos aleatorios: espacio muestral, sucesos y eventos.
- Definiciones de probabilidad (clásica y axiomática). Relaciones entre sucesos.
- Sucesos especiales. **Propiedades derivadas de la definición de probabilidad**.
- Probabilidad condicional e independencia. Propiedades de independencia.
- **Teorema de la probabilidad total. Teorema de Bayes.**

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 1

1. Para cada uno de los siguientes experimentos, se pide definir el espacio muestral:

   a) Se analiza un tubo de ensayo con una muestra para detectar la presencia o ausencia de una molécula contaminante.

   b) Se seleccionan sucesivamente dos artículos de cierta producción y se clasifica cada uno en normal o defectuoso.

   c) Se arroja una moneda hasta obtener una cara.

   d) Se seleccionan dos billetes de una billetera que contiene uno de $\$50$, uno de $\$10$ y uno de $\$5$. Considerar el experimento con y sin reposición.

   e) De una caja que contiene bolillas blancas y negras se extraen sucesivamente bolillas hasta obtener dos blancas o cuatro bolillas cualesquiera.

f) Se mide el tiempo en minutos de espera en la parada del colectivo 7 entre las 23 hs. y las 24 hs. en Campus.

2. Describir por extensión los siguientes eventos correspondientes a los experimentos aleatorios descriptos en el ítem (b) del ejercicio anterior.

   - A: el primer artículo seleccionado es defectuoso.
   - B: al menos uno de los artículos es defectuoso.
   - C: ambos artículos son defectuosos.

   Indicar el valor de verdad de las siguientes proposiciones:

   a) $A \subseteq B$

   b) $C \subseteq A$

   c) $A \cap C$ es un evento imposible.

   d) El complemento de un evento imposible es cierto.

   e) B^c y A son eventos incompatibles.

3. Supongamos que se lanza una moneda equilibrada tres veces y se observan las caras superiores registrando cara o cruz según corresponda.

   a) Establecer el espacio muestral de este experimento.

   b) Asignar una probabilidad a cada punto. ¿Se trata de un espacio de equiprobabilidad?

   c) Sea A el evento de observar exactamente una vez cara y B el evento de observar al menos una cara. Obtener los puntos muestrales de A y B.

   d) A partir de la respuestas anteriores, calcular $P(A)$, $P(B)$, $P(A \cup B)$ y $P(A \cap B)$.

4. Una moneda está cargada de modo tal que la probabilidad de que salga cara es el triple de la probabilidad de cruz. Calcular ambas probabilidades.

5. Los frascos de mermelada tienen por lo general dos tipos de fallas: peso insuficiente o tapa no hermética. El 12 % contiene menos cantidad de la informada en la etiqueta, el 8 % tiene problemas con la tapa y el 3 % presenta ambas deficiencias. Los frascos en los que se detecta alguna de estas fallas son descartados. Hallar la probabilidad de que un frasco elegido al azar:

   a) tenga exactamente una falla.

   b) no sea descartado.

6. Se dispone de un blanco cuadrado de $2{,}5$ m de lado. Un dispositivo electrónico realiza disparos independientes que impactan en forma aleatoria en cualquier punto del cuadrado. Hallar la probabilidad de que:

   a) Un disparo impacte a menos de 1m del centro del blanco.

   b) Un disparo impacte a menos de 1m del centro pero a más de $\frac{1}{2}$ m del centro del blanco.

c) Si se hacen dos disparos, ambos caigan en el semiplano superior.

d) Si se hacen dos disparos, los dos caigan en distinto cuadrante.

7. Se tiene un dado cargado tal que la probabilidad de cada número par es proporcional a ese número (con la misma constante de proporcionalidad en todos los casos) y las probabilidades de los números impares son las mismas que en un dado normal. Consideremos los eventos A = "número par", B = "divisor de 6" y C = "múltiplo de 5".

   a) Hallar la probabilidad de cada elemento del espacio muestral.

   b) Hallar $P(A)$, $P(A \cup C)$, $P(A^c \cap B^c)$, $P(B - C)$,

8. Una clase consta de seis varones y diez mujeres. La tercera parte del primer grupo y la quinta parte del segundo usa anteojos. Hallar la probabilidad de que un alumno seleccionado al azar:

   a) use anteojos o sea mujer.

   b) sea varón y no use anteojos.

9. Se eligen al azar tres motores de un conjunto de diez, entre los cuales cuatro son a inyección. Hallar la probabilidad de que:

   a) ninguno sea a inyección.

   b) a lo sumo uno sea a inyección.

   c) al menos dos sean a inyección.

10. Calcular la probabilidad de que en una reunión de ocho personas se encuentren al menos dos que cumplan años el mismo día. (Sugerencia: Considerar un año de 365 días y ordenar a los individuos). ¿Y si son 23 personas?

11. Usando la propiedad $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, probar la fórmula

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

12. Un estudio de la conducta después de una re-educación en leyes de tránsito de un gran número de infractores, sugiere que la probabilidad de reincidencia dentro de los seis meses siguientes a la re-educación podría depender de los años de educación formal recibida por el individuo.

    Las proporciones del número total de casos que caen dentro de las cuatro categorías de educación-reincidencia se presentan a continuación:

	Condición dentro del período de seis meses, después de la re-educación		
Educación	Reincidente	No reincidente	Totales
7 años o más	0.10	0.30	0.40
Menos de 7 años	0.23	0.37	0.60
Totales	0.33	0.67	1.00

Supóngase que se selecciona al azar una única persona del programa de tratamiento. Sean los eventos:

A: la persona seleccionada tiene siete años o más de educación.

B: la persona seleccionada reincide dentro del período de los seis meses posteriores a la reeducación.

a) Encontrar las probabilidades de los eventos: A, B, $A \cap B$ y A/B.

b) Encontrar las probabilidades de los eventos: $A \cup B$ y $(A \cup B)^c$.

c) ¿Resultaron independientes los eventos A y B?

13. Sea una oficina Of_1 trabajan 3 mujeres y un varón, y en otra oficina Of_2 trabajan dos mujeres y dos varones. Se decidió transferir un empleado al azar de la primer oficina a la segunda. Si se selecciona, después de este traspaso, de la segunda oficina un empleado, qué probabilidad hay de que sea mujer?

14. (T) Sean tres eventos A , B y C asociados a un espacio muestral Ω, demostrar que:

    a) Si A no es imposible, entonces :$P(B \mid A) + P(B^c \mid A) = 1$

    b) Si $P(B \mid A) > P(B)$ entonces $P(B^c \mid A) < P(B^c)$

    c) Si C no es imposible, entonces : $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$

15. Una caja contiene 7 cubos azules y 3 verdes y una segunda caja contiene 6 cubos azules y 4 verdes. Se elige al azar un cubo de la primera caja y se lo pone en la segunda caja. Luego se selecciona al azar un cubo de la segunda caja y se lo pone en la primera caja.

    a) Hallar la probabilidad de que durante el proceso se seleccione un cubo azul de la primera caja y un cubo verde de la segunda caja.

    b) Hallar la probabilidad de que al finalizar el proceso las cajas queden como estaban inicialmente.

16. Un dado equilibrado se arroja dos veces. Se definen los eventos:

    - A : "el primer resultado es par"
    - B : "el segundo resultado es par"
    - C : "la suma de los resultados es par"

    Se pide:

    a) Mostrar que los eventos A, B y C son independientes de a pares.

    b) Mostrar que los eventos A, B y C no son mutuamente independientes.

17. (T) Demostrar que si los eventos A, B y C son mutuamente independientes y no son imposibles entonces:

$$P(A \cup B \cup C) = 1 - P(A^c) \cdot P(B^c) \cdot P(C^c)$$

18. Sean A y B dos eventos con probabilidad $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ y $P(B-A) = \frac{1}{12}$. Hallar $P(A \cap B)$ y $P(A^c|B)$.

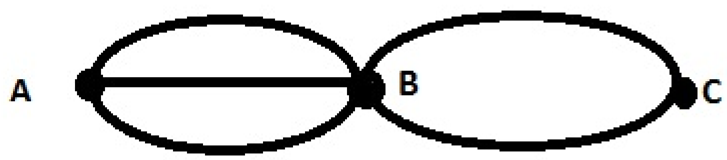
19. Calcular $P(B|A)$ si:

    a) A es un subconjunto de B
    b) A y B son incompatibles.

20. Se lanza un dado dos veces. Hallar la probabilidad de que la suma de sus números sea mayor o igual que ocho si

    a) aparece un cuatro en el primer dado.
    b) aparece un cuatro en, por lo menos, uno de los dados.

21. Existen tres caminos de A hasta B y dos caminos de B hasta C, como se muestra en la figura. Cada uno de estos caminos está bloqueado con probabilidad $0{,}1$ independientemente de los otros.



    Hallar la probabilidad de que:

    a) exista algún camino abierto que permita llegar de A hasta C.
    b) cuando el camino más corto que une A con B está cerrado, hallar la probabilidad de que se pueda llegar de C hasta A.

22. Una urna tiene siete bolas rojas y tres blancas. Si se sacan tres bolas, una tras otra sin reposición, ¿cuál es la probabilidad de que

    a) las dos primeras sean rojas y la tercera blanca?
    b) exactamente una sea blanca?

23. Tenemos dos urnas A y B. En A hay tres bolas rojas y dos blancas, y en B dos rojas y cinco blancas. Se elige una urna al azar y de ella una bola que se coloca en la otra urna. Luego se saca una bola de la segunda urna. Hallar la probabilidad de que las bolas sacadas sean de igual color.

24. (T) Sean A y B sucesos de un espacio muestral Ω. Probar que:

    a) Si A y B son independientes entonces son independientes: A y B^c (análogamente A^c y B). Deducir que A^c y B^c también son independientes.

b) Si los sucesos A y B son excluyentes ($A \cap B = \emptyset$) y $P(A)P(B) > 0$ entonces A y B **no** son independientes.

25. La probabilidad de que un hombre y una mujer vivan 10 años más, a partir de ahora, son $\frac{1}{3}$ y $\frac{1}{4}$ respectivamente. suponiendo independencia entre ambas superviviencias, calcular la probabilidad de que:

    a) ambos estén vivos dentro de diez años.

    b) al menos uno de ellos esté vivo dentro de diez años.

    c) ninguno esté vivo dentro de diez años.

    d) sólo la mujer esté viva dentro de diez años.

26. Se tienen tres cajoneras idénticas, cada cajonera a su vez tiene dos cajones. En una de las cajoneras hay una moneda de plata en cada cajón, en otra hay una de plata en un cajón y una de oro en el otro, mientras que en la última hay una de oro en cada cajón. Se elige una cajonera al azar y de ésta se saca una moneda de uno de los cajones: Si la moneda es de oro, ¿cuál es la probabilidad de que en el otro cajón haya una moneda de oro?

27. Uno de cada 25 adultos de cierta población está afectado de una enfermedad para la cual se ha desarrollado una prueba diagnóstica. La prueba es tal que, cuando un individuo padece la enfermedad, el resultado de la prueba es positivo en un $98\,\%$ de las veces, mientras que un individuo sano tendrá un resultado positivo solamente el $3\,\%$ de las veces. Eligiendo un individuo al azar de esta población:

    a) Hallar la probabilidad de que el test dé positivo.

    b) Hallar la probabilidad de que realmente padezca la enfermedad, sabiendo que el resultado de la prueba fue positivo.

    c) Hallar la probabilidad de que realmente esté sano, sabiendo que el resultado de la prueba fue negativo.

28. La caja A contiene 8 artículos de los cuales 3 son defectuosos, la caja B contiene 5 artículos de los cuales 2 son defectuosos y la caja C contiene 10 artículos de los cuales 4 son defectuosos. Se extrae al azar un artículo de cada caja.

    a) ¿Cuál es la probabilidad de que todos los artículos seleccionandos sean defectuosos?

    b) ¿Cuál es la probabilidad de que sólo un artículo de los seleccionados sea defectuoso?

    c) Si un solo artículo es defectuoso, ¿cuál es la probabilidad de que el artículo defectuoso proceda de la caja A?

29. En cierta facultad 40 % de los hombres y 55 % de las mujeres practican deporte. Además, el 70 % de los estudiantes son mujeres. Si se elige un estudiante al azar y hace deporte, ¿cuál es la probabilidad de que sea mujer?

30. De los pasajeros que viajan en Aerolineas Cordobesas, el $30\,\%$ viaja con familia, el $25\,\%$ con amigos y el resto solos. De los que viajan con familia, la mitad lo hace en vuelos nocturnos, de los que viajan con amigos el $70\,\%$ lo hace en vuelos nocturnos y de los que viajan solos, apenas el $25\,\%$ lo hace en vuelos nocturnos.

    a) Si se selecciona un pasajero al azar de Aerolíneas Cordobesas hallar la probabilidad de que no viaje en vuelo nocturno.

    b) Si viaja en vuelo nocturno, hallar la probabilidad de que viaje con amigos.

31. El Departamento de Ciencias Básicas de la Universidad tiene seis docentes titulares de Matemática: Euler, Cauchy, Leibnitz, Laplace, Gauss y Euclides, y debe seleccionar a tres docentes titulares de este área para integrar una comisión de revisión de contenidos. Debido a que el trabajo será tedioso, nadie desea hacerlo, luego se eligen los nombres de una urna al azar.

    a) Hallar la probabilidad de que alguno de los seleccionados tenga un apellido que comience con L.

    b) Si dos de los seleccionados tienen la misma inicial del apellido, ¿cuál es la probabilidad de que sea la L?

    c) Si las antigüedades docentes de los miembros titulares son respectivamente: 12, 15, 9, 16, 20 y 7 años; hallar la probabilidad de que la suma de antigüedades de dos de estos miembros elegidos al azar, resulte:

        1) 27 años
        2) mayor a 30 años.

Respuestas

1. *a*) $S_1 = \{$sí , no $\}$

    b) $S_2 = \{$NN, ND, DN, DD $\}$

    c) $S_3 = \{C,\ XC,\ XXC,\ XXXC,\ ...\}$

    d) $S_4 = \{(50;10),(50;5),(10;50),(10;5),(5;50),(5;10)\}$ sin reponer
    $S_5 = \{(50;50),(50;10),(50;5),(10;10),(10;50),(5;5),(10;5),(5;50),(5;10)\}$ con reposición

    e) $S_6 = \{BB, NBB, BNB, NNBB, NBNB, BNNB, BNNN, NBNN, NNBN, NNNB, NNNN\}$

    f) $S_7 = [0;60]$ en minutos.

2. $A = \{DN, DD\}$ $B = \{DN, DD, ND\}$ $C = \{DD\}$

    $(a)V$ $(b)V$ $(c)F$ $(d)V$ $(e)V$

3. *a*) $S = \{CCC,\ CCX,\ CXC,\ XCC,\ XXC,\ XCX,\ CXX,\ XXX\}$

    b) $S = \{s_1,\ s_2,\ s_3,\ s_4,\ s_5,\ s_6,\ s_7,\ s_8\}$ $p(s_i) = \frac{1}{8}$, siendo $1 \leq i \leq 8$

c) $A = \{CXX, XCX, XXC\}$ $B = \{CCC,\ CCX,\ CXC,\ XCC,\ XXC,\ XCX,\ CXX\}$

d) $P(A) = \frac{3}{8},\ P(B) = \frac{7}{8},\ P(A \cup B) = \frac{7}{8}; P(A \cap B) = \frac{3}{8}$

4. $P(X) = \frac{1}{4},\ \ P(C) = \frac{3}{4}$

5. a) 0.14

   b) 0.83

6. a) $\frac{4}{25}\pi$

   b) $\frac{3}{25}\pi$

   c) $\frac{1}{4}$

   d) $\frac{3}{4}$

7. a) $P(1) = P(3) = P(5) = P(4) = \frac{1}{6},\ P(2) = \frac{1}{12},\ P(6) = \frac{1}{4}$

   b) $P(A) = \frac{1}{2}, P(A \cup C) = \frac{2}{3},\ \ P(A^c \cap B^c) = \frac{1}{6}\ \ P(B - C) = \frac{2}{3}$

8. a) $P(A \cup M) = \frac{3}{4}$

   b) $P(V \cap A^c) = \frac{1}{4}$

9. a) $P(I = 0) = \frac{1}{6}$

   b) $P(I \leq 1) = \frac{2}{3}$

   c) $P(I > 2) = \frac{1}{3}$

10. $0{,}0743$

11. $P(A \cup B \cup C) = P(A \cup B) + P(C) - P(P((A \cup B) \cap C)$
$= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C))$
$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C).$

12. a) $P(A) = 0{,}4,\ P(B) = 0{,}33,\ P(A \cap B) = 0{,}1,\ P(A/B) = \frac{10}{33}$

    b) $P(A \cup B) = 0{,}63,\ P((A \cup B)^c) = 0{,}37$

    c) no

13. $\frac{11}{20}$

14. a) $P(B/A) + P(B/A^c) = \dfrac{P(B \cap A)}{P(A)} + \dfrac{P(B^c \cap A)}{P(A)} = \dfrac{P(A)}{P(A)} = 1$

    b) $P(B/A) > P(B) \Rightarrow \dfrac{P(A \cap B)}{P(A)} > P(B) \Rightarrow 1 - \dfrac{P(A \cap B)}{P(A)} < 1 - P(B)$
    $\Rightarrow \dfrac{P(A) - P(A \cap B)}{P(A)} < P(B^c) \Rightarrow \dfrac{P(A \cap B^c)}{P(A)} < P(B^c) \Rightarrow P(B^c/A) < P(B^c)$

c) $P\left((A \cup B)/C\right) = \dfrac{P((A \cup C) \cap (B \cup C))}{P(C)}$
$= \dfrac{P(A \cap C)}{P(C)} + \dfrac{P(B \cap C)}{P(C)} - \dfrac{P(A \cap B \cap C)}{P(C)}$
$= P(A/C) + P(B/C) - P(A \cup B/C)$

15. a) $P(\text{AV}) = \frac{14}{55}$
    b) $P(\text{igual al principio}) = \frac{32}{55}$

16. a) $P(A) = P(B) = P(C) = \frac{1}{2}$
    $P(A \cap B) = P(C \cap B) = P(A \cap C) = \frac{1}{4}$
    b) $P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{2}\frac{1}{2}\frac{1}{2}$

17. $P(A \cup B \cup C) + P\left((A \cup B \cup C)^c\right) = 1$
    $P(A \cap B \cup C) + P\left((A^c \cap B^c \cap C^c)\right) = 1$
    $P(A \cap B \cup C) + P(A^c)P(B^c)P(C^c) = 1$
    $\Rightarrow P(A \cup B \cup C) = 1 - P(A^c)P(B^c)P(C^c)$

18. $P(B \cap A) = \frac{1}{4},\ \ P(A^c/B) = \frac{1}{4}$

19. a) $A \subseteq B \Rightarrow P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(A)}{P(A)} = 1$
    b) $A \cap B = \emptyset \Rightarrow P(B/A) = 0$

