Inferencia Estadística

Ejercicios Parcial 1

Fecha de entrega: examen primer parcial

Opción múltiple

1. Si $0.0225E^2[X] > Var(X)$, entonces

   a) El coeficiente de variación es mayor a 1
   b) La distribución en asimétrica
   c) La mediana es menor a la media
   d) La dispersión en los datos es pequeña

2. Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas $Po(\lambda)$. ¿Cuál de las siguientes afirmaciones es correcta?

   a) $\frac{(n-1)s^2}{\lambda} \sim \mathcal{X}^2_{(n-1)}$
   b) $\bar{X} \sim N(\lambda, \lambda/n)$
   c) $\bar{X} \sim Po(n\lambda)$
   d) $\frac{(n-1)s^2}{\lambda} \sim \mathcal{X}^2_{(n)}$

3. Suponga que $X_i \sim U(0, \theta)$ y que se tiene una muestra aleatoria de tamaño n. Si se sabe que $\mathbb{P}(\bar{X} \leq 15) \approx 0.95$, entonces

   a) $\theta \approx 1$
   b) $\theta \approx 6$
   c) $\theta \approx 10$
   d) $\theta \approx 14$

4. Se lanza una moneda justa 100 veces, ¿cuál es la probabilidad aproximada de que el número de soles obtenidos esté entre 40 y 60?

   a) 56 %.
   b) 86 %.
   c) 97 %.
   d) 100 %.

5. Suponga que un diagrama de caja y brazos la amplitud intercuartílica es igual a 3 unidades y la mediana se encuentra a 2 unidades de distancia del primer cuartil, entonces es cierto que

   a) La distribución es simétrica

   b) La media de la distribución es 3/2

   c) La dispersión de los datos es pequeña

   d) La media es menor que la mediana

6. Sea $X_1, \ldots, X_n$ una muestra aleatoria de normales estándar y considere el estadístico $W = \sqrt{\frac{1-\tau}{\tau}\sum_{i=1}^{n} X_i^2}$, con $\tau \in \mathbb{R}^+$. ¿Cuál de las opciones es correcta?

   a) $Var(W) = \frac{2n(1-\tau)}{\tau}$

   b) $Var(W) = \frac{n(1-\tau)}{\tau}$

   c) $Var(W) = \frac{(1-\tau)}{n\tau}$

   d) $Var(W) = \frac{2(1-\tau)}{n\tau}$

7. Un examen de estadística cuenta con 48 preguntas de verdadero o falso. Juan elige la respuesta correcta con una probabilidad de 3/4 mientras que Mario simplemente adivina la respuesta correcta en cada pregunta. Si se necesitan 30 respuestas correctas para aprobar el examen, ¿por cuántos puntos porcentuales es más probable que apruebe el examen Juan?

   a) 2

   b) 7

   c) 93

   d) 95

8. En México, un camión de carga de 4 ejes puede transportar un máximo de 32 toneladas. Una empresa dedicada a la venta de productos lácteos debe transportar 87 contenedores y se sabe que el peso de dichos contenedores sigue una distribución con media 368 kgs y una desviación estándar de 3.87 kgs. ¿Cuál es la probabilidad de que los 87 contenedores puedan ser transportados sin violar el reglamento?

   a) 0.28

   b) 0.32

   c) 0.64

   d) 0.89

9. Sea X_n una sucesión de variables aleatorias tal que $F_{X_n}(x) = x - \frac{sen(2n\pi x)}{2n\pi}$ con $0 < x < 1$. Entonces es cierto que

   a) $X_n \xrightarrow{d} Exp(n\pi)$
   b) $X_n \xrightarrow{d} U(0,1)$
   c) $X_n \xrightarrow{d} Po(1)$
   d) $X_n \xrightarrow{d} Bin(n, \pi/n)$

10. Se cuenta con una muestra de tamaño $n = 20$ para la cual la media es 85 y la mediana es 80. Otra muestra de tamaño $n = 30$ tiene media 75 y mediana 72.

   a) La media de las muestras juntas es 80
   b) La media de las muestras juntas es 79
   c) La mediana de las muestras juntas es 76
   d) La mediana de las muestras juntas es 78