20. (a) $\frac{1}{2}$ (b) $\frac{5}{11}$

21. (a) $0{,}98901$ (b) $0{,}9801$

22. (a) $P(\text{RRB}) = \frac{7}{40}$ (b) $P(\text{exact 1B}) = \frac{21}{40}$

23. $P(\text{igual color}) = \frac{901}{1680}$

24. a) H: A y B independientes
    T: A y B^c independientes
    D: $P(A \cap B^c) + P(A \cap B) = P(A)$
    $P(A \cap B^c) + P(A)P(B) = P(A)$
    $P(A \cap B^c) = P(A)\,(1 - P(B))$
    $P(A \cap B^c) = P(A)P(B^c)$
    b) $P(A)P(B) > 0$, pero $P(A \cap B) = P(\emptyset) = 0$, no son independientes

25. (a) $\frac{1}{12}$ (b) $\frac{1}{2}$ (c)$\frac{1}{2}$ (d) $\frac{1}{6}$

26. $\frac{2}{3}$

27. (a) $0{,}068$ (b) $0{,}5765$ (c) $0{,}9991$

28. (a) $\frac{3}{50}$ (b) $\frac{87}{200}$ (c) $\frac{9}{29}$

29. $\frac{77}{101}$

30. (a) $0{,}5625$ (b) $0{,}4$

31. (a) $0{,}8$ (b) $\frac{1}{2}$ (c) i. $\frac{2}{15}$ ii. $\frac{4}{15}$

Capítulo 2

Variables aleatorias

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

- Variables aleatorias: discretas y continuas.
- Funciones de probabilidad puntual, de densidad y de distribución de probabilidad. Parámetros. Percentiles.
- Funciones de probabilidad conjunta y marginal. Covarianza. Coeficiente de correlación. Idea de independencia entre variables aleatorias.
- Momentos. **Esperanza y varianza: propiedades.**

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 2

1. Para cada variable aleatoria definida a continuación, indicar el recorrido y clasificarla:

   a) Q: "Número de estudiantes en la lista de un curso en particular, que están ausentes el primer día de clase de los 20 inscriptos"
   b) R: "Tiempo de espera en una caja de un banco antes de ser atendido"
   c) S: "Temperatura máxima y mínima medida en la estación meteorológica de La Plata un día cualquiera del año"
   d) T: "Cantidad de bicicletas en stock en una bicicletería al finalizar un día de la semana si comenzó la semana con 120 unidades y no realizó reposición"
   e) U: "Número de hijos que debe tener una pareja hasta tener 3 mujeres; siendo su tope 8 hijos"

f) V: "Cantidad de meses del año que una fábrica excede los límites permitidos de contaminación ambiental"

g) W: "Cantidad de dinero que se puede extraer sacando tres monedas de una caja que tiene 2 de 50 ctv, 3 de 25 ctv y 4 de $\$1$"

h) X: "Presión de un neumático de un coche que ha sido cargado con 32 libras y medido un día cualquiera"

i) Y: "Número de veces que se debe lanzar al aire una moneda para obtener dos caras o dos cecas consecutivas"

j) Z: "Número de ruedas que al recolocarlas al azar en un auto ocupan su posición original"

2. Se seleccionaron tres estaciones de servicio A, B y C de la ciudad de Buenos Aires que cuentan respectivamente con 5, 3 y 2 surtidores. En estas estaciones no siempre están todos los surtidores en uso. Dar el recorrido de las siguientes variables aleatorias:

   a) U: "Número total de surtidores entre las estaciones seleccionadas en uso"

   b) V: "Número de surtidores en uso de la estación B ; Número de surtidores en uso de la estación C"

   c) W: "Número de estaciones que tienen exactamente 2 surtidores en uso"

   d) X: "Diferencia en el número de surtidores en servicio entre la estación A y B"

3. De una caja con 6 bolillas azules y 2 rojas se extraen 3 sin reposición. Sea X: 'Cantidad de bolillas azules entre las tres extraídas'.

   a) Hallar la función de probabilidad puntual de X.

   b) Hallar la E(X) y V(X)

   c) Hallar $E(X^2),\ E(1/X),\ E(1/X^2)\quad \text{y}\quad V(X^2)$.

4. Se tienen dos urnas. La urna A tiene 6 bolitas rojas y 4 blancas. La urna B tiene 2 bolitas rojas y 7 blancas. Se extrae una bolita al azar de A y se coloca en B. A continuación se extraen de B **con reposición** 2 bolitas. Sea X: "cantidad de bolitas rojas extraídas de la urna B".

   a) Hallar la función de probabilidad de X.

   b) Graficar la función de probabilidad de X.

   c) Repetir (a) pero considerando **sin** reposición.

5. Sea X = número de neumáticos de un automóvil, seleccionado al azar, que tenga baja la presión.

   a) ¿Cuál de las siguientes tres funciones $p_i(x)$ es una función de probabilidad puntual para X?

   b) Obtener la función de distribución acumulada de X.

   c) Con la función de distribución de probabilidad seleccionada en (a), calcular: $P(2 < X < 4)$, $P(X < 2)$ y $P(X \neq 0)$.

x	0	1	2	3	4
$p_1(x)$	0.20	0.30	0.10	0.07	0.03
$p_2(x)$	0.40	0.10	0.10	0.10	0.30
$p_3(x)$	0.40	0.15	0.10	0.15	0.30

d) Si $p(x) = k(5 - x)$ para $x = 0, 1, 2, 3, 4$, ¿cuál debe ser el valor de la constante k para que p sea una función de de probabilidad?

6. La siguiente distribución de probabilidad corresponde a la variable aleatoria X: "cantidad de llegadas tarde a la clase de Probabilidad y Estadística en marzo de un alumno al azar".

x	0	1	2	3	4
$p(x)$	$0{,}3 + k$	$2k$	$0{,}2$	$0{,}1 + 5k^2$	$0{,}05$

a) Hallar el valor de k.

b) Calcular la probabilidad de que las tardanzas de un alumno elegido al azar sean 3.

c) Hallar el número más probable de tardanzas. ¿Coincide con el valor esperado de tardanzas?

7. En una fábrica se producen cadenas con 15, 25, 30 o 40 eslabones en proporciones 2:3:4:6. Se vuelca toda la producción en una sola cinta transportadora. Se eligen de la cinta al azar dos cadenas con reposición.

Se definen las variables aleatorias X: "promedio de eslabones de las cadenas elegidas" e Y: "longitud máxima de las cadenas elegidas". Determinar su recorrido.

a) Hallar la función de probabilidad puntual de cada una de las variables.

b) Calcular el valor esperado y la varianza de ambas variables.

8. Sea X la capacidad (en kg) de un lavarropas comprado por el próximo cliente que elige la marca *Cándida*. La función de probabilidad puntual de X está dada por:

x_i	5	7.5	10
$p(x_i)$	0.25	0.45	0.30

a) Calcular el valor esperado y la varianza de la variable X.

b) Si el precio del lavarropas es $Y = 20X - 7{,}5$, hallar el valor esperado y el desvío estándard del precio del lavarropas *Cándida*.

9. La cadena de gimnasios *Sporties* ofrece a sus socios un plan anual con opciones de pago. Para un socio seleccionado al azar, sea X = número de meses para pagar el plan. La función de

distribución acumulada de X es:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0{,}1 & \text{si } 1 \leq x < 2 \\ 0{,}4 & \text{si } 2 \leq x < 3 \\ 0{,}6 & \text{si } 3 \leq x < 6 \\ 0{,}9 & \text{si } 6 \leq x < 12 \\ 1 & \text{si } x \geq 12 \end{cases}$$

Utilizando la función de distribución, calcular las siguientes probabilidades:

a) $P(X < 5)$, $P(X > 2)$ $P(3 \leq X \leq 6)$ y $P(3 < X \leq 6)$

b) $P(X < 6/X \geq 3)$

10. Las máquinas tejedoras en una fábrica de elásticos utilizan rayo láser para detectar los hilos rotos. Cuando se detecta un hilo roto, se detiene la máquina tejedora. Consideremos la variable aleatoria X: "Cantidad de veces que se detiene la máquina por día". La función de probabilidad de X está dada por:

$$p(x) = \begin{cases} \frac{16}{31}\left(\frac{1}{2}\right)^x & x = 0, 1, 2, 3, 4 \\ 0 & \text{en otro caso.} \end{cases}$$

a) ¿Cuál es la probabilidad de que un día dado se detenga la máquina?.

b) Si las detenciones en dos días consecutivos son independientes, hallar la probabilidad de que el máximo de detenciones entre las del lunes y las del martes sea exactamente de 2 veces.

c) Hallar la función de probabilidad puntual del número máximo de detenciones por día considerando lunes y martes.

11. Un fabricante de automóviles tiene un programa de control de calidad que incluyen la inspección de materiales recibidos para verificar que no tengan defectos. Supongamos que recibe las cajas de cambio para los automóviles en lotes de 5 unidades. Se seleccionan al azar dos cajas de un lote para inspeccionarlas para decidir si son defectuosas o no.

a) Describir el espacio muestral asociado a este experimento.

b) Supongamos que el lote tiene dos cajas defectuosas y definamos la variable aleatoria W: "número de cajas defectuosas seleccionadas". Hallar las funciones de probabilidad y distribución de la variable W.

c) Hallar el valor esperado, el valor mediano y el valor más probable de la variable W.

12. La cantidad de veces que se usa en la fábrica, una herramienta por día para reparar ciertos equipos puede modelarse mediante una variable que tiene la siguiente función de probabilidad $P(x) = \dfrac{c}{(x+1)}$, siendo el recorrido de la variable $R_X = \{0, 1, 2, 3\}$.

a) Hallar el valor de la constante c.

b) Si esta herramienta se alquila a $\$100$ cada vez que se usa; hallar el valor esperado del costo de alquiler mensual (20 días hábiles).

13. La demanda semanal de taladros en cierto local comercial de la localidad de 'Arroyo Seco' sigue la siguiente función de distribución de probabilidades:

x	0	1	2	3	4
F(x)	0.2	0.55	0.8	0.95	1

Cada taladro vendido origina una ganancia de \$ 350, pero los que no se venden originan una pérdida de \$ 80 por unidad. ¿Cuántos taladros convendría disponer en stock para maximizar el beneficio semanal esperado ?

14. Indicar cuál/es de las siguientes son funciones de densidad de probabilidad:

a) $f(x) = 3x^2$ si $0 \leq x \leq 1$, $f(x) = 0$ en otro caso.

b) $f(x) = 3e^{-x/3}$ si $x > 0$, $f(x) = 0$ en otro caso.

c) $f(x) = \frac{2}{3}(x-1)$ si $0 \leq x \leq 3$, $f(x) = 0$ en otro caso.

15. El porcentaje de fallas de una producción industrial está dado por la variable aleatoria:

$$f_X(x) = \begin{cases} a(x - x^3) & 0 < x \leq 1 \\ 0 & \text{otro caso.} \end{cases}$$

a) Hallar el valor de a.

b) Hallar la función de distribución de probabilidades.

c) Hallar el valor esperado y la varianza.

16. Sea

$$f(x) = \begin{cases} \dfrac{1}{72}(x+1)^2 & \text{si } -1 \leq x \leq m \\ 0 & \text{en otro caso} \end{cases}$$

a) Hallar el número real $m > 0$ para que $f(x)$ sea la función de densidad de una variable aleatoria X.

b) Hallar la función de distribución acumulada de X.

c) Utilizando la función calculada, hallar $P(-1 \leq X < 3)$ y $P(-1 < X < 3)$.

d) Calcular el percentil 75 o tercer cuartil de la variable.

e) Calcular la función de densidad de la variable aleatoria $Y = 2X + 1$ y su valor mediano.

f) Hallar el valor esperado de X y la varianza de Y.

17. Sea X una variable aleatoria con densidad:

$$f_X(x) = \begin{cases} |x| & \text{si } -5c \leq x \leq 5c \\ 0 & \text{otro caso} \end{cases}$$

a) Hallar el valor de la constante c de modo tal que resulte una función de densidad de probabilidad.

b) Considerar los eventos: $A = \left\{x/x > -\frac{1}{2}\right\}$ y $B = \left\{x/x < \frac{1}{2}\right\}$ e indicar si se trata de eventos independientes.

18. Una empresa fabrica unos componentes eléctricos cuya duración (en años) está dada por una variable aleatoria T, cuya función de densidad es:

$$f_T(t) = \begin{cases} 1 & si \quad 0 \leq t < \frac{1}{2} \\ \frac{1}{2}e^{-(t-\frac{1}{2})} & si \quad t > \frac{1}{2} \\ 0 & \text{en otro caso} \end{cases}$$

a) Hallar la función de distribución acumulada F_T

b) El producto se considera *regular* si dura menos de tres meses, *bueno* si dura entre tres meses y tres años y *muy bueno* si dura más de tres años. Calcular los porcentajes de componentes *regulares*, *buenos* y *muy buenos* de la producción.

c) Se empaquetan los componentes en cajas de 20 unidades. Si un comprador encuentra en una caja 1 o más artículos *regulares*, la fábrica le proporciona una caja nueva en forma gratuita. Cierto usuario adquirió una caja, ¿cuál es la probabilidad de obtener otra de regalo?

19. Sea X una variable aleatoria con función de densidad $f(x) = \frac{1}{2}(3-x)$ para $1 \leq x \leq 3$ y 0 en caso contrario.

a) Hallar la función de distribución acumulada y graficarla.

b) Determinar r sabiendo que $P(X < r) = 2P(X > r)$.

c) Calcular $P\left(X \leq \dfrac{5}{2} | 2 \leq X \leq \dfrac{7}{2}\right)$.

20. Sea X una variable aleatoria con función de densidad definida por

$$f(x) = \begin{cases} ax & \text{si } 0 \leq x < 1 \\ a & \text{si } 1 \leq x < 2 \\ -ax + 3a & \text{si } 2 \leq x < 3 \\ 0 & \text{en otro caso} \end{cases}$$

a) Determinar la constante a.

b) Hallar la función de distribución acumulada de X y graficarla.

21. Sea X una variable aleatoria con función densidad definida por $f(x) = 1$ para $0 < x < 1$ y 0 en caso contrario. Calcular las funciones densidad de las variables aleatorias $Y = \ln(X)$ y $Z = 3X + 4$.

22. Sea X una una variable aleatoria con función densidad dada por $f(x) = 3x^2$ para $-1 \leq x \leq 0$ y 0 en caso contrario. Sea b un número real tal que $-1 < b < 0$. Calcular $P(X > b \mid X < b/2)$.

23. El tiempo que tarda un proceso electrónico es una variable aleatoria con media 2 hs y varianza $0{,}5$ hs^2. El costo del proceso es de \$3 por hora más un costo fijo de \$8. Hallar el costo esperado y su varianza.

24. La función de distribución de la demanda de combustible en miles de litros por día X en cierta boca de expendio es:

$$F(x)\begin{cases} 0 & \text{si } x < 0 \\ bx^2 & \text{si } 0 \leq x < 1 \\ b[2(x-1)+1] & \text{si } 1 \leq x < 3 \\ 1 - b(x-4)^2 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

a) Hallar la función de densidad de la demanda de combustible.

b) Hallar la demanda superada sólo el 20 % de los días.

25. Un ecologista desea marcar una región circular de muestreo de 10 m de radio, sin embargo el radio de la región resultante es una variable aleatoria cuya función de densidad está dada por:

$$f(r) = \begin{cases} \dfrac{3}{4}[1 - (10-r)^2] & \text{si } 9 \leq r < 11 \\ 0 & \text{en otro caso} \end{cases}$$

a) Hallar la probabilidad de que el radio difiera del deseado por el ecologista en a lo sumo 30 cm.

b) ¿Cuál es el área esperada de la región resultante?

26. Cierto laboratorio atiende análisis clínicos regulares y de urgencia. Sea X_1 el número de pacientes que se atiende por un análisis regular en un momento particular del día, y X_2 el número de pacientes que demandan atención de urgencia en este mismo tiempo. Si la función de probabilidad conjunta de X_1 y X_2 está dada por:

a) Hallar la probabilidad de que el número de pacientes regulares sea igual al número de pacientes de urgencia.

b) Hallar la probabilidad de que haya exactamente dos pacientes de urgencia.

c) Hallar la probabilidad de que haya dos pacientes de urgencia y al menos uno regular.

$X_1 \downarrow, X_2 \rightarrow$	0	1	2
0	0.08	0.07	0.04
1	0.06	0.15	0.09
2	0.05	0.04	0.16
3	0	0.14	0.12

d) Hallar las funciones de probabilidad marginales de X_1 y X_2. ¿Son estas variables independientes? Justificar la respuesta.

27. Siendo la función de probabilidad conjunta

$$f(x;y) = \begin{cases} c(x+y) & \text{con } x, y \in \{1,2,3\} \\ 0 & \text{en otro caso} \end{cases}$$

Hallar:

a) El valor de c para que $f(x;y)$ resulte una función de probabilidad conjunta.

b) $P(X = 1 \wedge Y < 4), P(Y = 1), P(X < 2/Y < 2)$.

c) $Cov(X;Y),\ \rho(X;Y)$

28. Se seleccionan al azar dos cartuchos de bolígrafo de una caja que contiene tres azules, dos rojos y tres verdes. Si X: "número de cartuchos azules elegidos" e Y: "números de cartuchos rojos seleccionados", calcular:

a) La función de probabilidad conjunta $p(x,y)$.

b) $P(A)$, siendo $A = \{(x;y)/x + y \leq 1\}$.

c) $P(X = 0/Y = 1)$ y $P(Y = 1/X = 0)$.

29. Considerando la función de probabilidad conjunta de (X,Y) dada por la siguiente tabla:

$X \downarrow, Y \rightarrow$	-1	0	1
-1	b	c	b
0	c	0	c
1	b	c	b

Sabiendo que $b + c = 0{,}25$; demostrar que $E(XY) = E(X)E(Y)$. ¿Son independientes las variables X e Y? ¿Cuánto vale el coeficiente de correlación lineal?

30. Supongamos que X_1 y X_2 son variables aleatorias de las que se sabe que:

$$E(X_1) = 1 \qquad E(X_2) = -2 \qquad \rho(X_1;X_2) = \frac{\sqrt{6}}{3} \qquad V(X_1) = 4 \qquad V(X_2) = 9$$

Hallar:

a) $E(3X_1 - 2X_2)$

b) $V(3X_1)$

c) $Cov(X_1; X_2)$

d) $V(X_1 + X_2)$

e) $V(2X_1 - 3X_2)$

Respuestas

1. Denotamos $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

   a) $R_Q = \{x \in \mathbb{N}_0 / 0 \leq x \leq 20\}$

   b) $R_S = [0; 300]$ min

   c) $R_S = \{(x_1; x_2) / -10^o \leq x_2 \leq x_1 \leq 40^o\}$

   d) $R_T = \{x \in \mathbb{N}_0 / 0 \leq x \leq 120\}$

   e) $R_U = \{3, 4, 5, 6, 7, 8\}$

   f) $R_V = \{x \in \mathbb{N}_0 / 0 \leq x \leq 12\}$

   g) $R_W = \{0{,}75,\ 1,\ 1{,}25,\ 1{,}5,\ 1{,}75,\ 2,\ 2{,}25,\ 2{,}5,\ 3\}$

   h) $R_X = [0; 32]$

   i) $R_Y = \{x \in \mathbb{N} / x \geq 2\}$

   j) $R_Z = \{0, 1, 2, 4\}$

2. *a*) $R_U = \{x \in \mathbb{N}_0 / 0 \leq x \leq 10\}$

   b) $R_V = \{(x_1; x_2) / 0 \leq x_1 \leq 3 \wedge 0 \leq x_2 \leq 2\}$

   c) $R_W = \{0, 1, 2, 3\}$

   d) $R_X = \{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$

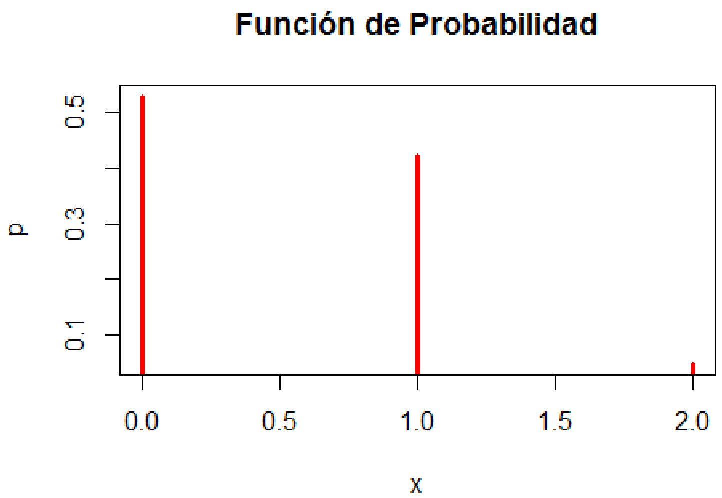
3. *a*) X: “cantidad de azules”, $P(X = 1) = \frac{3}{28},\ P(X = 2) = \frac{15}{28}, P(X = 3) = \frac{5}{14}$.

   b) $E(X) = \frac{9}{4},\ \ V(X) =, \frac{45}{112}$

   c) $E(X^2) = \frac{153}{28}, E(\frac{1}{X}) = \frac{83}{168},\ E(\frac{1}{X^2}) = 0{,}2808,\ V(X^2) = 7{,}7487$

4. *a*) $P(X = 0) = 0{,}55,\ P(X = 1) = 0{,}38,\ P(X = 2) = 0{,}07$

   b) Gráfico



c) $P(X = 0) = \frac{119}{225}$, $P(X = 1) = \frac{19}{45}$, $P(X = 2) = \frac{11}{225}$

5. a) $p_2(x)$

   b) $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0{,}4 & \text{si } 0 \leq x < 1 \\ 0{,}5 & \text{si } 1 \leq x < 2 \\ 0{,}6 & \text{si } 2 \leq x < 3 \\ 0{,}7 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$

   c) $P(2 < X < 4) = 0{,}1$, $P(X < 2) = 0{,}5$, $P(X \neq 0) = 0{,}6$

   d) $k = \frac{1}{15}$

6. (a) 0.1 (b) 0.15 (c) $Mo = 0$, $E(X) = 1{,}25$, no

7. X: "promedio de las dos longitudes", Y: "longitud máxima"

   a) Funciones de probabilidad

L	$p(L)$
15	2/15
25	0.20
30	4/15
40	0.40

   b) $E(X) = 31$, $E(Y) = 35{,}2$,$V(X) = 37$, $V(Y) = 38{,}2$

8. a) $E(X) = 7{,}625$, $V(X) = 3{,}4219$

   b) $E(Y) = 145$, $\sigma_Y = 37$

X	$p(X)$
15	4/225
20	12/225
22.5	16/225
25	9/225
27.5	48/225
30	16/225
32.5	36/225
35	48/225
40	36/225

Y	$p(Y)$
15	4/225
25	21/225
30	56/225
40	144/225

9. *a*) $P(X < 5) = 0{,}6$, $P(X > 2) = 0{,}6$, $P(3 \leq X \leq 6) = 0{,}5$, $P(3 < X \leq 6) = 0{,}3$

   b) $\frac{1}{3}$

10. *a*) $\frac{15}{31}$

    b) $\frac{208}{961}$

    c) V: "máximo entre lunes y martes"

X	$p(X)$
0	256/961
1	320/961
2	208/961
3	116/961
4	61/961

11. *a*) $S = \{BB,\ BD,\ DB,\ DD\}$

    b) Función de probabilidad

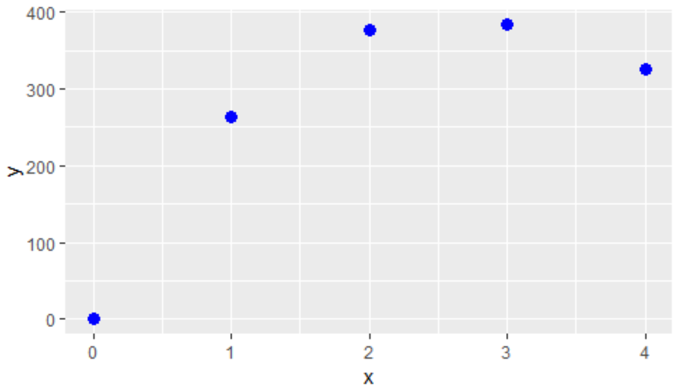
W	$p(W)$
0	0.3
1	0.6
2	0.1

    c) $E(W) = 0{,}8$, $med(W) = Mo(W) = 1$

12. *a*) $c = \frac{12}{25}$

b) $E(\text{alq.mens}) = \$1840$

13. Conviene tener un stock de 3 taladros.



14. (a) sí (b) no (c) no

15. a) a=4

b)

$$f(x) = \begin{cases} 0 & si \quad x < 0 \\ 2x^2 - x^4 & si \quad 0 \leq x < 1 \\ 1 & si \quad x \geq 1 \end{cases}$$

c) $E(X) = \frac{8}{15}$ $V(x) = \frac{11}{225}$

16. *a*) $m = 5$

b) $F(x) = \begin{cases} 0 & \text{si } x < -1 \\ \dfrac{(x+1)^3}{216} & \text{si } -1 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$

c) $\frac{8}{27}$

d) $x_{0,75} = \sqrt[3]{162} - 1$

e) $f_Y(y) = \dfrac{1}{576}(y+1)^2$ si $-1 \leq y \leq 11$ y 0 en otro caso.

f) E(X) = 3.5 y V(Y) = 5.4

17. (a) $c = \frac{1}{5}$ (b) no son independientes.

18. *a*) $F_T(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t < \frac{1}{2} \\ 1 - \frac{1}{2}e^{-(t-\frac{1}{2})} & \text{si } t \geq \frac{1}{2} \end{cases}$

b) $P(\text{regular}) = 0{,}25$, $P(\text{bueno}) = 0{,}75 - 0{,}5e^{-2{,}5}$, $P(\text{muy mueno}) = 0{,}5e^{-2{,}5}$

c) $P(X \geq 1) = 1 - 0{,}75^{20} = 0{,}9968$

19. *a*) $F(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{3}{2}x - \frac{1}{4}x^2 - \frac{5}{4} & \text{si } 1 \leq x \leq 3 \\ 1 & \text{si } x > 3 \end{cases}$

    b) $r \simeq 1{,}845$

    c) $\frac{3}{4}$

20. *a*) $a = \frac{1}{2}$

    b) $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{4}x^2 & \text{si } 0 \leq x < 1 \\ \frac{1}{4} + \frac{1}{2}(x-1) & \text{si } 1 \leq x < 2 \\ \frac{3}{2}x - \frac{1}{4}x^2 - \frac{5}{4} & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$

21. *a*) $f_Y(y) = \begin{cases} e^y & \text{si } y < 0 \\ 0 & \text{en otro caso} \end{cases}$

    b) $f_Z(z) = \begin{cases} \frac{1}{3} & \text{si } 4 \leq z \leq 7 \\ 0 & \text{en otro caso} \end{cases}$

22. $\dfrac{-7b^3}{b^3+8}$

23. $E(C) = 14$, $V(C) = \frac{9}{2}$

24. *a*) $f(x) = \begin{cases} \frac{1}{3}x & \text{si } 0 \leq x < 1 \\ \frac{1}{3} & \text{si } 1 \leq x < 3 \\ -\frac{1}{3}x + \frac{4}{3} & \text{si } 3 \leq x < 4 \\ 0 & \text{en otro caso} \end{cases}$

    b) $x_{0,8} = 2{,}9$

25. a) $0{,}4365$ b) 100π

26. *a*) $P(X_1 = X_2) = 0{,}39$

    b) $P(X_2 = 2) = 0{,}41$

    c) $P(X_2 = 2 \wedge X_1 \geq 1) = 0{,}37$

    d) no son independientes, $0{,}3 \cdot 0{,}4 \neq 0{,}15$

27. *a*) $c = \frac{1}{36}$

    b) $0{,}25,\ 0{,}25\ \frac{2}{9}$

    c) $cov(X,Y) = -\frac{1}{36}, \rho_{XY} = -\frac{1}{23}$

28. *a*) Tabla

$p(x;y)$	$Y=0$	$Y=1$	$Y=2$
$X=0$	3/28	6/28	1/28
$X=1$	9/28	6/28	0
$X=2$	3/28	0	0

b) $\frac{9}{14}$

c) $P(X=0/Y=1)=0{,}5$, $P(Y=1/X=0)=0{,}6$

29. a) no son independientes
b) $\rho_{XY}=0$

30. *a*) $E(3X_1-2X_2)=7$

b) $V(3X_1)=36$

c) $Cov(X;Y)=2\sqrt{6}$

d) $V(X_1+X_2)=13+4\sqrt{6}$

e) $V(2X_1-3X_2)=97-24\sqrt{6}$

Capítulo 3

Variables aleatorias especiales

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

- Distribuciones discretas especiales: Binomial, Hipergeométrica y Poisson.
- **Propiedad que vincula a las distribuciones Binomial y Poisson.**
- Distribuciones continuas especiales: Uniforme, Normal, Exponencial, Gamma, Chi Cuadrado y t de Student.
- **Propiedad de falta de memoria de la distribución Exponencial.**
- **Propiedad que vincula a las distribuciones Exponencial y Poisson.**
- Para las distribuciones Binomial, Poisson, Exponencial, Uniforme y Normal: **verificación de condición de función de probabilidad o densidad y cálculo de Esperanza y Varianza**.

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 3

1. Diez refrigeradores de cierto tipo han sido devueltos a un distribuidor debido a la presencia de un ruido oscilante agudo cuando el refrigerador está funcionando. Supongamos que 4 de estos 10 refrigeradores tienen compresores defectuosos y los otros 6 tienen problemas más leves. Si se examinan al azar 5 de estos 10 refrigeradores, y se define la variable aleatoria X: "el número entre los 5 examinados que tienen un compresor defectuoso". Indicar:

   a) la distribución de la variable aleatoria X.

   b) la probabilidad de que no todos tengan fallas leves.

   c) la probabilidad de que a lo sumo cuatro tengan fallas de compresor.

2. Un grupo de amigos del secundario se reúnen en la casa de Laura para comer un asado. En este grupo hay 6 varones y 8 mujeres, contando a Laura. De las mujeres 5 estudian Letras y el resto Exactas, mientras que de los varones sólo uno estudia Letras y el resto Exactas.

   a) Si las primeras en llegar a la casa son tres chicas, ¿cuál es la probabilidad de que las tres estudien lo mismo que Laura?

   b) Si tres cualesquiera de ellos hacen el asado, ¿cuál es la probabilidad de que estudien lo mismo?

   c) Si se seleccionan 2 al azar de este conjunto de amigos y se define la variable aleatoria X: "cantidad de amigos que estudia letras entre los dos elegidos", hallar el valor esperado y la varianza de X.

3. En una partida de truco, asumiendo que el mazo se encuentra bien mezclado, se reparte una mano de cartas.

   a) Hallar la probabilidad de que el jugador que recibe las primeras tres cartas tenga envido (dos cartas del mismo palo y una diferente).

   b) Hallar la probabilidad de que el primer jugador tenga flor (tres cartas del mismo palo).

   c) Hallar la probabilidad de que ambos tengan flor.

   d) Hallar la probabilidad de que el primer jugador no tenga ni flor ni envido.

4. La compañía de aviación *GranJet* ha determinado mediante un estudio estadístico que el $4\,\%$ de los pasajeros que reservan un viaje Buenos Aires - Misiones no se presentan a tomar el vuelo. Un día de mucha demanda de pasajes la empresa decide vender 72 pasajes de un vuelo con capacidad para 70 pasajeros. ¿Cuál es la probabilidad de que puedan viajar todos los pasajeros que se presentan a tomar el vuelo?

5. Se ha probado que el 25 % de los neumáticos de motocicleta sufren pinchaduras en caminos de ripio durante los primeros 1000km de uso. ¿Cuál es la probabilidad de que entre los próximos 6 neumáticos que se prueben:

   a) Al menos 2 sufran pinchaduras?

   b) A lo sumo 3 no sufran pinchaduras?

   c) No se supere el número esperado de pinchaduras?

6. Al probar motores para cuatriciclos se encontró que el 5 % no superaban la prueba.

   a) Hallar la probabilidad de que en los próximos 5 motores, al menos uno no pase la prueba?

   b) Hallar el número de motores que se espera pasen la prueba entre los próximos 20.

   c) Hallar la probabilidad de que no todos pero sí la mayoría de los próximos 7 motores no pasen la prueba.

7. En una empresa el $30\,\%$ de los empleados (con más de 2000 empleados) están en un proyecto de calidad, el $50\,\%$ están en un proyecto de expansión y el $70\,\%$ está trabajando en al menos uno de estos proyectos. Se seleccionan al azar para una encuesta de satisfacción cinco empleados de esta empresa.

a) Hallar la probabilidad de que al menos dos de los seleccionados estén exactamente en un proyecto.

b) Hallar la probabilidad de que a lo sumo tres estén en ambos proyectos.

c) Hallar la probabilidad de que todos los seleccionados estén en algún proyecto.

8. (T) Sea X una variable aleatoria con distribución Binomial de parámetros n y p.

   a) Demostrar que $E(X) = np$

   b) Demostrar que $V(X) = np(1-p)$

   c) ¿Existen valores de probabilidad p para los cuales se cumpla que $V(X) = 0$?

   Considerando un valor de n fijo:

   d) ¿Para qué valores de p es máxima $E(X)$?

   e) ¿Para qué valores de p es máxima $V(X)$?

9. (T) Sea X una variable aleatoria binomial con parámetros n y p. Demostrar que

$$P(x = k+1) = \frac{(n-k)p}{(k+1)(1-p)} P(x = k)$$

10. Los autos que pasan por cierta cabina de peaje siguen un proceso de Poisson con intensidad $\lambda = 20$ autos por hora. El peaje cuesta $\$45$ pero todos los automovilistas pagan con billetes de $\$50$. Al empezar el día, Fabián (que trabaja en la cabina) cuenta solamente con un billete de $\$5$.

    a) ¿Cuál es la probabilidad de que algún automovilista se quede sin vuelto en los primeros 5 minutos?

    b) Fabián quiere salir de la cabina para tomar un café. ¿Cuánto debería tardar como mucho si quiere la probabilidad de que en su ausencia no aparezca ningún auto sea de $\frac{1}{10}$?

11. La cantidad de pacientes que asisten a la sala de emergencias de la localidad de 'Moquehue' semanalmente, sigue una distribución de Poisson con media 3. Se pueden asistir 4 pacientes por semana, los que no pueden ser asistidos se derivan a la localidad de 'Villa Pehuenia'.

    a) ¿Cuál es la probabilidad de derivar algún paciente esta semana?

    b) ¿Cuál es el número esperado de pacientes que se derivan semanalmente?

    c) ¿Cuántos pacientes se deberían poder asistir para garantizar que no se realicen derivaciones el 90 % de las semanas?

12. Las llegadas a la fila de espera de una oficina del ANSSES ocurren según un proceso de Poisson de modo tal que la probabilidad de que no haya arribos en 5 minutos es e^{-1}

    a) Determinar qué cantidad de personas se espera que lleguen a la fila en una hora.

    b) Calcular la probabilidad de que pasen más de 6 minutos entre el arribo de dos personas consecutivas de la fila.

13. (T) Demuestre que si $X \sim Po(\lambda)$ se cumple que: $P(x = k+1) = \frac{\lambda}{(k+1)} P(x = k)$.

14. De una variable aleatoria uniforme se sabe que su valor esperado es 15 y que $P(13 \leq x \leq 18{,}5) = 0{,}55$, hallar su varianza y $P(x \leq 13)$.

15. El peso de ciertos bultos se distribuye uniformemente en el intervalo (a, b). Supongamos que el $20\,\%$ de los bultos pesa menos de 4 kg y el $40\,\%$ más de 8 kg.

    a) Hallar la probabilidad de que un bulto elegido al azar pese entre 5 y 9 kg.

    b) Si se eligen ocho bultos de estos aleatoriamente, hallar la probabilidad de que ninguno pese entre 5 y 9 kg.

    c) Hallar la cantidad de bultos de estos ocho, que se espera pesen entre 5 y 9 kg.

16. La duración de un amplificador electrónico está modelada como una variable Exponencial. Si el $10\,\%$ de los amplificadores tiene una duración media de 20000 horas, mientras que la del resto es de 50000 horas, hallar:

    a) la proporción de amplificadores que fallan antes de las 60000 horas.

    b) la probabilidad de que si la duración de un amplificador supera las 20000 horas, supere las 40000 horas.

17. La duración de ciertas lámparas eléctricas se distribuye exponencialmente. Se sabe además que el 80 % dura más de 800 horas. ¿Cuántas lámparas se deben extraer como mínimo para que la probabilidad de que alguna funcione más de 600 horas sea al menos del 90 %?

18. El ingreso anual de los jefes de familia de una cierta ciudad se puede modelar con una distribución Exponencial con $\lambda = 0{,}00005$. Para clasificar a los hogares de esa ciudad se ha decidido dividir a la población en 5 grupos igualmente numerosos: clase baja, clase media-baja, clase media, clase media-alta y clase alta, de modo que el 20 % de la población pertenezca a cada uno de ellos; es decir, el 20 % de los hogares con menores ingresos entran dentro de la clase baja, el segundo 20 % será clasificado dentro de la clase media-baja, etc. Hallar los salarios que indican el salto de categoría.
    Observación: Los valores hallados representan los cuantiles correspondientes a 0.20, 0.40, 0.60 y 0.80 respectivamente.

19. Sabiendo que $X \sim Exp(\lambda)$, hallar la distribución de $Y = cX$.

20. La duración de ciertas pilas se distribuye en forma Exponencial. Se sabe además que el $20\,\%$ tiene una duración inferior a 400 horas. ¿Cuál es la probabilidad de que entre cinco pilas elegidas al azar todas duren menos de 500 horas?.

21. Las baterías de una marca de celulares pueden provenir de dos fábricas diferentes, H y M. El $60\,\%$ de las pilas vendidas son producidas por la fábrica H y tienen una duración (en miles de horas) Exponencial cuyo valor esperado es $\mu = 100$. El resto son producidas por la otra fábrica, con una duración (en miles de horas) con distribución Uniforme $[80, 130]$.

a) Si una batería del primer fabricante dura más de 70.000 horas, ¿qué probabilidad hay de que dure más de 90.000 horas?

b) Calcular la probabilidad de que una batería vendida del stock dure a lo sumo 90.000 horas.

c) Si efectivamente la batería seleccionada dura a lo sumo 90.000 horas, ¿cuál sería la probabilidad de que proviniera de la fábrica H?