11. En un conjunto de datos el mínimo dista 2 unidades del primer cuartil y además se sabe que la distancia entre el segundo y el tercer cuartil es 3 veces mayor que la distancia entre el segundo y el primero. Con base en lo anterior se sabe con certeza que:

   a) La media de los datos es menor al primer cuartil
   b) La media de los datos es mayor al primer cuartil
   c) La distribución es simétrica
   d) La distribución es asimétrica

12. Se tiene una moneda en la cual una cara tiene el valor de 5 y la otra el valor de 10. La moneda es una moneda justa y ésta se lanza de manera continua hasta que la suma total de los lanzamientos supera los 600 puntos. La probabilidad aproximada p de que 75 lanzamientos sean necesarios para lograr lo anterior es

   a) $p \approx 0.01$
   b) $p \approx 0.02$
   c) $p \approx 0.03$
   d) $p \approx 0.04$

13. Una aerolínea afirma que de acuerdo a sus datos históricos, el 5 % de los pasajeros no se presenta a los vuelos trasatlánticos. Si la aerolínea vende 160 boletos para un vuelo MEX-CDG con solamente 155 asientos disponibles, la probabilidad de que todos aquellos pasajeros con reservación puedan volar es aproximadamente del

   a) 56 %.
   b) 86 %.
   c) 97 %.
   d) 99 %.

14. Un estudio sobre las deficiencias de hierro en la sangre de los bebés comparó un grupo de bebés cuyas madres eligieron alimentarlos exclusivamente con leche materna, con otro grupo en el que fueron alimentados exclusivamente con fórmula láctea. La información sobre los niveles de hemoglobina en la sangre de los bebés es la siguiente

grupo	n	media	d.e.
Leche materna	26	13.3	1.7
Fórmula láctea	31	12.4	3.3

Con base en la información provista previamente podemos concluir que

a) La dispersión relativa a la media es mayor en el grupo de leche materna

b) La dispersión relativa a la media es menor en el grupo de leche materna

c) La mediana del grupo de fórmula láctea es menor a su media

d) La mediana del grupo de fórmula láctea es mayor a su media

15. Una cadena de cines desea introducir un nuevo sabor de palomitas. Para ello realiza un estudio que le permita identificar cuál de los cuatro posibilidades de nuevo sabor es la mejor opción. En dicho estudio se considera una muestra de tamaño n. ¿Cuál de los siguientes argumentos es correcto?

a) El tamaño de muestra n debe ser menor a 30

b) Mientras más entrevistados se tengan, mayor será la variación en la media muestral

c) La varianza muestral se distribuye $\mathcal{X}^2_{(n-1)}$

d) A partir de la muestra se deberá proponer un estadístico apropiado para dar respuesta a la pregunta de negocio

16. Si $n = 150$ entonces el mejor estimador para la varianza es

a) $T(\underline{X}) = \frac{1}{n}\sum_{i=1} X_i^2$

b) $T(\underline{X}) = \frac{1}{n-1}\sum_{i=1} X_i^2$

c) $T(\underline{X}) = \frac{1}{n}\sum_{i=1}(X_i - \bar{X})^2$

d) $T(\underline{X}) = \frac{1}{n-1}\sum_{i=1}(X_i - \bar{X})$

17. La media muestral de un conjunto de datos es x mientras que su mediana muestral es y y su desviación estándar muestreal es z. ¿Cuál de los siguientes argumentos es correcto?

   a) Si $x > y$ entonces necesariamente $z < y$
   b) Si $x = z$ entonces y es mínima
   c) Si $x = z$ entonces la distribución de los datos es simétrica
   d) Si $\frac{y}{z}$ es pequeño, entonces los datos son muy variables

18. Las distribuciones de frecuencias

   a) Sólo se pueden obtener para datos con escalas de medición nominales y ordinales.
   b) Se refiere a las diferentes mediciones observadas de un fenómeno.
   c) Son métodos gráficos útiles para descubrir patrones en los datos.
   d) Caracteriza la variedad de valores que toma una variable sobre una población.