22. El número de clientes que atiende un cajero de banco sigue una distribución de Poisson con intensidad de 2 clientes cada 15 minutos.

a) ¿Cuál es la probabilidad de que el tiempo de atención de un cliente resulte superior a 20 minutos?

b) ¿Cuál es la probabilidad de que se demore más de 10 minutos en atender un cliente?

c) Si se demoró más de media hora en atender un cliente, hallar la probabilidad de que esa atención se demore mas de 40 minutos.

d) Comparar los resultados de los dos ítems anteriores y explicar.

e) ¿Cuál es la probabilidad de que en menos de 4 horas se atiendan 30 clientes?

23. Los mensajes que llegan al nodo A de en un sistema de comunicación de datos, son puestos en paquetes antes de ser transmitidos por la red. Si la llegada de mensajes a este nodo sigue una distribución de Poisson con una intensidad de 30 mensajes por minuto y sabiendo que se utilizan 6 mensajes para formar un paquete,

a) ¿Cuál es el tiempo promedio necesario para formar un paquete, esto es, el tiempo que transcurre hasta que lleguen los seis mensajes al nodo?

b) ¿Cuál es la varianza del tiempo necesario para formar el paquete?

c) ¿Cuál es la probabilidad de formar un paquete en menos de 12 segundos?

24. El tiempo semanal Y (en horas) durante el cual cierta máquina industrial no funciona, tiene una distribución Gamma con $\alpha = 3$ y $\lambda = 0{,}5$. La pérdida semanal L (en cientos de pesos) para la industria debido a esta baja está dada por:

$$L = 30Y + 8$$

(3000 pesos por hora en que la máquina no funciona más 800 de costo fijo mensual). Calcular la probabilidad de que se pierdan más de 18800 pesos en una semana.

25. El tiempo semanal Y (en horas) durante el cual cierta máquina industrial no funciona tiene distribución Γ con parámetros $\alpha = 100$ y $\lambda = 20$. La pérdida en pesos para la operación industrial esta dada por $L = 20Y - 3Y^2$. Calcular el valor esperado de L.

26. Dos componentes idénticas, A y B, funcionan en forma independiente en un sistema S. El tiempo hasta la ocurrencia de una falla para cada una de estas componentes puede considerarse una variable aleatoria con distribución Exponencial de parámetro $\alpha = (100hs)^{-1}$. Obtener las funciones de distribución y de densidad de probabilidad para el tiempo, T, hasta la ocurrencia de falla del sistema S:

a) considerando que las componentes están conectadas en serie.

b) considerando que las componentes están conectadas en paralelo.

27. Sea X una variable aleatoria Normal con $\mu = 5$ y $\sigma = 10$. Hallar:

    a) $P(X < 0)$, $P(X > 10)$, $P(X \geq 15)$.

    b) $P(-20 < X < 15)$, $P(-5 \leq X \leq 30)$.

    c) el valor de x tal que $P(X > x) = 0{,}05$.

    d) el valor de x tal que $P(X < x) = 0{,}23$.

28. Sea $X \sim N(\mu, \sigma^2)$.

    a) Mostrar que $P(|X - \mu| \leq 1{,}24\sigma) = 0{,}785$.

    b) Hallar el valor de c (en términos de σ) que cumple $P(\mu - c \leq X \leq \mu + c) = 0{,}95$.

29. Se dice que un proceso de la calidad seis sigma que es una metodología de mejora de procesos, centrada en la reducción de la variabilidad de los mismos, consiguiendo reducir o eliminar los defectos o fallas en la entrega de un producto o servicio al cliente. El proceso es de la calidad seis sigma si la media del proceso está a menos de seis desviaciones estándar de la especificación más cercana. Si la distribución de las mediciones es Normal.

    a) Si la esperanza del proceso está centrada entre las especificaciones superior e inferior a una distancia de tres desviaciones estándar de cada una de ellas, ¿cuál es la probabilidad de que el producto no cumpla con las especificaciones?

    b) Si la media del proceso se desplaza hacia arriba respecto la del ítem (a) en $1{,}5$ desviaciones estándar, ¿cuál es la probabilidad de que un producto no cumpla con las especificaciones? Expresar las respuestas en parte por millón.

30. La distribución de los pesos de los alumnos varones de una facultad es aproximadamente Normal, con media $\mu = 75$ kg. y desviación típica 7 kg.

    a) Hallar la probabilidad de que un alumno, elegido al azar, pese más de 95 kilos.

    b) Estimar el número de alumnos , entre los 15000 de esta Facultad, con peso entre 80 y 95 kilos.

    c) Calcular el peso no superado por el $10\,\%$ de los alumnos.

    d) Si se seleccionan diez alumnos de esta población, hallar la probabilidad de que al menos la mitad tengan pesos superiores a 80kg.

31. La longitud de ciertas piezas fabricadas por una máquina se distribuye normalmente. Se sabe que el $15\,\%$ de las piezas mide menos de 13mm y el $10\,\%$ de las piezas mide más de 14mm. Una pieza se considera no apta para un proceso de ensamble si mide menos de 12.8mm o más de 13.8mm. Además el embalaje del producto se realiza en cajas que contienen 12 bolsas de 10 piezas cada una.

   a) ¿Cuál es la probabilidad de que una pieza seleccionada al azar no sea apta?
   b) Un cliente que recibe una bolsa de 10 piezas, decide la aceptación del pedido si encuentra a lo sumo una pieza no apta. ¿Cuál es la probabilidad de que el pedido no sea rechazado?

32. La dureza Rockwell de un metal se determina al presionar con un punto acerado la superficie del metal y después medir la profundidad de la penetración del punto. Suponga que la dureza Rockwell de cierto metal está normalmente distribuida con media de 50HR y desvío estándar de 3HR.
    a) Una pieza de metal será considerada aceptable si su dureza está entre 46 y 55. ¿Cuál es la probabilidad de que una pieza seleccionada al azar tenga una dureza aceptable?
    b) Si se desea que el $95\,\%$ de las piezas resulten aceptables respecto de su dureza y se quiere que el intervalo de aceptabilidad sea de la forma $(50-c, 50+c)$, ¿cuál deberá ser el valor de c?
    c) Considerando independencia entre la dureza de las piezas y seleccionando 8 piezas al azar de este metal, ¿qué cantidad se espera que resulte aceptable, con el criterio establecido en (a)?
    d) Si se seleccionan al azar tres piezas, hallar la probabilidad de que la máxima de las durezas no supere 52.

33. (T) Demuestre que si X tiene una distribución Normal con parámetros μ y σ^2, entonces $Y = aX+b$ también tiene una distribución Normal. ¿Cuáles son los parámetros de la distribución de Y?

Respuestas

1. *a*) $X \sim H(N=10, R=4, n=5)$
   b) 0.9762
   c) 1
2. *a*) $\frac{4}{35}$
   b) 0.2088
   c) $E(X) = \frac{6}{7}$, $V(X) = \frac{288}{637}$
3. *a*) $\frac{135}{247}$
   b) 0.04858
   c) 0.00247
   d) 0.40486
4. 0.78836
5. a) 0.466 b) 0.9624 c) 0.534

6. a) 0.22622 b) 19 c) 0.00019

7. X: "está exactamente en un proyecto", Y: "está en ambos proyectos", W: "está en algún proyecto"
   - *a*) $X \sim Bi(n = 5, p = 0{,}6)$, $P(X \geq 2) = 0{,}91296$
   - *b*) $Y \sim Bi(n = 5, p = 0{,}1)$, $P(Y \leq 3) = 0{,}99954$
   - *c*) $W \sim Bi(n = 5, p = 0{,}7)$, $P(W = 5) = 0{,}16807$

8. - *a*) Sugerencia: considerar la Binomial como suma de variables aleatorias Bernoulli independientes
   - *b*) Idem anterior
   - *c*) $p = 0$ o $p = 1$
   - *d*) $E(X) = np$ es máxima para $p = 1$
   - *e*) $V(X) = np(1 - p)$ es máxima para $p = 0{,}5$.

9. $P(X = k + 1) = C_{n,k+1}p^{k+1}(1 - p)^{n-k-1}$

$$= \frac{n!}{(k+1)!(n-k-1)!}\frac{p}{1-p}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}\frac{(n-k)p}{1-p}p^k(1-p)^{n-k}$$

$$\Rightarrow P(X = k + 1) = \frac{(n-k)p}{1-p}P(X = k)$$

10. - *a*) 0.4963
    - *b*) 6.9 min

11. a) 0.185 b) 0.32 c) $n \geq 5$

12. - *a*) $\lambda = 12$ por hora
    - *b*) $0{,}3012$

13. $P(X = k + 1) = \dfrac{e^{-\lambda}\lambda^{k+1}}{(k+1)!} = \dfrac{\lambda}{k+1}\dfrac{e^{-\lambda}\lambda^k}{k!}$

    $\Rightarrow P(X = k + 1) = \dfrac{\lambda}{k+1}P(X = k)$

14. $\frac{25}{3}$ $0{,}3$

15. (a) 0.4 (b) 0.0168 (c) $E(X) = 3{,}2$

16. (a) 0.7239 (b) 0.6529

17. $n \geq 2$

18. Los valores límites son: $4462{,}87$, $10216{,}51$, $18325{,}81$, y $32188{,}75$

19. $Y \sim \epsilon(\alpha/c)$

20. 0.00085

21. (a) 0.8187 (b) 0.436 (c) 0.8165

22. (a) 0.06948 (b) 0.2636 (c) 0.2636 (d) Prop. falta de memoria (e) 0.5939

23. *a*) 12seg
    b) $V(T) = 24\text{seg}^2$
    c) $0{,}5543$

24. 0.42319

25. $E(L) = 24{,}25$

26. *a*) en serie $T \sim \epsilon(1/50hs)$
    b) en paralelo $F_T(t) = (1 - e^{-\frac{t}{100}})^2$, $f_T(t) = \frac{1}{50}e^{-t/100} - \frac{1}{50}e^{-t/50}$

27. *a*) $P(X < 0) = 0{,}3085$; $P(X > 10) = 0{,}3085$; $P(X > 15) = 0{,}1587$
    b) $P(-20 < X < 15) = P(-5 < X < 30) = 0{,}83513$
    c) 21.45
    d) -2.39

28. (a) 0.78502 (b) $c = 1{,}96\sigma$

29. (a) 0.0027 (b) 0.0668

30. (a) 0.0021 (b) $\simeq 3531$ (c) 66 kg (d) 0.0699

31. (a) 0.27315 (b) 0.2019

32. (a) 0.861 (b) $c = 5{,}88$ (c) $E(X) = 6{,}9$ (d) 0.4177

33. Considerar la relación entre $F(Y)$ y $F(X)$.

Capítulo 4

Estimación

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

- Combinación lineal de variables aleatorias normales.
- Sucesiones de variables aleatorias. La Ley de los grandes números.
- Teorema del Límite Central.
- Estimación Puntual: Método de los momentos, máxima verosimilitud.
- Muestras. Estimadores puntuales: insesgamiento, consistencia, eficiencia y suficiencia.
- **Error cuadrático medio y su relación con el sesgo**.
- Distribución de la media y la varianza muestral para el caso Normal.
- Estimación por intervalos (para la media, diferencia de medias, la varianza y la proporción). **Deducción del intervalo de confianza para la media con varianza conocida y desconocida.**

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 4

1. Los siguientes datos corresponden a las temperaturas de unión de los O-rings ($^\circ F$), en una muestra de prueba de lanzamientos reales del motor de un cohete de un transbordador espacial:

$$84 \ 49 \ 60 \ 83 \ 66 \ 47 \ 45 \ 70 \ 68 \ 91 \ 65 \ 58 \ 33 \ 53 \ 78 \ 82 \ 55 \ 61 \ 49 \ 70$$

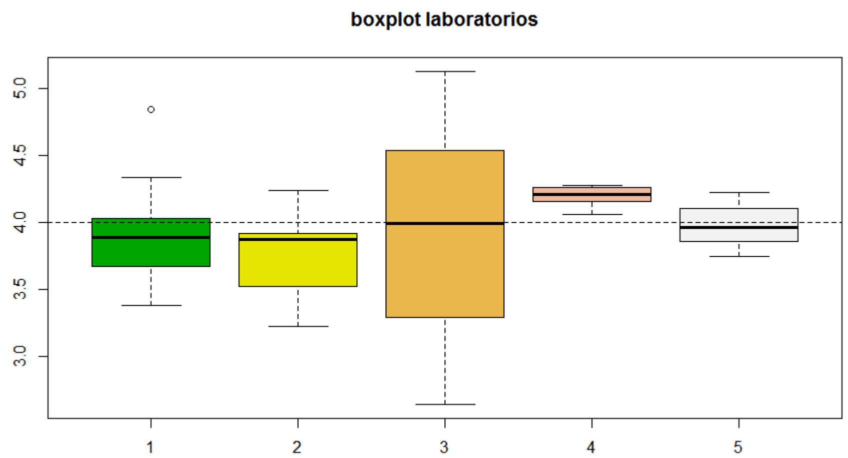
   a) Hallar la media aritmética y la mediana de la muestra.

b) ¿Qué indicaría un valor de la media muy superior al de la mediana?

c) Calcular la varianza y el desvío estándar muestrales.

d) Eliminar la observación mayor, repetir los cálculos y comparar los resultados.

2. Se desea elegir un laboratorio para enviar a procesar muestras de agua. Se postulan para esto cinco candidatos. Los boxplots correspondientes a 10 mediciones de cada laboratorio de una muestra de agua que contiene un $4\,\%$ de impurezas se exhiben a continuación:



a) ¿Cuál/es de los laboratorios tienen sesgo en sus mediciones?

b) Entre los menos sesgados: ¿cuál es el más preciso? Justificar.

3. Un circuito funciona con varias resistencias independientes en serie, cada una de las cuales tiene distribución Normal con media 20 ohms y varianza $0{,}36\text{ohm}^2$.

a) Si se instalan 6 de estas resistencias, ¿cuál es la probabilidad de que la resistencia del circuito sobrepase los 118 ohm?

b) ¿Cuántas resistencias deberán instalarse de manera que la probabilidad de que la resistencia total exceda los 200 ohm sea aproximadamente 0.05?

4. Una empresa agroalimentaria tiene tres sucursales en cierta región, cada una de ellas manejada por dos empleados. Los sueldos mensuales de estos empleados en miles de unidades monetarias se exhiben en la siguiente tabla:

Sucursal	1	1	2	2	3	3
Empleado	1	2	3	4	5	6
Sueldo	19.5	23.8	20.1	23.8	15.6	19.5

a) Se seleccionan al azar dos empleados entre estos seis con reemplazo y se calcula el promedio de sus sueldos. Hallar la distribución muestral del sueldo medio.

b) Se selecciona una de las tres sucursales al azar y se considera la variable aleatoria Z: "promedio de los sueldos de los dos empleados de la misma". Hallar la distribución de Z.

c) Compare los resultados de (a) y (b) con la media y la varianza del sueldo de un empleado elegido al azar de la empresa.

5. Se supone que el espesor de cierta pintura al aceite (X) en mm; tiene distribución Normal. Se realizaron las siguientes observaciones de espesores de esta pintura:

0.89	0.88	0.88	1.03	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.64	1.71	1.76	1.83

a) Calcular a partir de la muestra la media, la varianza, el desvío estándar y la mediana.

b) ¿Cuál/es de estos valores utilizaría para estimar el valor central de los espesores de esta pintura?

c) Si la muestra tuviera tamaño 1000, ¿qué estimador puntual elegiría para la media? ¿En qué basa su elección?

6. (T) Probar que si $X_1,\ X_2,\ \ldots,\ X_n$ son variables aleatorias independientes idénticamente distribuidas con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$, entonces:

a) $E(X_1 + X_2 + \cdots + X_n) = \text{n}\mu$

b) $V(X_1 + X_2 + \cdots + X_n) = \text{n}\sigma^2$

7. Una fábrica envasa paquetes de café cuyo peso es una variable aleatoria Normal de media 495 g y desvío estándar 18 g. Los paquetes se venden en cajas de 9 unidades. El peso de la caja de cartón es una variable aleatroria Normal de media 50 g y desvío estándard 4 g Si se selecciona una caja al azar, ¿cuál es la probabilidad de que el peso no supere los 4.500 g? (considerar los pesos de los paquetes como variables aleatorias independientes).

8. En una población determinada, la estatura de una persona adulta tiene un promedio de 165 cm con una desviación estándar de 7 cm.

a) Si se toma una muestra aleatoria de 100 personas en esta población, calcular la probabilidad de que la estatura media de la muestra exceda 168 cm.

b) Si se supone distribución normal de las estaturas, hallar la probabilidad de que en una muestra de 36 personas el desvío de la estatura resulte superior a 6 cm.

c) Si llamamos p a la propoción de personas de la población que superan el $1{,}70$ m, hallar la probabilidad de que dicha proporción en una muestra de tamaño 36 supere el $20\,\%$.

9. (*) La pintura para autopista de color amarillo tiene una especificación que recomienda que la varianza del tiempo de secado no supere los 20min^2. Sabiendo que la distribución del tiempo de secado es aproximadamente Normal con varianza 16min^2:

a) hallar la probabilidad de que la varianza del tiempo de secado en una muestra de tamaño 10 exceda los $42{,}77\text{min}^2$.

b) hallar la variabilidad superada por el $20\,\%$ de las muestras de tamaño 10 de tiempos de secado.

10. En un supermercado se decide redondear el cobro de las ventas al entero más cercano. Si en un día se realizan 800 ventas y se supone que el redondeo tiene distribución Uniforme, hallar la probabilidad de que en ese día, el supermercado pierda en el redondeo más de $\$9$.

11. La primera tarea en un curso introductorio de programación por computadora implica correr un breve programa. Si la experiencia indica que el $40\,\%$ de todos los estudiantes principiantes no cometerán errores tipográficos, calcular la probabilidad aproximada de que en un grupo de 80 estudiantes:

    a) por lo menos $35\,\%$ no cometan errores.

    b) entre el $45\,\%$ y el $55\,\%$ cometan errores.

12. Se examinan 180 piezas recién fabricadas y se registra el número de imperfecciones por pieza. Sería ideal que las piezas fabricadas no tuvieran imperfecciones. Sea X: "número de imperfecciones en una pieza seleccionada al azar" y supongamos que X tiene distribución Poisson de parámero λ. Se obtuvieron las siguientes observaciones:

Número de imperfecciones por pieza	0	1	2	3	4	5	6	7
Frecuencia observada	18	47	52	40	13	7	2	1

    a) Hallar las expresiones de los estimadores de momentos y de máxima verosimilitud correspondientes al valor del parámetro λ desconocido.

    b) Analizar si estos estimadores son insesgados.

    c) Con la estimación lograda, aproximar el valor de $P(X < 4)$.

13. Sea $X_1,\ X_2,\ X_3,\ X_4$ una muestra aleatoria de tamaño 4 de una población exponencial de parámetro α desconocido. Considere $f(x) = \alpha e^{-\alpha x}$.

    Se proponen los siguientes estadísticos para estimar el valor desconocido α. (Tener en cuenta el valor esperado de esta variable.)

    - $T_1 = \frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_3 + X_4)$
    - $T_2 = \frac{1}{5}(X_1 + 2X_2 + 3X_3 + 4X_4)$
    - $T_3 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$

    a) Seleccionar el o los estimadores insesgados entre los propuestos.

    b) Elegir el de menor varianza entre los seleccionados en (a).

    c) Ver si el error cuadrático medio del estimador elegido en (b) es el mínimo.