19. Sea X_n una sucesión de variables aleatorias tal que $F_{X_n}(x) = \left(\frac{x}{1+x}\right)^n$ con $x > 0$. Entonces,

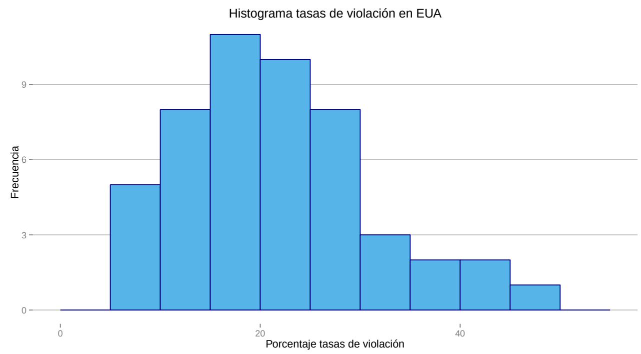
   a) $F_{X_n}(x) \xrightarrow{d} Exp(n)$
   b) $F_{X_n}(x) \xrightarrow{p} Exp(n)$
   c) $F_{X_n}(x) \xrightarrow{2} 0$
   d) $F_{X_n}(x) \xrightarrow{p} 0$

20. En cierta gasolinera ubicada al sur de la CDMX, la cantidad de gasolina Magna que se vende a la semana sigue una distribución Normal con media 50,000 galones y desviación estándar 10,000. El abastecimiento inicial es de 74,000 galones y se tiene programado un abastecimiento semanal de 47,000 galones. La probabilidad p de que, después de 11 semanas, el abastecimiento esté por debajo de los 20,000 galones es de

   a) $p \approx 0.10$
   b) $p \approx 0.21$
   c) $p \approx 0.79$
   d) $p \approx 0.89$

Preguntas abiertas

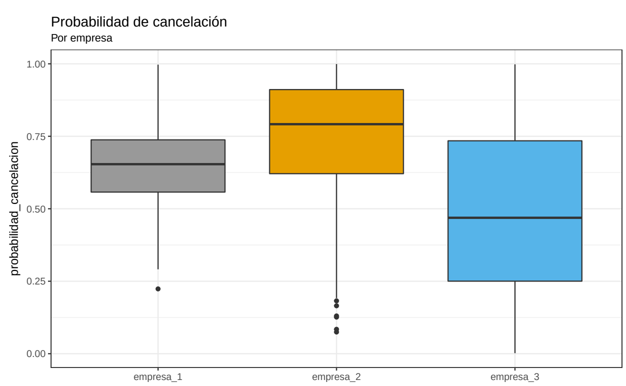
1. En 1973 los Estados Unidos realizaron un estudio sobre violencia y crimen en su país. Se recolectó información de 50 estados diferentes y se calcularon tasas por cada 100,000 habitantes de arrestos, asaltos, homicidios y violaciones. Una organización en particular está interesada en estudiar la evolución de las tasas de violaciones de finales del siglo XX.



Con base en el histograma de las tasas de violación obtenidas, ¿es razonable pensar que las tasas de violación siguen una distribución normal?. En caso contrario ¿qué tipo de distribución sugeriría?. Justifique su respuesta, en ambos casos, con argumentos contundentes.

2. En un estudio de mercado se evaluó la probabilidad de cancelación del servicio de telefonía, cable e internet en la CDMX para 3 diferentes empresas. Esta probabilidad de cancelación se obtuvo para 500 usuarios de cada una de estas compañías proveedoras del servicio.

   La probabilidad de cancelación se estimó con base en variables como la satsfacción con el servicio, la antigüedad, el tipo de paquete contratado y el número de quejas reportado. En el siguiente diagrama se muestran los resultados obtenidos