14. (T) Verificar que el error cuadrático medio de un estimador es la suma de la varianza del mismo y el cuadrado de su sesgo.

15. (T) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población, de la que se sabe que $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Demostrar que $\bar{X}^2$ no es un estimador insesgado de μ^2.

16. Se denota con X la proporción de tiempo que un estudiante, seleccionado al azar, emplea trabajando en cierta prueba de aptitud. Suponiendo que X tiene una distribución cuya función de densidad está dada por:

$$f(x) = \begin{cases} (\theta+1)x^{\theta} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

    a) Obtener el estimador de momentos de θ.

    b) Se toma una muestra aleatoria de 12 estudiantes, obteniendo las siguientes observaciones: $0{,}92 - 0{,}79 - 0{,}90 - 0{,}69 - 0{,}86 - 0{,}49 - 0{,}73 - 0{,}97 - 0{,}94 - 0{,}77 - 0{,}93 - 075$, calcular la estimación de θ para esta muestra.

17. Se quiere estimar el valor medio de las horas de estudio semanales de los alumnos recién ingresados a la UTN. Se toma una muestra de 100 alumnos y se obtiene un promedio 4 hs 20 min. Sabiendo que la distribución de estos tiempos es aproximadamente Normal con desvío de 55 min,

    a) construir un intervalo de confianza de nivel: $90\,\%$, $95\,\%$ y $99\,\%$ para el valor medio de las horas semanales de estudio.

    b) comparar las longitudes de los intervalos construidos en (a) y dar una conclusión.

18. Se quiere estimar la estatura promedio en la población que se supone tiene una distribución Normal con $\sigma = 4$ cm. Se toma una muestra de 144 personas y se obtiene que el promedio muestral es de 168 cm.

    a) Calcular el intervalo de confianza para la altura promedio de la población con un nivel de confianza del $98\,\%$.

    b) ¿De qué tamaño debería ser la muestra para mantener el mismo error muestral (que el obtenido en (a)) pero con un nivel de confianza del $95\,\%$?

19. (T) Supongamos que a partir de una muestra aleatoria $X_1, X_2, \ldots, X_{100}$, donde X_i es el tiempo de duración de cierto procedimiento que tiene distribución Normal con media de 20 minutos. Se construyó un intervalo de confianza para la varianza poblacional de nivel 0.98 y este intervalo resultó: $[5{,}28, 10{,}2]$min^2. Indicar si las siguientes afirmaciones son verdaderas o falsas, justificando adecuadamente en cada caso.

    a) La probabilidad de que σ^2 se encuentre en 5.28 y 10.2 es de aproximadamente $98\,\%$.

    b) Si en lugar de hacer 100 observaciones se hubieran tomado 1000 y se hubiera construido un intervalo de confianza de nivel $0{,}98$, este último con certeza habría resultado incluido en $[5{,}28, 10{,}2]$min^2.

c) Si se tomaran muchas muestras independientes de tamaño 100 y para cada una de ellas se calculara un intervalo de confianza con el mismo procedimiento, alrededor del $98\,\%$ de estos intervalos contendrían a σ^2.

d) Si se duplica el tamaño de la muestra, la longitud del intervalo se reduciría a la mitad.

e) Si se tomaran muchas muestras independientes de tamaño 100 los promedios muestrales de las mismas estarían contenidos en el intervalo: $[20 - 1{,}96 * \frac{\sqrt{5{,}28}}{10}; 20 + 1{,}96 * \frac{\sqrt{10{,}2}}{10}]$.

20. Un artículo de la revista American Ceramic Society Bulletin contiene información acerca de la distribución de las resistencias a la fractura, en MPa, de ciertas barras de cerámica quemadas en determinado horno: se sabe que tienen distribución Normal y que el desvío poblacional es $\sigma = 3{,}73$ MPa. Una muestra de tamaño $n = 121$ arrojó un promedio de $\bar{X} = 89{,}10$ MPa.

    a) Calcular un intervalo de confianza del $95\,\%$ para la resistencia real promedio.

    b) ¿Cómo se modificaría este intervalo, si se desconociera la varianza poblacional y la muestra arrojara una desviación estándar de 4.2 MPa?

21. Un artículo del Journal of Sports Science expone un intervalo de confianza para el promedio del nivel de hemoglobina de los jugadores de hockey sobre hielo (en g/dl) en la Olimpíada de Canadá de 20 jugadores, fue estimado con un 95 % de confianza, obteniéndose el siguiente intervalo (15.01,15.59)g/dl.

    a) Indicar los valores de $\bar{x}$ y s para esta muestra. Suponga que el nivel de hemoglobina de los jugadores de hockey sobre hielo en la Olimpíada de Canadá sigue una distribución Normal.

    b) Si se conociera la varianza poblacional y fuera de $0{,}36\ (g/dl)^2$, ¿cuál sería el nivel de confianza del intervalo construido?

22. La resistencia a la rotura, expresada en kg de 12 ejemplares de cuerda es de 280, 242, 270, 285, 273, 275, 266, 257, 285, 274, 258, 264. Estimar la resistencia media mediante un intervalo de confianza con nivel $0,95$ suponiendo distribución Normal.

23. Se tomó una muestra de n = 36 resistencias eléctricas resultando la desviación estándar muestral de las lecturas $s = 2{,}81$ ohm. Suponiendo normalidad, en la distribución de las resistencias hallar un intervalo de confianza de $95\,\%$ para σ.

24. Se midió la altura de un río en un punto geográfico dado (en m) durante 20 días al azar del verano, obteniéndose los siguientes valores en cm:

69.5	71.9	72.6	73.1	73.3	73.5	75.5	75.7	75.8	76.1
79.7	79.5	80.1	82.6	83.7	92.7	78.0	77.9	78.1	79.6

   a) ¿Qué condiciones se necesitan para que resulte válida la construcción de un intervalo de confianza para la varianza poblacional?

   b) Construir un intervalo de confianza de $99\,\%$ para la desviación estándar de la distribución de la altura de este río en verano.

25. Se quiere estimar la proporción de personas que han visto una cierta obra de teatro en cartel. ¿Qué tamaño de muestra se deberá tomar para asegurar un nivel de confianza del $99\,\%$ y un margen de error de $0{,}02$? Para el tamaño de muestra obtenido, hallar el intervalo de confianza para la proporción en cuestión, asumiendo que la proporción muestral es de $0{,}35$. ¿Se logró el error objetivo?

26. En una muestra de 81 piezas de tela 'modal' se obtuvo una elongación media muestral de $8{,}17\,\%$ y un desvío estándar muestral del $1{,}42\,\%$. Además la tercera parte de las telas observadas tuvo elongaciones superiores a $8{,}4$. Suponiendo que la elongación sigue una distribución Normal, estimar mediante un Intervalo del $95\,\%$ de confianza:

    a) el verdadero valor medio de elongación μ.
    b) el desvío estándar poblacional de la elongación σ.
    c) la proporción poblacional de telas con elongación superior a $8{,}4$.

27. Un artículo sobre "Evaluaciones de neumáticos bajo condición de esfuerzo" reporta que de una muestra aleatoria de 350 neumáticos, 54 mostraron daños severos, 48 manchas y 16 ambos tipos de defecto.

    a) Calcular un intervalo de confianza del $99\,\%$ para la proporción de neumáticos con algún defecto.
    b) Calcular un intervalo de confianza del $95\,\%$ para la proporción de neumáticos con más de un defecto.
    c) Calcular un intervalo de confianza del $95\,\%$ para la proporción de neumáticos con un único defecto.

28. Considere las siguientes muestras de contenido de grasa (en porcentaje) de salchichas seleccionadas al azar de dos marcas A y B:

A	25.2	21.3	22.8	17	29.8	21	25.5	16	20.9	19.5
B	24.2	20.7	21	18.9	29.1	20.5	25.2	16.3		

    Si se supone que los porcentajes de grasa de ambas marcas tienen distribución Normal:

    a) Hallar un intervalo de confianza del $95\,\%$ para la diferencia entre los contenidos medios de grasa de las dos marcas.
    b) ¿Es necesario pedir el cumplimiento de alguna otra condición?
    c) Si se conoce que el desvío estándar poblacional del contenido de grasa de la primera marca es de $2{,}8$ g, mientras que el de la segunda es de $2{,}3$ g, construir nuevamente el intervalo pedido en (a).

29. Se tomaron dos muestras aleatorias de 1000 individuos de Capital Federal y 1500 de la provincia de Buenos Aires y se las encuestó respecto a si acuerda con la construcción de bicisendas en su barrio o localidad. Como resultado de la encuesta, 400 de los individuos de Capital Federal manifestaron estar de acuerdo mientras que 900 de la provincia de Buenos Aires acuerdan con la medida.

a) Hallar un intervalo de confianza de nivel asintótico $95\,\%$ para la diferencia de proporciones de acuerdo con la bicisenda en las dos poblaciones.

b) ¿En cuánto se reduce la longitud del intervalo si se considera una confianza del $90\,\%$?

30. Se piensa que la concentración del ingrediente activo de un antibiótico , es afectada por el tipo de instrumental utilizado en el proceso de fabricación. Se sabe que la desviación estándar de la concentración del ingrediente activo es de 3 g/l, independientemente del tipo de instrumental empleado en su producción. Se realizan 10 observaciones de una muestra de un antibiótico con cada uno de los equipos disponibles, y se obtienen los siguientes datos:

Equipo 1: 57.9 - 66.2 - 65.4 - 65.4 - 65.2 - 62.6 - 67.6 - 63.7 - 67.2 - 71.0

Equipo 2: 65.4 - 70.7 - 70.1 - 69.3 - 63.8 - 68.6 - 68.6 - 69.4 - 65.3 - 68.8

a) Suponiendo normalidad de las distribuciones, encontrar un intervalo de confianza del $95\,\%$ para la diferencia entre los valores medios de las concentraciones logradas por cada uno de los equipos.

b) Considerar ahora la dispersión poblacional desconocida pero igual para ambos equipos.

31. Se desea estimar el porcentaje de vuelos demorados en Aeroparque. Las líneas de interés para el estudio son Golazo y LAP. En un determinado mes se seleccionan al azar 100 vuelos de la primera y 121 de la segunda. Resultaron 20 demoras en Golazo y 28 demoras en LAP. Se construyó un intervalo de confianza (LAP - Golazo) para la diferencia de proporción siendo su límite superior 0,1605. Hallar el nivel de confianza utilizado.

Respuestas

1. a) $\bar{x} = 63{,}35$, $\tilde{x} = 63$
   b) presencia de outlier/s
   c) $s^2 = 232{,}55$, $s = 15{,}25$
   d) eliminando el 91, $\bar{x} = 61{,}89$, $\tilde{x} = 61$, $s^2 = 200{,}77$, $s = 14{,}17$

2. a) Todos menos L_3 y L_5.
   b) Elegiría L_5 por ser el más preciso.

3. (a) 0.91317 (b) $n \geq 10$

4. a) $E(\bar{X}) = 20{,}38 = E(X)$, $V(\bar{X}) = 3{,}989 = \frac{1}{2}V(X)$
   b) $P(Z = 17{,}55) = P(Z = 21{,}65) = P(Z = 21{,}95) = \frac{1}{3}$
   c) $E(Z) = 20{,}38 = E(X) \quad V(Z) = 4{,}028 \neq \frac{1}{2}V(X)$

5. a) $\bar{x} = 1{,}35$,$\tilde{x} = 1{,}395$, $s^2 = 0{,}1107$, $s = 0{,}3327$
   b) $\hat{\mu} = \bar{x} = 1{,}35$

c) $\bar{x}$ por la Ley de los Grandes Números.

6. a) $E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ...E(X_n) = n\mu$
   b) $V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ...V(X_n) = n\sigma^2$ usando la condición de independencia de las observaciones
7. 0.46322
8. (a) 0.00001 (b) 0.87403 (c) 0.71341
9. (a) 0.00421 (b) $s^2 = 21{,}76$
10. 0.1351
11. (a) 0.81835 (b) 0.17846
12. (a) $\hat{\lambda}_{mom} = \hat{\lambda}_{mv} = \bar{x} = 2{,}09$ (b) s'í (c) 0.84053
13. a) $E(T_1) = E(T_3) = \mu$ insesgados, $E(T_2) = 2\mu$ sesgado
    b) $V(T_1) = \frac{5}{18\alpha^2}$, $V(T_3) = \frac{1}{4\alpha^2}$, $V(_3) < V(T_1)$
    c) $ECM(T_1) = V(T_1)$, $ECM(T_3) = V(T_3)$, sí es el mínimo.
14. $ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2$
    $ECM(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 + [E(\hat{\theta})]^2 - 2\theta E(\hat{\theta}) + \theta^2$
    $ECM(\hat{\theta}) = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$
    $ECM(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2$
15. $E(\bar{X}^2) - [E(\bar{X})]^2 = V(\bar{X}) = \dfrac{\sigma^2}{n} \Rightarrow E(\bar{X}^2) = \mu^2 + \dfrac{\sigma^2}{n} \neq \mu^2$
16. (a) $\hat{\theta} = \dfrac{1 - 2\bar{X}}{\bar{X} - 1}$ (b) $\hat{\theta} = 3{,}309$
17. a) $I_{90} = [250{,}925, 269{,}075]$, $I_{95} = [249{,}22, 270{,}78]$, $I_{99} = [245{,}81, 274{,}19]$
    b) $L_1 < L_2 < L_3$ cuando aumenta la confianza se pierde precisión
18. (a) $I_{98} = [167{,}22, 168{,}78]$ (b) $n \simeq 102$
19. (a) F (b) F (c) V (d) F (e) F
20. (a) $IC_{95} = [88{,}45, 89{,}75]$ (b) $IC_{95} = [88{,}34, 89{,}86]$
21. (a) $\bar{x} = 15{,}3$, $s = 0{,}62$ (b) 0.97
22. $IC_{95} = [261{,}09, 277{,}06]$
23. $IC_{95} = [2{,}28, 3{,}67]$
24. a) distribución Normal e independencia de las observaciones
    b) $IC_{99} = [3{,}602, 8{,}553]$

25. (a) $n \geq 4161$ (b) $IC_{99} = [0{,}331, 0{,}369]$

26. *a*) $IC_{\mu} = [7{,}856, 8{,}484]$
    b) $IC_{\sigma} = [1{,}23, 1{,}68]$
    c) $IC_{p} = [0{,}231, 0{,}436]$

27. *a*) $IC_{99} = [0{,}1863, 0{,}3051]$
    b) $IC_{95} = [0{,}0238, 0{,}0676]$
    c) $IC_{95} = [0{,}1581, 0{,}2419]$

28. *a*) $IC_{95} = [-4{,}183, 4{,}023]$
    b) independencia y varianzas iguales
    c) $IC_{95} = [-2{,}436, 2{,}276]$

29. *a*) $IC_{95} = [0{,}1608, 0{,}2392]$
    b) $IC_{90} = [0{,}135, 0{,}233]$ se reduce un $16\,\%$

30. *a*) $IC_{95} = [0{,}15, 5{,}41]$
    b) $IC_{95} = [0{,}02, 5{,}53]$

31. 0.98

Capítulo 5

Test de Hipótesis

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

- Hipótesis: científica y estadística, nula y alternativa.
- Tipos de error: nivel de significación. Potencia de la prueba.
- Pruebas para media y varianza de una población normal.
- Pruebas para el parámetro p de la Binomial.
- Valor p y su relación con el nivel de significación
- Diseño en función de la significación y la potencia de la prueba.

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 5

1. Dados los siguientes pares de aseveraciones, indicar cuál/es no cumple/n con las reglas para establecer hipótesis acerca del valor de la media poblacional μ en una prueba basada en una muestra aleatoria $\{X_1, X_2, ..., X_n\}$. Justificar la respuesta.

   a) $H_0 : \mu = 100$ vs. $H_1 : \mu = 200$
   b) $H_0 : \mu = 100$ vs. $H_1 : \mu < 100$
   c) $H_0 : \mu = 100$ vs. $H_1 : \mu \neq 100$
   d) $H_0 : \mu < 100$ vs. $H_1 : \mu \geq 100$
   e) $H_0 : \mu \leq 100$ vs. $H_1 : \mu > 100$
   f) $H_0 : \mu \neq 100$ vs. $H_1 : \mu = 100$

g) $H_0 : \bar{x} = 100$ vs. $H_1 : \bar{x} > 100$

Asociar los siguientes enunciados con las hipótesis que les correspondan.

a) Para determinar si las soldaduras efectuadas en tubos de una planta de energía nuclear cumplen con las especificaciones, se selecciona una muestra al azar de soldaduras y se realizan pruebas en cada soldadura de la muestra. La resistencia de la soldadura se mide como la fuerza requerida para romper la soldadura. Supongamos que en las especificaciones se establece que la resistencia media de soldaduras debe rebasar 100lb/pulg^2.

b) Interesa controlar si una balanza está correctamente calibrada, para ello se pesará una pesa patrón de 100 kg, veinte veces.

c) Se conduce una investigación para estudiar si el nivel de ventas promedio de una sucursal por día es inferior a 100 unidades.

d) ¿Qué significaría el error de tipo I en (a)?

e) ¿Qué implicaría el error de tipo II en (b)?

2. El estadístico de contraste de un ensayo de hipótesis, Z, tiene distribución Normal estándard indicar para cada una de las regiones de rechazo, las hipótesis correspondientes y el nivel de significación considerado:

   a) $\{Z > 2{,}33\}$

   b) $\{Z < -2{,}58\}$

   c) $\{Z < 1{,}96 \vee Z > -1{,}96\}$

3. Para cada uno de los siguientes pares de hipótesis, considerando una prueba cuyo estadístico de contraste tenga distribución normal estándar para los valores observados en cada caso, indicar el valor p de la prueba.

   a) $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$; $Z_{obs} = 2{,}5$

   b) $H_0 : \mu = 12$ vs. $H_1 : \mu < 12$; $Z_{obs} = -1{,}2$

   c) $H_0 : \mu = 7$ vs. $H_1 : \mu \neq 7$; $Z_{obs} = 2{,}8$

   ¿En qué casos rechazaría la hipótesis nula con un nivel de significación del $5\,\%$?

4. Dado el valor observado para una prueba basada en un estadístico t y sus grados de libertad (gl), aproximar el p valor y la decisión de una prueba con nivel de significación $\alpha = 0{,}05$:

   a) unilateral derecha, con gl $= 8$, $t_{obs} = 2{,}84$

   b) unilateral izquierda, con gl $= 10$, $t_{obs} = -2{,}5$

   c) bilateral, con gl $= 15$, $t_{obs} = 1{,}95$

   En alguno de los casos le parece que cambiaría la decisión si se considerara un nivel de significación con $\alpha = 0{,}01$.

5. Se dispone de una muestra de tamaño 1 para decidir entre dos hipótesis sobre la media μ de una distribución Exponencial. La hipótesis H afirma que $\mu \leq 2$ y la hipótesis K afirma que $\mu > 2$.

   a) Diseñar una regla de decisión que cumpla la condición que la máxima probabilidad de equivocarse al rechazar la hipótesis H cuando H es verdadera sea 0.1.

   b) Calcular la probabilidad de decidir en forma errónea cuando el verdadero valor de la media es $\mu = 2{,}1$.

6. Una empresa estudia introducir un nuevo sistema de producción para mejorar su productividad media establecida actualmente en 42 unidades por persona y por día. Se estima que el cambio no será rentable si no consigue elevar dicho número por encima de 45 u. Realizada una prueba con la nueva tecnología, aplicada a 35 personas, se obtuvo una producción media de 46.5 y no se observó ningún cambio apreciable en la dispersión que estaba establecida en $\sigma = 3$ u por día. Considerando esta información y utilizando un nivel de significación del $1\,\%$, ¿se debería efectuar el cambio tecnológico?.

7. Se desea calibrar un fotómetro para medir la proporción de monóxido de carbono (CO) de un gas. Se toman 25 mediciones de una muestra patrón con 60 ppm de proporción de CO y sabiendo que los errores de medición tienen distribución Normal con dispersión $\sigma = 0{,}5$ ppm. Designando con μ al verdadero valor medio de la calibración del fotómetro:

   a) ¿Qué hipótesis probaría?

   b) Si se rechaza la hipótesis de nulidad, para valores de $\bar{x}$ superiores a 60.196 ppm o inferiores a 59.804 ppm. ¿Cuál es la probabilidad de que la recalibración se realice sin ser necesaria?

   c) ¿Cuál es la probabilidad de no recalibrar si el fotómetro mide 60.3 ppm?

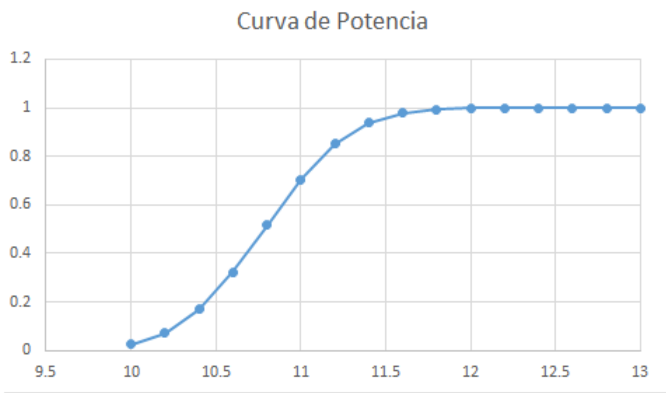
8. Un fabricante sostiene que el modelo de auto A, tiene un rendimiento promedio menor a 13 kilómetros por litro de nafta. Se selecciona una muestra de 9 de estos vehículos, y cada uno es conducido con un litro de nafta en las mismas condiciones. La muestra proporciona una media de 12.34 km/l. Se sabe que la distribución de la cantidad de kilómetros recorridos por litro de nafta es Normal con un desvío estándar de 1.26 km/l.

   a) Plantear las hipótesis a contrastar, el estadístico del test junto con su de distribución y la decisión tomada a nivel 0.05.

   b) ¿El test elegido es de nivel exacto o aproximado? Justificar.

   c) ¿Cuál es la probabilidad de no decidir que se cumple la hipótesis del fabricante, cuando el verdadero rendimiento medio del modelo de auto A es de 12 km/l.

9. (*) La siguiente gráfica corresponde a la curva de potencia de una prueba para la media de una población normal, conocida la varianza $\sigma^2 = 4$ basada en una muestra de tamaño $n = 25$:



a) Indicar qué hipótesis se están testeando.

b) Sabiendo que $\beta_{\mu=11} = 0{,}2946$, ¿cuál es el nivel de significación de la prueba?

c) Hallar la probabilidad del error de tipo II para $\mu = 11{,}2$, $\mu = 11{,}5$, $\mu = 12$.

10. El coeficiente de compresión del cemento curado sigue una distribución Normal con una media de 50 kg/cm^2 (en cientos). Utilizar un cemento de menor coeficiente promedio podría ocasionar problemas en la construcción por lo que se requiere tener una probabilidad de $0{,}95$ de descartar las partidas si el coeficiente de compresión promedio proviniera de una población con coeficiente de compresión promedio de 47 kg/cm^2 y una probabilidad de $0{,}01$ de rechazar las partidas que cumplan con la especificación. Supongamos que el desvío de la población es de 5 kg/cm^2.

    a) Plantear un test adecuado. Justificar las hipótesis elegidas.

    b) Calcular el tamaño de la muestra necesario para que se cumplan las condiciones requeridas y establecer la zona de rechazo y la regla de decisión.

11. Una compañía de productos para el consumidor está desarrollando un nuevo shampoo, y está interesada en la altura de la espuma (en mm). La altura de la espuma tiene una distribución aproximadamente Normal. La compañía quiere verificar si la altura de la espuma supera los 180 mm y si se considera una muestra de tamaño 24. Los valores observados son:

122.72	182.53	156.03	200.31	219.86	187.20	213.00	133.47
203.39	207.56	212.64	132.57	204.95	181.51	184.27	194.70
184.46	169.48	207.28	186.65	195.85	181.26	181.78	211.32

    a) Establecer las hipótesis correspondientes al test.

    b) Interpretar la salida del programa R. Concluir.

```
data: muestra
t = 0.81221, df = 23, p-value =
0.2123
alternative hypothesis: true mean
is greater than 180
sample estimates:
mean of x 185.62
```

12. La calidad de ciertas piezas de un aparato depende de su esfericidad. Para mejorar el control de calidad se determinó un nuevo método para medir el índice de esfericidad (IE) y se aplicó en 91 piezas.

Variable	n	Media	D.E.	W^*	p (una cola)
IE	91	1.45	0.22	0.93	0.2569

Tabla 5.1: Test de Shapiro Wilks

El método de medición anterior tiene un desvío estándar de 0.32 plg. Interesa determinar si el nuevo método es más preciso, es decir tiene menor desvío estándar.

a) Analizar el cumplimiento de los supuestos para realizar un test y usar un nivel de significación del $5\,\%$.

b) Realizar la prueba y concluir en el contexto del problema.

13. Un capataz afirma que bajo su supervisión el armado de una pieza tiene una duración promedio inferior a 25 minutos y la dispersión de 2 minutos. El tiempo medio arrojado por una muestra de 25 obreros con este nuevo proceso resultó de 23 minutos y el desvío estándard muestral es de 3.9 minutos, si los tiempos se distribuyen en forma aproximadamente Normal.

a) Plantear un test para decidir si la media y la variabilidad del tiempo del armado de una pieza coinciden con la afirmación del capataz.

b) Concluir en función de los datos muestrales, considerando un nivel de significación del $5\,\%$.

14. El fabricante de neumáticos *Mirelli* sostiene que en promedio sus productos tienen una duración superior a 50.000 km antes de necesitar reemplazo. Un grupo de consumidores quiere poner a prueba dicha afirmación. Para ello seleccionan una muestra de 50 neumáticos y resulta que en promedio duraron 50.300 km con una dispersión de 1000 km.

a) Definir con claridad cuál sería el parámetro de interés en este problema.

b) Establecer la hipótesis nula y la alternativa en términos de dicho parámetro.

c) Explicar el significado de los errores de tipo I y II en el contexto de este problema.

d) Responder si la afirmación del fabricante es aceptable con un nivel de significación del $0{,}01$.

15. Dos empresas distintas compiten para brindar servicio de internet en cierta región. Denotamos con p a la proporción de pobladores que prefieren a la primera de las empresas por encima de la segunda. Se toma una muestra aleatoria de 50 personas, para probar la hipotesis de que la primera empresa tiene conquistada más de la mitad del mercado.

    a) Considerando que X es la cantidad de individuos encuestados que optaron por la primer empresa, ¿cuál de las siguientes regiones de rechazo le parece más adecuada?
    - $H_1 : \{x \geq 32\}$
    - $H_1 : \{x \leq 18\}$
    - $H_1 : \{x \leq 15 \vee x \geq 45\}$

    b) ¿Cuál es la distribución de probabilidad de X siendo cierta H_0?

    c) ¿Qué se puede concluir, al 1 %, si 12 personas estuvieron a favor de la segunda empresa? ¿Cuál es el p valor de la prueba?

    d) ¿Cuál es la probabilidad de cometer error de tipo II si $p = 0{,}6$?

16. Una empresa de cosmética afirma que una nueva crema reduce las manchas del cutis tras un uso prolongado (3 meses o más) en más de un $60\,\%$. A fin de verificar la veracidad de esta afirmación, un centro de belleza que pretende adquirir este producto, utilizará el mismo en 120 manchas de cutis de clientas elegidas al azar. Si se eliminan más de 82 manchas tomará la decisión de adquirir el nuevo producto.

    a) Evaluar el nivel de significación utilizado.

    b) Hallar el error de tipo II que se cometería si el producto lograra limpiar el $75\,\%$ de las manchas.

17. Se supone que el $10\,\%$ de los consumidores de cierta localidad prefiere la marca A de café. Se realizó una campaña publicitaria y después de la misma se entrevistó a 100 habitantes de esta localidad para determinar la efectividad de la campaña. El resultado de esta encuesta mostró que 13 encuestados prefieren la marca A.

    a) ¿Existe evidencia a nivel aproximado 0.05, de un aumento en la preferencia por la marca A?

    b) Calcular el valor p.

    c) ¿Cuál es la probabilidad aproximada de decidir que la campaña publicitaria no fue efectiva, cuando en realidad la proporción de preferencia por la marca A después de la campaña es 0.2?

    d) ¿Qué tamaño muestral le parece apropiado para que la probabilidad de (c) fuese a lo sumo 0.05?

18. Las piezas producidas por dos máquinas se encuentran en el depósito de un taller mecánico. Se sabe que el diámetro de estas piezas es una variable aleatoria Normal con parámetros conocidos $\mu_1 = 10{,}2$ mm y $\sigma_1 = 0{,}45$ mm para la máquina 1, y $\mu_2 = 10{,}4$ mm y $\sigma_2 = 0{,}5$ mm para la máquina 2. Se ha recibido un trabajo que requiere de 2500 de estas piezas y se preparó el material de trabajo con todas la piezas producidas por una sola de estas dos máquinas, pero

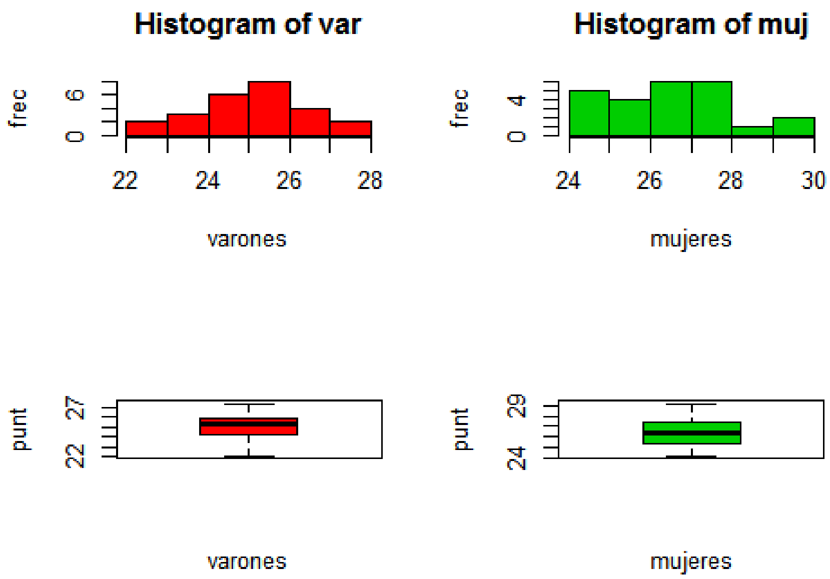
no se sabe cuál. Para decidir si es de una u otra, se realiza una prueba. Diseñar la prueba de tal manera que la probabilidad de decidir que las piezas corresponden a una máquina cuando en realidad no es así, a partir de los valores medios de las muestras, sea la misma para ambas máquinas.

19. Un productor está interesado en reducir el tiempo de secado de una pintura impermeabilizante. Se prueban dos fórmulas de pintura, la fórmula A tiene el contenido químico tradicional, mientras que la fórmula B tiene un nuevo ingrediente que reduce el tiempo de secado. De la experiencia se sabe que la desviación estándar del tiempo de secado es de 8 minutos, para ambas pinturas y que la distribución es normal. Se pintan diez especímenes con cada una de estas fórmulas resultando las medias muestrales de 121 minutos y 112 minutos respectivamente.

    a) ¿A qué conclusión puede llegar el productor sobre la eficacia del nuevo ingrediente considerando un nivel de significación del $5\,\%$?

    b) Si la verdadera diferencia de tiempos promedio de secado fuera de 10 minutos, ¿cuál sería el tamaño muestral requerido (suponiendo ambas muestras del mismo tamaño) para detectar dicha diferencia con una probabilidad igual o mayor al 0.90?

20. En un estudio sobre el desarrollo psicomotriz de niños que cursan el prescolar se seleccionaron dos muestras aleatorias de 24 niños y 25 niñas de 5 años de edad. Entre otras pruebas se realizó un test que evalúa habilidades psicomotoras. Los investigadores estaban interesados en comparar las medias de los puntajes en las poblaciones de varones y mujeres de 5 años. En el siguiente gráfico se aprecian los histogramas y boxplots de ambas muestras. Se sabe que la dispersión de los puntajes de este test en cualquier población es de 1.5 puntos y se pueden considerar normales.



    a) La media de los varones resultó $25{,}04$ y la de las mujeres $26{,}40$, indicar si las diferencias de rendimiento pueden considerarse significativas al $5\,\%$.

b) ¿Se podría haber inferido lo mismo de la observación de los graficos?

21. En un laboratorio de especialidades medicinales están comparando la efectividad de dos medicamentos A y B para el tratamiento de la mastitis. Para ello se trataron dos grupos de vacas afectadas con síntomas similares. Se registró el tiempo en días que tardaron en desaparecer los síntomas en cada caso. Si las distribuciones de estos tiempos son normales con varianza común:

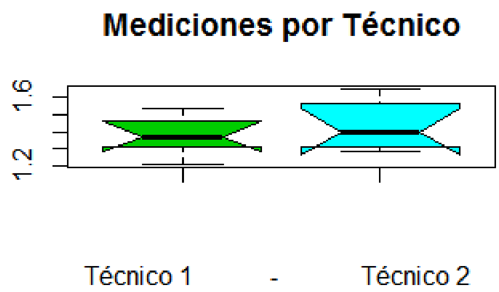
A	5	7	6	7	7	8	6	8	7	6
B	7	9	10	8	6	5	7	9	9	5

a) Plantear las hipótesis a contrastar, el estadístico del test junto con su función de distribución, la región de rechazo y la decisión tomada nivel aproximado 0.05.

b) ¿Cuál es el p-valor de la prueba?

22. Dos técnicos de control de calidad miden el acabado de la superficie de una pieza metálica, y obtienen los siguientes resultados:

Técnico 1	1.45	1.37	1.21	1.54	1.48	1.29	1.34	
Técnico 2	1.55	1.43	1.58	1.37	1.29	1.34	1.28	1.65



a) ¿Qué observa en el boxplot paralelo respecto de las posiciones centrales de los grupos?

b) Se aplicó el test de normalidad de Shapiro Wilks a ambas muestras obteniéndose respectivamente los siguentes p valores 0.966 y 0.368. Determinar si existe diferencia significativa entre los valores medios de las mediciones obtenidas con $\alpha = 0{,}05$ sabiendo que las mediciones de ambos técnicos tienen igual dispersión.

23. (T) Decidir si son verdaderas o falsas las siguientes afirmaciones y justificar adecuadamente cada respuesta.

a) Si el p-valor para ciertas hipótesis estadísticas es igual a $0{,}08$, indica que se rechaza la hipótesis nula a nivel $5\,\%$.

Tabla 5.2: Estadística descriptiva

Técnico	media	sd	n
Técnico 1	1.383	0.115	7
Técnico 1	1.436	0.141	8

b) Al aplicar cualquier prueba de hipótesis se verifica que: $P(\text{error tipo I}) = 1 - P(\text{error tipo II})$.

c) Se denomina potencia de una prueba al riesgo que voluntariamente asume el investigador, de equivocarse al rechazar la hipótesis de nulidad cuando en realidad esta es cierta.

d) Si se realiza dos veces un ensayo de hipótesis, la conclusión será la misma en ambos casos independientemente de la muestra seleccionada.

e) El nivel de significación de un test es la probabilidad de que la hipótesis nula sea cierta.

f) Si el p-valor es 0.3, el test correspondiente no rechaza al nivel 0.01.

g) Si un intervalo de confianza de nivel 0.99 para la media poblacional de una distribución Normal calculado a partir de una muestra da como resultado [-2, 3], entonces el test para las hipótesis $H_0 : \mu = -3$ vs. $H_1 : \mu \neq -3$ basado en los mismos datos no rechazaría la hipótesis de nulidad con nivel 0.01.

24. Una muestra aleatoria de 250 automóviles que circulan por la ciudad y que tienen más de 10 años o más de uso indica que 98 de ellos no pasan un chequeo básico de su sistema de frenos mientras que en una muestra de 300 autos con menos de 10 años de antigüedad, 79 no pasan un chequeo básico de su sistema de frenos.

    a) ¿Existe evidencia significativa, al nivel 1 %, de que la proporción de autos con desgaste en su sistema de frenos de 10 o más años de uso es mayor que esta proporción en autos con menos de 10 años de uso?

    b) Hallar el valor p de la prueba e interpretarlo.

    c) ¿El nivel de esta prueba es exacto o asintótico?

Respuestas

1. (d), (f) y (g) no cumplen

    a) Asociado a (e)

    b) Asociado a (c)

    c) Asociado a (b)

    d) Decir que cumple la especificación cuando no es así.

    e) No detectar que la balanza está descalibrada.

2. *a*) $\alpha = 0{,}01$, $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$

    b) $\alpha = 0{,}005$ $H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$

c) $\alpha = 0{,}05$ $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

3. a) p valor $= P(Z > 2{,}5) = 0{,}00621 \Rightarrow$ se rechaza H_0
   b) p valor $= P(Z < -1{,}2) = 0{,}1157 \Rightarrow$ no se rechaza H_0
   c) p valor $= 2P(Z > 2{,}8) = 0{,}00512 \Rightarrow$ se rechaza H_0

4. a) p-valor= 0.0109
   b) p-valor= 0.0157
   c) p-valor= 0.0701

   No.

5. a) Se rechaza H_0 si $x_{obs} > 4{,}6$
   b) $0{,}8884$

6. Se rechaza H_0; es decir, se debe implementar el cambio tecnológico.

7. (a) $H_0 : \mu = 60$ vs. $H_1 : \mu \neq 60$ (b) $\alpha = 0{,}05$ (c) 0.1492

8. a) $H_0 : \mu \geq 13$ vs. $H_1 : \mu < 13$; $\bar{x} \sim N(13; 0{,}42)$, no se rechaza H_0 al $5\,\%$
   b) Exacto, pues se conoce la distribución exacta del estadístico.
   c) 0.231

9. a) $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$
   b) $\alpha = 0{,}025$
   c) $\beta_{11,2} = 0{,}1492$, $\beta_{11,5} = 0{,}03$, $\beta_{12} = 0{,}0012$

10. a) $H_0 : \mu = 50$ vs $H_1 : \mu < 50$. Se rechaza H_0 si $\bar{x} < VC = 50 - 2{,}33 * \frac{5}{\sqrt{n}}$.
    b) $n = 44$

11. a) $H_0 : \mu \leq 180$ vs. $H_1 : \mu > 180$
    b) No se rechaza H_0, $\bar{x} = 184{,}3288$, p-valor $= 0{,}2123 > \alpha = 0{,}05$. No existe evidencia empírica a favor de que la altura mínima supere los 180 mm.