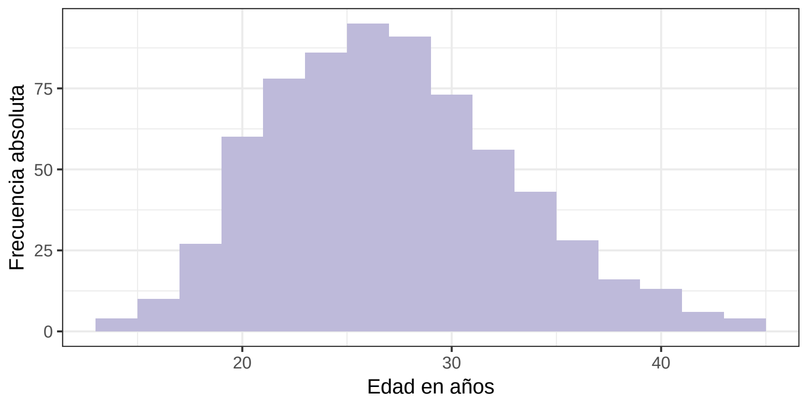


   a) Describa el comportamiento distribucional de la probabilidad de abandono para cada empresa *b*) Determine cuál de las tres empresas se encuentra en una mejor posición y cuál se encuentra en la peor. Justifique

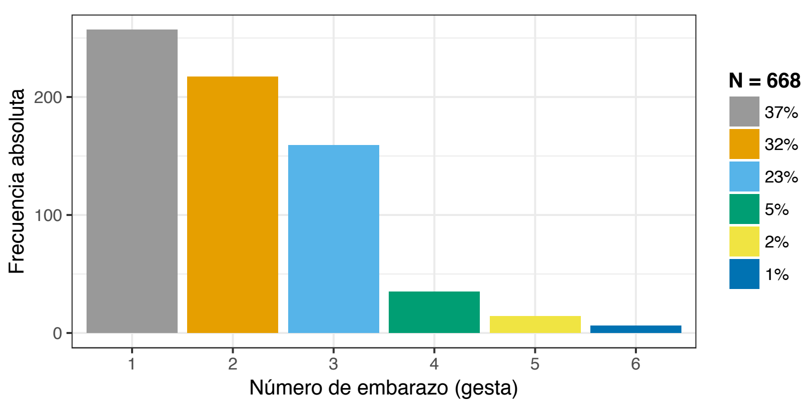
c) ¿Qué podemos decir sobre la industria de comunicaciones en la CDMX desde la perspectiva del consumidor con base en su probabilidad de cancelación?

3. En un Hospital General de Zona del IMSS, el departamento de Ginecología y Obstetricia realizó un estudio sobre los procedimientos quirúrgicos que se llevan acabo en el quirófano (tococirugía). Se cuenta con información de 688 cesáreas que se realizaron en el hospital durante el año 2015.

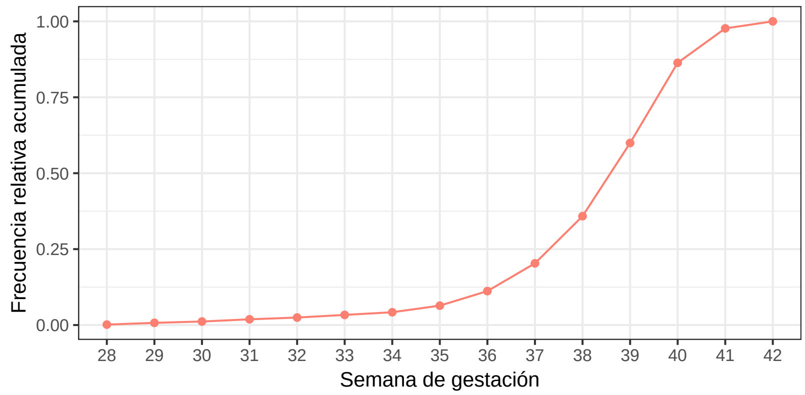
a) Describa el comportamiento de las edades de las madres



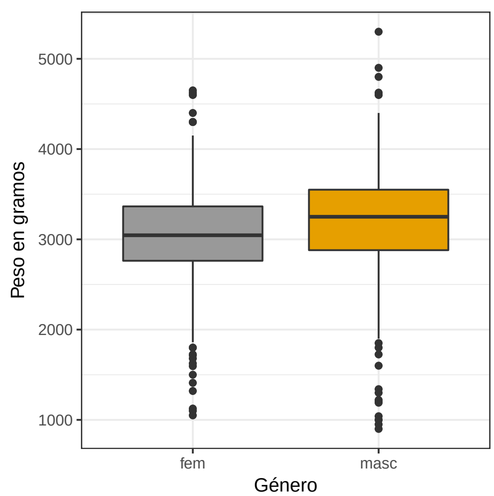
b) Se denomina primigesta a aquellas mujeres en su primer embarazo y multigesta en otro caso. ¿Qué se puede concluir acerca del comportamiento entre primigestas y multigestas en esta cohorte de estudio? ¿Existe evidencia para afirmar que las primigestas son, en general, más jóvenes?