12. a) $H_0 : \sigma^2 = 0{,}32^2$ vs. $H_1 : \sigma^2 < 0{,}32^2$. El test de Shapiro Wilks no rechaza la normalidad, luego puede aplicarse la prueba basada en el estadístico $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}$.
    b) $\chi_{obs} = 42{,}539 < 69{,}126 = \chi^2_{90,0,95}$. Se rechaza H_0, con lo cual hay evidencia a favor de que el nuevo método es más preciso.

13. a) $T_{obs} = -2{,}56 < T_{crit} = -1{,}71$ entonces se rechaza H_0 y la afirmación de la media es correcta
    b) $\chi_{obs} = 91{,}26 > 39{,}36 = \chi_{crit}$ entonces se rechaza H_0 y la afirmación de la varianza no es correcta

14. a) μ: "Verdadero valor medio de la duración de los neumáticos *Mirelli* antes de requerir reemplazo"

b) $H_0 : \mu \leq 50000$ vs. $H_1 : \mu > 50000$

c) Error tipo I: confiar en que la duración es mayor a 50000 cuando en realidad no lo es
Error tipo II: no detectar que la duración es superior a 50000

d) $T_{obs} = 2{,}12 < T_{crit} = 2{,}405$ entonces no se rechaza H_0 al $1\,\%$, no hay evidencia a favor de que la media supera los 50000

15. a) $\{X \geq 32\}$

    b) $X \sim Bi(50; 0{,}5)$ bajo $H_0 : p = 0{,}5$

    c) $Z_{obs} = 3{,}76$, $p-$ valor $< 0{,}0001$, la conclusión es que se rechaza $H_0 : p = 0{,}5$

    d) 0.718

16. (a) $\alpha = 0{,}03144$ (b) $\beta_{0{,}75} = 0{,}04551$

17. (a) no se rechaza H_0 (b) p-valor=0.1587 (c) $\beta_{0{,}2} = 0{,}1038$ (d) $n \geq 167$

18. $VC = 10{,}29$, si $H_0 = \mu = 10{,}2$, se rechaza H_0 para valores superiores a VC.

19. a) $z_{obs} = -2{,}52 < -1{,}65 = Z_{crit}$ entonces se rechaza H_0 al $5\,\%$ y hay evidencia a favor de la eficacia del nuevo ingrediente

    b) $n \geq 11$

20. a) Sí, las diferencias son significativas al $5\,\%$.

    b) No, la observación de gráficos no constituye evidencia estadística.

21. a) $H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$, estadístico: $\dfrac{\bar{x}_A - \bar{x}_B}{s_P\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim T_{n_A+n_B-2}$ $T_{obs} = 1{,}25 < T_{crit} = 2{,}1$, entonces no se rechaza H_0

    b) p-valor = 0.2273

22. a) La media del Técnico 2 parece superior a la del Técnico 1.

    b) No se rechaza H_0, no hay evidencia empírica a favor de la suposición realizada a partir de la observación.

23. (a) F (b) F (c) F (d) F (e) F (f) V (g) F

24. (a) Hay evidencia empírica al 1 % de significación. (b) 0.0012 (c) aproximado o asintótico.

Capítulo 6

Regresión y Correlación

Contenidos

Los contenidos que aparecen resaltados, incluyen deducción o demostración.

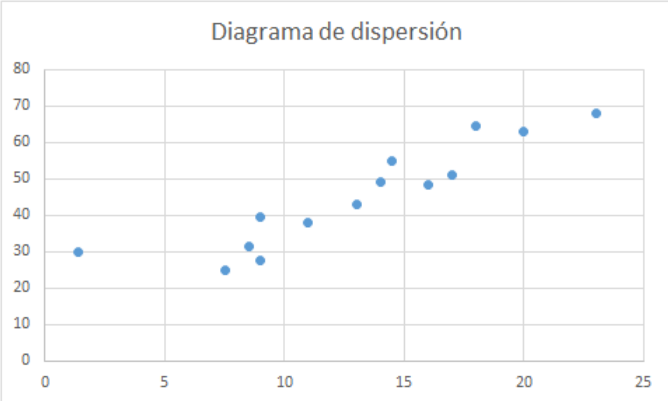
- El modelo de regresión lineal simple: supuestos, condiciones y parámetros.
- Estimadores de cuadrados mínimos de los parámetros de la regresión .
- El coeficiente de determinación. Interpretación.
- Estimación del coeficiente de correlación. Prueba de significación de la regresión.
- Intervalos de confianza para el valor esperado de la respuesta correspondiente a una nueva observación e intervalos de predicción.
- Análisis diagnóstico.

Notación:
(*) ejercicios que requieren aplicación informática o de celular.
(T) ejercicios teóricos.

Práctica 6

1. Se realiza un estudio para establecer una ecuación mediante la cual se pueda utilizar la concentración de cobre de una pieza (X) para predecir el porcentaje de níquel de la misma (Y). Se extrajeron al azar los siguientes datos de 14 piezas fabricadas en una misma planta durante el mes pasado:

X	1.4	7.5	8.5	9	9	11	13	14	14.5	16	17	18	20	23
Y	30	25	31.5	27.5	39.5	38	43	49	55	48.5	51	64.5	63	68




```
Análisis de regresión lineal
Variable      N     R²    R² Aj
Y            14    0.84    0.82

Coeficientes de regresión y estadísticos asociados
Coef   Est.  E.E.  LI(95%)    LS(95%)
const 15.85 4.09      6.95      24.76
X      2.26 0.29      1.63       2.89

Cuadro de Análisis de la Varianza (SC tipo III)
 F.V.           SC       gl       CM
Modelo.     2167.05       1    2167.05
X           2167.05       1    2167.05
Error        426.32      12      35.53
Total       2593.38      13
```

Sabiendo que X e Y tienen distribución Normal:

a) Graficar sobre el diagrama de dispersión la recta de regresión y señalar un par ordenado observado y un residuo.

b) Verificar gráfica y analíticamente que la recta de regresión pasa por el punto: $(\bar{x}, \bar{y})$.

c) Indicar a partir de la salida de computadora, los parámetros estimados para el modelo y el porcentaje de variabilidad de Y que logra explicar X a través de la recta de regresión.

2. En una planta química, se sospecha que la cantidad de vapor utilizada por mes (Y en miles de libras) está relacionada con la temperatura ambiente promedio del mismo mes (X en °F). La siguiente tabla presenta, para todo un año, el uso de vapor y la temperatura del mes correspondiente:

X	21	24	32	47	50	59	68	74	62	50	41	30
Y	240	220	320	325	280	230	200	275	270	250	200	270

Shapiro-Wilks (modificado)

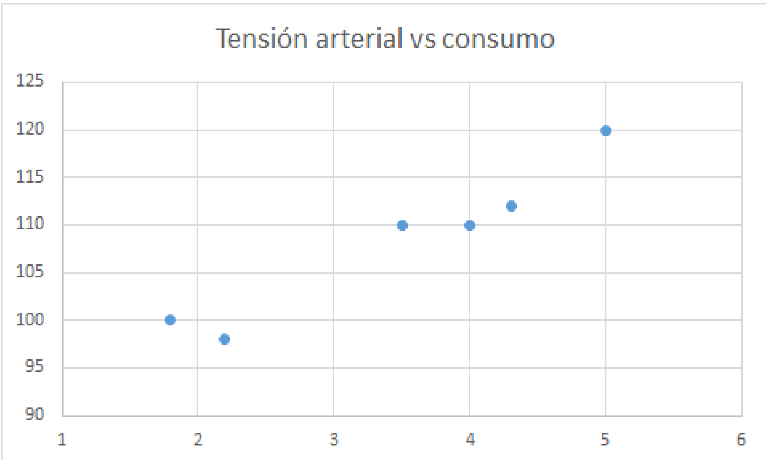
Variable	n	Media	D.E.	W^*	p (Unilateral D)
X	12	46.50	17.34	0.93	0.5413
Y	12	256.67	41.25	0.92	0.4005

a) Indicar las hipótesis a testear para responder a la sospecha de la planta química indicando el estadístico de contraste y su distribución.

b) Indicar si se satisfacen los supuestos para realizar esta prueba (para el test de Shapiro Wilks utilizar un nivel del 0.10) y, en caso afirmativo, realizarla y concluir en el contexto del problema con un nivel del $0{,}05$.

3. Se quiere estudiar la asociación lineal entre consumo diario de sal y tensión arterial. A una serie de voluntarios se les interroga sobre la dosis de sal en su dieta y se mide simultáneamente su tensión arterial. Los datos están en la tabla:

Consumo de sal (mg)	$1{,}8$	$2{,}2$	$3{,}5$	$4{,}0$	$4{,}3$	5
Tensión arterial (mm de Hg)	100	98	110	110	112	120



Si se sabe que ambas variables tienen distribución Normal conjunta,

a) de la observación del diagrama de dispersión, ¿le parece razonable testear la hipótesis de asociación lineal? Justificar.

b) ¿qué test correspondería aplicar en caso afirmativo?

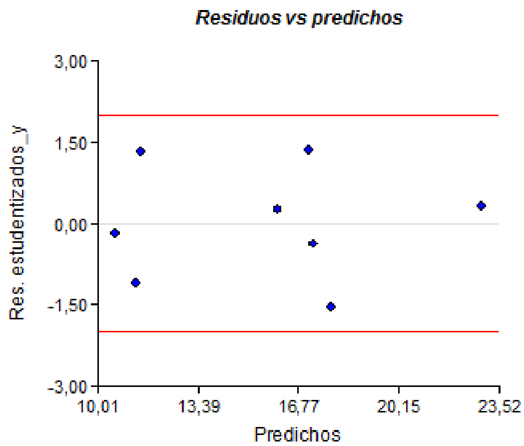
c) Establecer las hipótesis, indicar el estadístico de contraste, la región crítica y la decisión al $5\,\%$.

4. Los siguientes datos proporcionan información acerca del contenido de agua nieve (X) en cm y la afluencia de abril a junio en pulgadas (Y) en la cuenca del río Salado. Se sabe que ambas variables tienen distribución Normal.

Shapiro-Wilks (modificado)

X	23.1	32.8	31.8	32	30.4	24	39.5	24.2
Y	10.5	16.7	18.2	17	16.3	10.5	23.1	12.4

Variable	n	Media	D.E.	W^*	p-valor
RESIDUO	8	0.00	0.78	0.94	0.6941



a) Ajustar el modelo de regresión lineal simple para estimar Y en función de X, basándose en los datos disponibles, e interpretar los coeficientes obtenidos.

b) ¿Se satisface el supuesto de normalidad de los errores para este modelo? Utilizar un $10\,\%$ de significación. Justificar.

c) ¿Le parece que se satisface el supuesto de homocedasticidad de los errores para este modelo? (Observe en el gráfico si se presenta alguna estructura, es decir si al aumentar el valor de la variable predictora aumenta la variabilidad de los residuos, o dismimuye) o bien si se detecta residuos muy altos en valor absoluto.

d) En caso de haber respondido afirmativamente a las últimas dos preguntas, testear la significación del modelo con nivel 0.05.

5. Una compañía productora de energía eléctrica está interesada en desarrollar un modelo que relacione la demanda máxima por hora (D, en kw) con el uso de la energía total al mes (U, en kwh) las variables se distribuyen en forma Normal bivariada.

   La siguiente tabla muestra los datos obtenidos de una muestra de 11 clientes elegidos al azar entre los de esta compañía:

U	679	292	1012	493	582	1156	997	2189	1087	2077	2078
D	0.79	0.44	0.56	0.79	2.7	3.64	4.73	9.5	5.34	6.88	6.85

a) Ajustar al modelo de regresión lineal simple a los datos e indicar si el ajuste es adecuado con un nivel de significación del $1\,\%$.

b) Encontrar un intervalo de predicción para un $u_0 = 900$ kwh con un nivel del $99\,\%$. Interpretarlo e indicar la diferencia con un intervalo de confianza para la media con el mismo valor de u_0.

c) ¿Puede estimarse a partir de estos datos el valor esperado de D para $U = 156$?

6. Se realizó un estudio para estimar los efectos producidos en las personas debidos a la exposición al ruido. Ocho personas participaron en este estudio. Se registraron los aumentos en la presión sanguínea del individuo y el nivel de presión sonora, distribuido normalmente, a que se sometieron. Los datos se expresan en la tabla:

X: Nivel de presión sonora (Db)	60	63	70	80	85	90	94	100
Y: Aumento presión sanguínea (Hg)	1	0	4	3	5	8	7	6

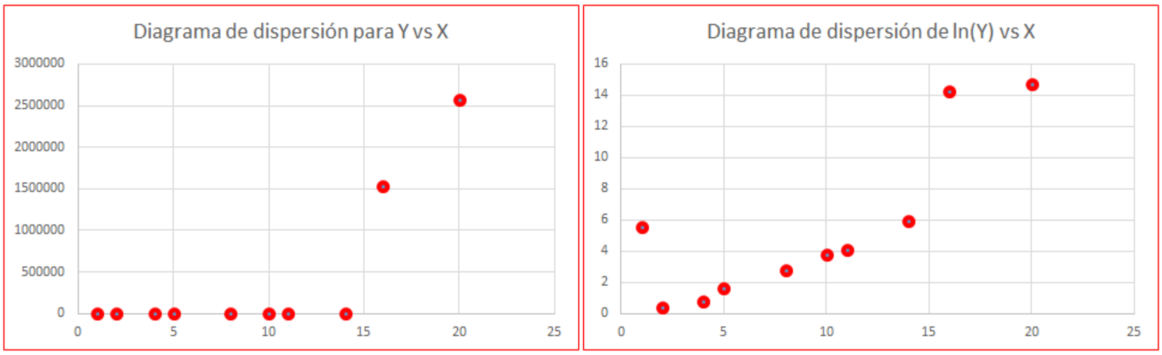
$$\sum_{i=1}^{n} x_i = 642 \qquad \sum_{i=1}^{n} y_i = 34 \qquad \sum_{i=1}^{n} x_i^2 = 53030 \qquad \sum_{i=1}^{n} y_i^2 = 200 \qquad \sum_{i=1}^{n} x_i y_i = 2983$$

a) Hallar la recta de regresión estimada.

b) Hallar el coeficiente de determinación e interpretarlo.

7. Una empresa de servicios informáticos está preocupada por la proliferación de cierto tipo de virus. La siguiente tabla registra el número de días que han transcurrido desde que se ha detectado un nuevo virus (X) y el número de ordenadores (Y) que han requerido los servicios de esta empresa a causa del virus.

x_i	y_i	$\ln(y_i)$
1	254	5.54
2	1.5	0.41
4	2.105	0.74
5	5.05	1.62
8	16.2	2.79
10	45.32	3.81
11	58.57	4.07
14	375.8	5.93
16	1525640	14.24
20	2577000	14.76



a) A partir de los dos diagramas de dispersión presentados, decidir cuál es la regresión lineal más adecuada y estimar los coeficientes correspondientes.

b) Estimar puntualmente el número de consultas que tendrá la empresa a los 15 días de la detección del virus.

8. *a*) (T) Deducir los estimadores de mínimos cuadrados de los coeficientes del modelo de regresión lineal para el caso general.

   b) (T) Existen aplicaciones importantes en las que, debido a restricciones científicas conocidas, la recta de regresión debe pasar por el origen de coordenadas, es decir el $(0,0)$ debe pertenecer a la recta. Por ello, el modelo tan sólo se requiere estimar un parámetro (Modelo de regresión al origen). Demostrar que el estimador de mínimos cuadrados para la pendiente es:

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

9. (T) Demostrar que la suma de cuadrados totales es igual a la adición de la suma de cuadrados residuales y la suma de cuadrados de la regresión; es decir,

$$\sum_{i=i}^{n} (Y_i - \bar{Y})^2 = \sum_{i=i}^{n} \left(Y_i - \hat{Y}_i\right)^2 + \sum_{i=i}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

10. Con respecto a los resultados de un análisis regresión lineal simple, ¿cuál de las afirmaciones siguientes puede ser cierta? (Justificar adecuadamente.)

    - $SC_{Total} < SC_{error} + SC_{regres}$
    - $R^2 = -0{,}65$
    - $s_b = -1{,}25$
    - $t_{obs} = -2{,}45$

11. En varios estudios se ha demostrado que los líquenes, organismos compuestos por un alga y un hongo, son buenos indicadores biológicos de la contaminación del aire. Los siguientes datos

X	0.05	0.1	0.11	0.12	0.31	0.37	0.42	0.58	0.68	0.68	0.73	0.85	0.92
Y	0.48	0.55	0.48	0.5	0.58	0.52	1.02	0.86	0.86	1	0.88	1.04	1.7

corresponden a las variables X (depósito de nitratos en g/cm^2) e Y (nitrógeno en el liquen, en % de peso seco):

Análisis de regresión lineal

Variable	N	R²	R² Aj
Y	13	0.72	0.69

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T
const	0.37	0.10	0.15	0.58	3.69
X	0.97	0.18	0.56	1.37	5.29

a) Indicar el modelo estimado para estos datos por mínimos cuadrados.

b) Completar los p valores de las salidas e indicar la significación del modelo.

c) Construir un intervalo de confianza de nivel 95 % para el valor esperado de nitrógeno en el líquen para un depósito de nitrato de 0.4 g/cm^2.

12. En la siguiente tabla se han registrado las superficies y los pesos de ciertas piezas metálicas producidas por una máquina.

SUPERFICIE DE LA PIEZA	44.09	36.67	51.72	36.04	38.97	61.4	42.06	53.33	50.01
PESO DE LA PIEZA	47.6	45	43.04	56.8	63.11	68.7	41.8	34.3	31.2

a) Graficar el diagrama de dispersión aproximado de Peso vs. Superficie. ¿Hay alguna observación que le llame la atención?

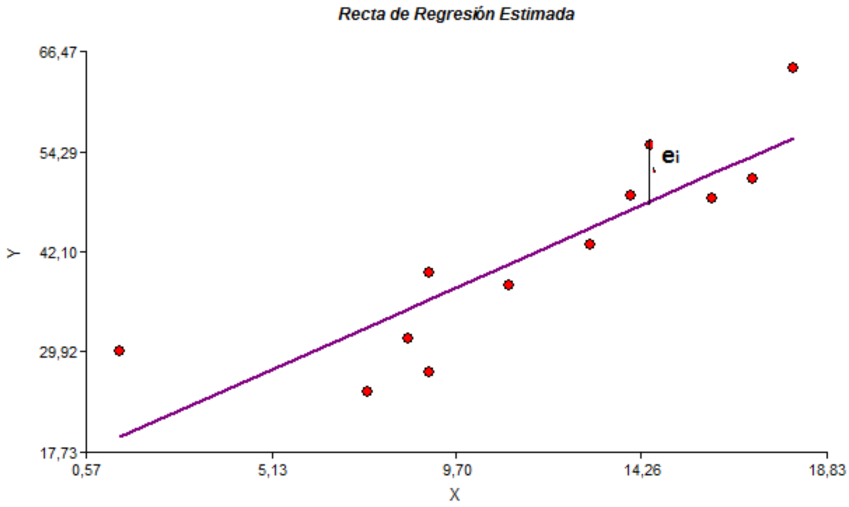
b) Estimar los coeficientes del modelo de regresión lineal utilizando toda la información disponible.

c) Calcular los residuos. ¿Hay alguno que le parezca especial?.

d) Eliminar la sexta observación y repetir los dos ítems anteriores. Comparar los resultados.