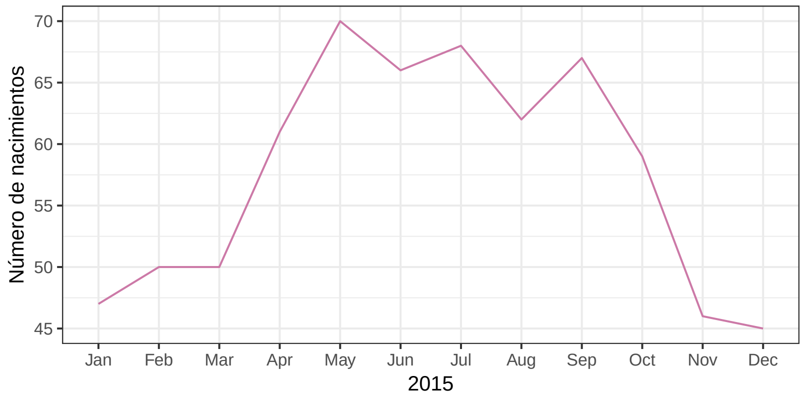
c) Un parto o cesárea se considera de término si éste ocurre entre la semana 38 y 42, mientras que un embarazo que ocurre antes de las 37 semanas se denomina prematuro. ¿Qué porcentaje (aproximadamente) de mujeres tuvieron un parto de término? ¿Con qué probabilidad (aproximadamente) se presentan embarazos prematuros?



d) Existe la creencia popular de que, al momento del nacimiento, los niños pesan más que las niñas. A partir de siguiente diagrama determine si existe evidencia para afirmar dicha creencia.



e) Otra creencia popular es que existen meses en el año en los cuales sistemáticamente hay más nacimientos que en otros. A partir de la siguiente información ¿se puede corroborar lo anterior o solamente es posible describir el comportamiento del año analizado?



4. Suponga que X es una variable aleatoria tal que

X	2	4	6
$P(X = x)$	$1/6$	$1/2$	$1/3$

Obtenga la distribución de muestreo de $\bar{X}$ si se realiza un muestreo aleatorio simple con reemplazo de tamaño 2.

5. Si $\mu = 6$, $\sigma^2 = 0.8$ y $n = 31$, calcule $P(s^2 < 1/2)$.

6. Sean $y_1, \ldots, y_n$ mediciones observadas de una muestra aleatoria. Demuestre que

$$\sum_{i=1}^{n} (y_i - \bar{y}) = 0.$$

7. Un biólogo especializado en el estudio de árboles está interesado en estimar la media basal de cierto tipo de pino. Por otro estudios se sabe que las mediciones de estos árboles se distribuyen de manera Normal con desviación estándar igual a 4 pulgadas cuadradas. Si el biólogo toma una muestra de tamaño $n = 9$, encuentre la probabilidad de que la media muestral se encuentre a lo más a dos unidades de la media poblacional.

8. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ dos muestras aleatorias independientes con $X_i \sim N(\mu_1, \sigma_1^2)$ y $Y_i \sim N(\mu_2, \sigma_2^2)$ y sea $W = \bar{X} - \bar{Y}$ la diferencia de las medias muestrales.

   a) Demuestre que W sigue una distribución normal.
   b) Obtenga $E[W]$.
   c) Obtenga $Var(W)$.
   d) Suponga que $n = m$, $\sigma_1^2 = 3$ y $\sigma_2^2 = 2.5$. Calcule el tamaño de muestra necesario para que la distancia entre W y su media poblacional sea a lo más de 1.5 unidades con una probabilidad de 95 %.

9. Sea X una variable aleatoria tal que $X \sim \mathcal{X}^2_{(\nu)}$ donde ν representa los grados de libertad. Obtenga $E[X]$ y $Var(X)$.

10. En promedio, cierta planta automotriz produce 14 piezas de carrocería al día con una desviación estándar de 2 unidades. Si se toma una muestra aleatoria de tamaño 100,

    a) ¿Cuál es la probabilidad de que la producción media de carrocerías exceda 15 unidades?

    b) Encuentre un intervalo que incluya con 95 % de probabilidad la producción media de carrocerías.

11. Sea $X \sim U(0, 1)$ y $X_n \sim U(0, 1 + 1/n)$. Demuestre que $X_n \xrightarrow{d} X$.

12. Sea $X \sim U(0, 2)$ y $X_n \sim U(1/n, 2)$. Demuestre que $X_n \xrightarrow{d} X$.

13. Sea $S_n = \text{mín}\{X_1, \dots, X_n\}$ y $\{X_i\}_{i=1}^n$ una sucesión de variables aleatorias tal que $X_i \sim U(0, 1)$. Demuestre que la sucesión $\{S_i\}_{i=1}^n$ converge a 0 en media cuadrática y en probabilidad. *Hint:* $\int_0^1 t^m(1-t)^n \, dt = (m!n!)/(m+n+1)!$.

14. Sean $X_1, \dots, X_n$ variables aleatorias independientes e idénticamente distribuidas de manera uniforme en el intervalo (0,1). Definimos $Z_n = \text{máx}\{X_1, \dots, X_n\}$.

    a) Demuestre que $P(Z_n \leq z) = z^n$, $0 < z < 1$.

    b) Sea $U_n = n(1 - Z_n)$. Demuestre que $U_n \xrightarrow{d} Exp(1)$.

15. El plan económico federal de los Estados Unidos buscaba que los contribuyentes ahorraran una parte del saldo a favor que recibían después de hacer su declaración ante el fisco. Suponga que las primeras estimaciones de 35 economistas sobre el porcentaje de ahorro arrojaron como resultados $\mu = 0.26$ y $\sigma = 0.12$.

    a) ¿Cuál es la probabilidad de que la media muestral se encuentre a un 1 % de la media poblacional?

    b) ¿Considera que la estimación de los economistas sobre el porcentaje de ahorro de los contribuyentes es igual al verdadero valor que ahorrarán? Justifique su respuesta.

16. De acuerdo al INEGI, basados en el CENSO de 2010, se sabe que la edad mediana de la población mexicana es de 26 años. Si se encuestan a 100 mexicanos al azar, ¿cuál es la probabilidad aproximada de que al menos 60 de ellos tengan menos de 26 años?

17. Apple asegura que sus productos poseen tan buena calidad que el 85 % de todos los iPhones, iPods y iPads que produce pasaron el control de calidad y, por tanto, no tienen defectos. Si un inspector toma una muestra de tamaño 100 ¿cuál es la probabilidad de que 15 o más gadgets presenten algún defecto?

18. Sea $X_1, \ldots, X_{10}$ una muestra aleatoria de tamaño 10 proveniente de una distribución normal con media 0 y varianza σ^2.

   a) Obtenga la distribución de $10\bar{X}^2/S^2$.

   b) Obtenga la distribución de $S^2/10\bar{X}^2$.

   c) Calule
   $$P\left(-c \leq \frac{S}{\bar{X}} \leq c\right) = .95.$$

19. Sean $X_1, \ldots, X_n$ variables aleatorias independientes distribuidas como una $\mathcal{X}^2_{(1)}$ y sea
   $$Y = \sum_{i=1}^{n} X_i.$$

   a) Demuestre que $Y \sim \mathcal{X}^2_{(n)}$.

   b) Demuestre que $Z = (Y - n)/\sqrt{2n}$ se distribuye asintóticamente como una normal estándar.

20. Sea $\{X_i\}_{i=1}^{n}$ una sucesión de variables aleatorias tal que $X_n \sim U\left(\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right)$ y sea X una variable aleatoria degenerada en $\frac{1}{2}$. Demuestre que $X_n \xrightarrow{d} X$.