Respuestas

1. *a*) Gráfico



b) $\bar{x} = 13$, $\bar{y} = 45{,}25$, $\hat{y}_i = 15{,}85 + 2{,}263x_i$, se verifica que $\bar{y} = 15{,}85 + 2{,}263\bar{x}$

c) $\hat{\alpha} = 15{,}85$, $\hat{\beta} = 2{,}26$, $\hat{\sigma}^2 = 35{,}53$, $R^2 = 0{,}84$

2. a) $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$, $T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$

   b) Sí, pues puede suponerse distribución Normal conjunta. No se rechaza la hipótesis de nulidad al $5\,\%$, no hay evidencia en favor de la asociación lineal de las variables X (temperatura) e Y (vapor).

3. a) Sí.

   b) Test de correlación lineal.

   c) $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$, $T^{obs} = 7{,}54 > t_{crit} = 2{,}77$ entonces se rechaza la hipótesis nula

4. a) $\hat{y}_i = -6{,}68 + 0{,}749x_i$, $\hat{\alpha} = -6{,}68$ se interpreta como el valor medio estimado de y cuando $x = 0$; $\hat{\beta} = 0{,}749$ se interpreta como el aumento medio estimado de Y cuando X aumenta en una unidad.

   b) Sí, por el p valor del test de Shapiro Wilks.

   c) Sí, no se aprecia en el gráfico estructura ni creciente ni decreciente de la variabilidad.

   d) $T_{obs} = 13{,}24 > T_{crit} = 2{,}4469$ entonces se rechaza la hipótesis nula; es decir, que la regresión es estadísticamente significativa

5. a) $\hat{d}_i = -0{,}881 + 0{,}0041u_i$, $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$, $T_{obs} = 6{,}06 > T_{crit} = 3{,}24$ entonces la regresión es significativa al $1\,\%$

   b) $I_{pred} = [-2{,}1, 7{,}7]$ siempre tiene mayor longitud que el intervalo de confianza.

   c) No, pues está alejado del recorrido de U.

6. a) $\hat{y}_i = -9{,}28 + 0{,}1686x_i$

   b) $R^2 = 0{,}7731$ indica que el $77{,}31\,\%$ de la variabilidad de Y queda explicado por la variabilidad de X a través de la recta de regresión.

7. (a) la de $\ln(Y)$ vs. X (b) $\hat{y}(15) = 11488{,}33$

8. *a)* $l(X_1, X_2, ..., X_n) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2}\sum_{i=1}^{n}(Y_i - a - bx_i)2$

   Luego, el logaritmo de la función de verosimilitud difiere en una constante de la función que minimiza el método de mínimos cuadrados, y así $\hat{\alpha}_{MV} = \hat{\alpha}_{MC}$ y $\hat{\beta}_{MV} = \hat{\beta}_{MC}$.

   b) $\hat{b} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

9. $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$. Luego debemos probar que $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y}) = 0$. Considerando que $\hat{Y}_i = a + bXi$ y la segunda ecuación normal, queda demostrado.

10. La última.

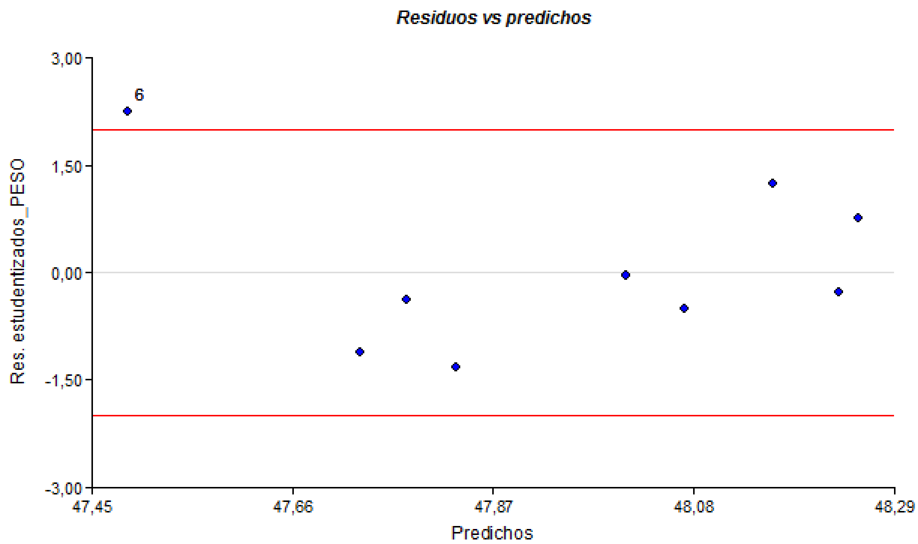
11. *a)* $\hat{y}_i = 0{,}37 + 0{,}97x_i$

    b) $p - \text{valor}_1 = 0{,}0036$, $p - \text{valor}_2 = 0{,}0003$

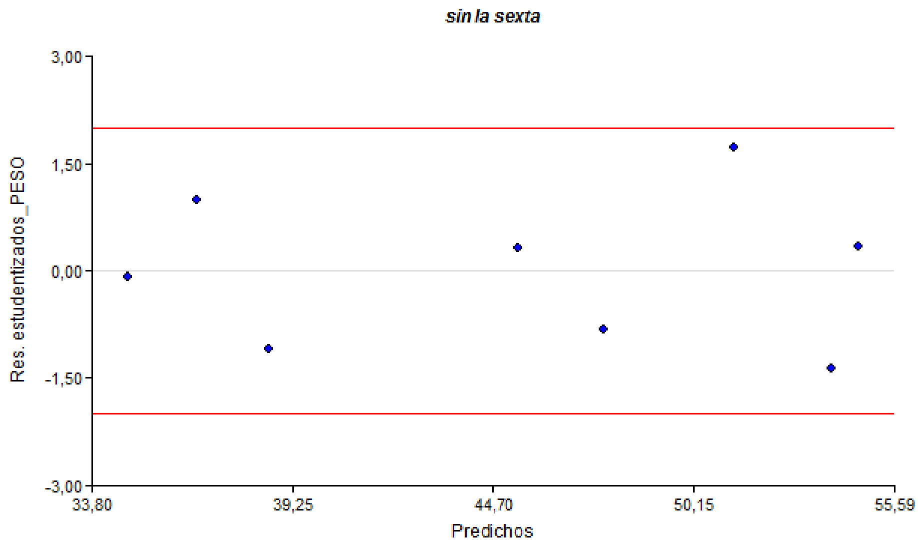
    c) $IC = [0{,}536, 0{,}98]$

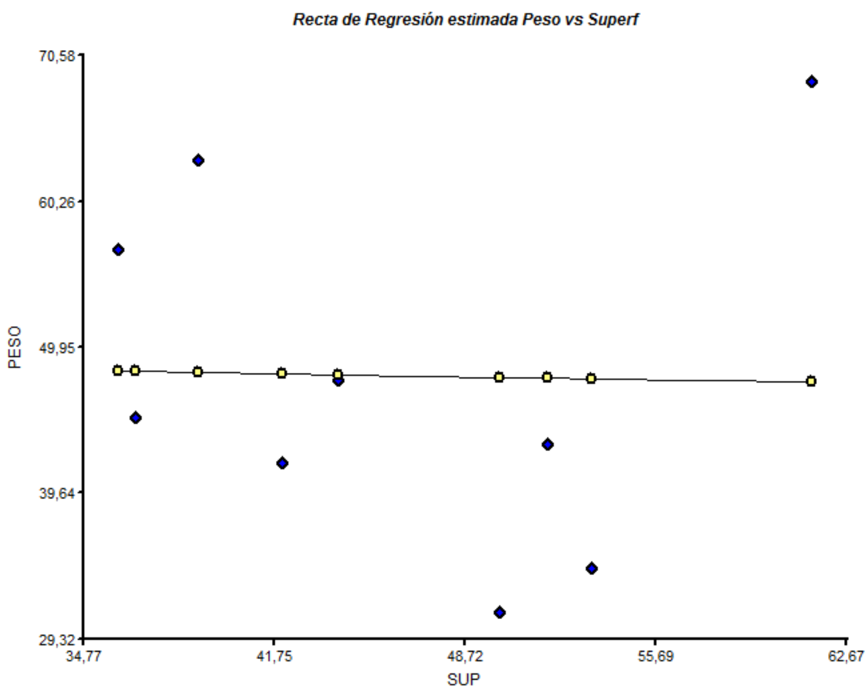
12. *a)* $\hat{\alpha} = 49{,}35$, $\hat{\beta} = -0{,}03$

    b) La sexta observación

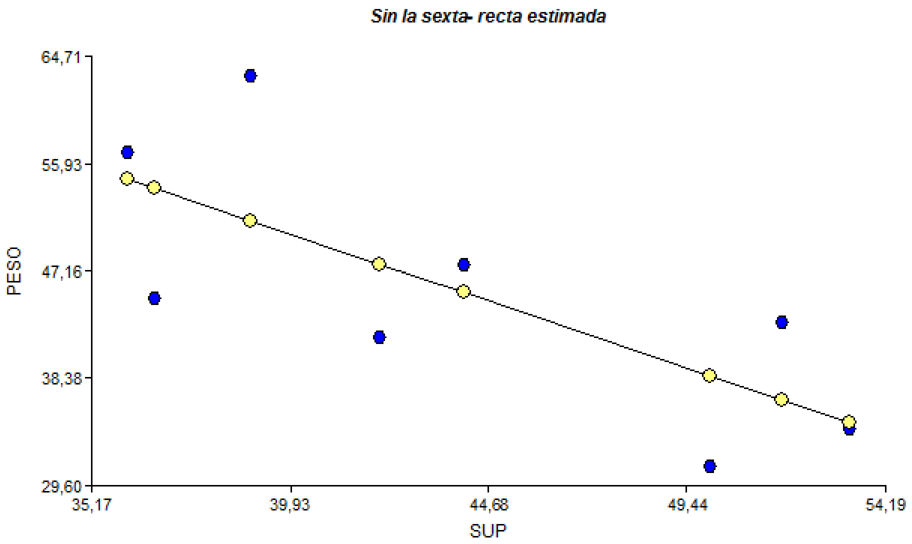


    c) El sexto residuo





d) $Y_i = 95{,}9 - 1{,}146X_i$. Cambian fundamentalmente los coeficientes


Sin la sexta- recta estimada
64,71
55,93
47,16
38,38
29,60
PESO
35,17
39,93
44,68
49,44
54,19
SUP

Capítulo 7

Modelos de Exámenes

7.1. Modelo para el Primer Parcial

Universidad Tecnlógica Nacional Facultad Regional Buenos Aires
Primer Exámen Parcial de Probabilidad y Estadística

La condición mínima de aprobación es dos prácticos y un teórico correctos. Todas las respuestas deberán estar debidamente justificadas.

Ejercicio 1

Dos fábricas de tomates enlatados (A y B) abastecen con su producción el consumo de la CABA, en una proporción de 40 % y 60 % respectivamente. El peso en gramos tiene una distribución normal. Las latas de la fábrica A tienen una media de 250 gramos y una varianza de 16 g^2. Los de la fábrica B se sabe que el la mediana es de 247 gramos y una varianza de 36 g^2.

(a) Se compra al azar una lata, ¿cuál es la probabilidad de que su peso sea mayor que 248 gr.

(b) Si el peso de la lata elegida al azar es mayor que 248 gramos, ¿cuál es la probabilidad de que sea de la fábrica B?

Ejercicio 2

La proporción de impurezas que aparecen en un compuesto de resinas es una variable aleatoria con la siguiente función de densidad:

$$f_X(x) = \begin{cases} ax & 0 < x \leq 2 \\ 0 & \text{otro caso.} \end{cases}$$

(a) Hallar la función de distribución de probabilidades.

(b) Hallar el valor mediano y la varianza de la distribución.

Ejercicio 3

Las distancia entre dos fallas de hilado se distribuye exponencialmente de modo tal que la probabilidad de que haya más de 1800 metros entre dos fallas es 0.082. Calcular la probabilidad de que entre dos fallas consecutivas tomadas al azar haya entre 1000 y 1500 metros.

Ejercicio 4

De los pernos manufacturados por cierta aplicación el 90 % satisface la longitud especificada y se puede utilizar inmediatamente, 6 % está demasiado largo y debe recortarse, y el restante 4 % demasiado corto y debe desecharse.

(a) Hallar la probabilidad de que al seleccionar tres pernos al azar todos puedan utilizarse inmediatamente.

(b) Hallar la probabilidad de que al seleccionar cinco pernos al azar al menos uno deba recortarse.

(c) Cuántos se espera de una muestra de 30 pernos que no deban desecharse? Cuántos utilizarse inmediatamente?

Teórico 1

(a) Enuncie dos propiedades de esperanza matemática y demuestre una de ellas.

(b) Sabiendo que $E(X) = 2$ y que $E(X^2) = 8$, hallar:$E[(2 + 4X)^2]$ y $E[X^2 + 2(1 + X)]$.

Teórico 2

Una urna contiene cinco bolillas negras, dos blancas y cuatro rojas. Se extrae aleatoriamente una muestra de 3 bolillas, de a una por vez y reponiendo cada bolilla en la urna antes de extraer la siguiente. Sea X el número de bolillas negras en la muestra, e Y el número de rojas antes de la primera blanca.

(a) Hallar la distribución conjunta de probabilidades de X e Y

(b) Decidir si las variables X e Y son dependientes.

7.2. Modelo para el Segundo Parcial

Universidad Tecnlógica Nacional Facultad Regional Buenos Aires
Segundo Exámen Parcial de Probabilidad y Estadística

La condición mínima de aprobación es dos prácticos y un teórico correctos. Todas las respuestas deberán estar debidamente justificadas.

Ejercicio 1

La siguiente tabla exhibe las antigüedades de los empleados de una cierta fabrica medidas en años y el numero de llegadas tarde de un mes elegido al azar del 2017:

Antigüedad (años)	4	4,6	5,3	6,4	6,7	7
Llegadas tardes (en el mes)	2	3	3	4	5	5

(a) Hallar un modelo lineal que permita estimar la cantidad de llegadas tarde en función de la antigüedad del empleado.

(b) Se puede estimar con este modelo las llegadas tarde de un individuo que tiene 20 años en la empresa?

(c) Estudiar mediante un test de nivel 0.05 la validez de este modelo lineal.

Ejercicio 2

El fabricante de neumáticos sostiene que en promedio sus productos tienen una duración no menor a 50mil km antes de necesitar reemplazo. Un grupo de consumidores quiere poner a prueba dicha afirmación, suponiendo que la distribución de la duración es normal

(a) Plantee las hipótesis a contrastar, el estadístico del test junto con su función de distribución, la región de rechazo y la decisión tomada nivel aproximado 0.05.

(b) Si una muestra de 30 neumáticos arrojó un promedio de 49mil km y un desvío de 3mil km ¿cuál es el pvalor de la prueba?.

Ejercicio 3

Se desea estimar el tiempo medio de espera en una cola de supermercado los días domingos al mediodía, para ello se tomo el tiempo de espera de 100 clientes en diferentes días domingo en el mismo supermercado y la muestra arrojó un promedio de 20 minutos con una dispersión de 3 minutos. Hallar un intervalo del 98 % de confianza para el tiempo promedio. Hallar el tamaño de muestra necesario para reducir a la mitad la longitud de dicho intervalo.

Ejercicio 4

Se sabe por experiencias anteriores que el peso de los recién nacidos en cierta localidad se distribuye normalmente con media 3kg y desviación 400g. ¿Cuál es la probabilidad de que el promedio de los pesos de los 16 bebes que nacieron ayer supere los 3,2kg?

Teórico 1

Definir los tipos de errores de una prueba estadística y vincular estos conceptos con potencia de la prueba y p valor.

Teórico 2

Explicar dos métodos de estimación y ejemplificar. Cómo puede seleccionar entre dos estimadores de un parámetro?

7.3. Modelo para el Final

Probabilidad y Estadística **Examen Final-2018**

La condición de aprobación es al menos el 50 % de la práctica y al menos el 50 % de la teórica.

Ejercicio 1

En una empresa se sabe que el consumo telefónico tiene distribución normal y que su promedio histórico es de 2520 pulsos mensuales. El gerente considera que debido al uso de los celulares el promedio de pulsos de línea ha bajado. Propone para contestar a esta cuestión un test de hipótesis basado en una muestra de los últimos 12 meses que le facilita su asistente:

$$2477 \quad 2443 \quad 2748 \quad 2662 \quad 2531 \quad 2578 \quad 2744 \quad 2301 \quad 2049 \quad 2237 \quad 2198 \quad 2602$$

a) ¿Cuál es la conclusión del gerente a nivel 0.05?

b) Cinco años después el gerente quiere rehacer la misma prueba de hipótesis con los datos de los 5 años previos. Le pide a su asistente que le facilite los datos y éste (como ya era la hora de retirarse) decide darle los mismos datos que le dio cinco años atrás pero repetidos cinco veces. Si el gerente rehace el test con esos datos, ¿qué conclusión obtendrá al mismo nivel que antes?

Ejercicio 2

Se diseña un ascensor de carga cuyo límite es 1000 kg. El peso de cada caja sigue una distribución normal con un peso medio de 32 kg y un desvío estándar de 10 kg.

a ¿Cuál es la probabilidad de que un grupo de 30 cajas exceda el límite de carga?

b Se toma una muestra aleatoria simple de 2 cajas ¿cuál es la probabilidad de que el mínimo de la muestra sea inferior a 30kg?

Ejercicio 3

Sea X(mm) una variable aleatoria que describe la profundidad de un disparo. Su función de densidad de probabilidad es:

$$f_X(x) = \begin{cases} 2x^{-1} & c \leq x \leq c+2 \\ 0 & \text{otro caso} \end{cases}$$

a) Indicar si los eventos $\{x/x > c+1\}$ y $\{x/x < c+1{,}5\}$ son independientes.

b) Si el diámetro de la perforación del disparo Y en mm, se puede definir como $Y = 10X + 1$, hallar el valor del diámetro esperado.

Ejercicio 4

La siguiente información resultó de un estudio realizado para examinar la relación entre una medida de la corrosión del acero y la concentración de $NaPO_4$.

corrosión	7,68	6,95	6,30	5,75	5,01	1,43	0,93	0,72	0,68	0,65	0,56
$NaPO_4$	2,5	5,03	7,6	11,6	13,0	19,6	26,2	33,0	40,0	50,0	55,0

a) Estime el modelo lineal para explicar la corrosión conociendo el nivel de $NaPO_4$. Indique los supuestos del modelo.

b) Estime el coeficiente de determinación e interprételo en el contexto del problema.

Teórico 1

a) Defina eventos independientes y eventos mutuamente excluyentes. Ejemplifique.

b) Demuestre que si dos eventos son independientes también lo son sus complementos.

Teórico 2

a) Defina la distribución exponencial, señalando el soporte de la variable.

b) Halle su valor esperado y explique la relación con la distribución de Poisson.