21. Con base en la Encuesta Intercensal 2015 realizada por el INEGI se obtuvo la siguiente información sobre fuentes de ingreso y trabajo:

¿Cuál fue se ocupación la semana pasada?	**Porcentaje**
Comerciantes, empleados en ventas y agentes de ventas	15
Funcionarios, directores y jefes	5
Operadores de maquinaria industrial, ensambladores, choferes y conductores de transporte	8
Profesionistas y técnicos	28
Trabajadores artesanales	8
Trabajadores auxiliares en actividades administrativas	10
Trabajadores en actividades agrícolas, ganaderas, forestales, caza y pesca	1
Trabajadores en actividades elementales y de apoyo	18
Trabajadores en servicios personales y vigilancia	9

   A partir de esta información

   a) Determine el tipo de variable y la escala de medición para la ocupación de la semana pasada

   b) Describa el comportamiento distribucional de la variable

   c) Realice un gráfico adecuado para presentar esta información

22. Si X es una variable aleatoria Poisson con parámetro λ, pruebe que la función generadora de momentos de $Y = (X - \lambda)/\sqrt{\lambda}$ está dada por

$$m_Y(t) = \exp\left\{\lambda e^{t/\sqrt{\lambda}} - \sqrt{\lambda}t - \lambda\right\}.$$

23. Los tiempos de llegada (en minutos) de vuelos internacionales al aeropuerto JFK de la ciudad de Nueva York son un una variable aleatoria X tal que

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\{\beta x\} \mathbb{I}_{(0,\infty)}(x)$$

con $\alpha > 0$ y $\beta > 0$.

a) Calcule $E[X]$ y $Var(X)$

b) Se toma una muestra aleatoria de 15 vuelos y se obtienen los siguientes valores 3, 3, 4, 3, 4, 4, 4, 6, 5, 3, 4, 5, 5, 5 y 4. Calcule la media muestral y la varianza muestral

c) Si ahora se toma una muestra de tamaño n = 100 y se observa que la media muestral es igual a 4 y la varianza muestral es igual a 1. Calcule la probabilidad de que, en promedio, el tiempo de arribo de los vuelos sea menor a 5 minutos.

24. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de distribución $F_X(x)$ y suponga que $E[X_i] = \mu$. Se define una nueva muestra aleatoria $Y_1, \ldots, Y_n$ como sigue

$$Y_i = \begin{cases} 1 & \text{si } X_i > \mu \\ 0 & \text{si } X_i \leq \mu \end{cases}$$

Obtenga la distribución de $\sum_{i=1}^{n} Y_i$

25. Sean X y Y variables aleatorias independientes tal que $X \sim N(0,1)$ y $Y \sim N(0,1)$. Sea $Z = \min\{X, Y\}$, demuestre que

$$Z^2 \sim \mathcal{X}_{(1)}.$$

Recuerde que la función de densidad de una distribución ji-cuadrada con k grados de libertad está dada por

$$f_W(w) = \frac{1}{2^{k/2}\Gamma\left(k/2\right)} w^{k/2-1} \exp\{-w/2\} \mathbb{I}_{(0,\infty)}(w) \quad \text{con } k > 0.$$

Hint: $\Gamma\left(1/2\right) = \sqrt{\pi}$.

26. Sea $\{X_i\}_{i=1}^{n}$ una sucesión de variables aleatorias tales que

$$f_{X_n}(x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^{x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \mathbb{I}_{\{0,1,2,\ldots\}}(x)$$

Demuestre que $X_n \xrightarrow{d} Po(\lambda)$.

27. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de variables distribuidas $U(0, \theta)$ y sea

$$T(\underline{X}) = \left(\frac{n+1}{n}\right) X_{(n)}$$

donde $X_{(n)} = \text{mín}\{X_1, \ldots, X_n\}$. Obtenga la distribución de $T(\underline{X})$.

28. Sea Y el número de veces que un contribuyente acude a solicitar asesoría fiscal. Considere la siguiente distribución de probabilidad para Y

y	1	2
$P(Y = y)$	0.4	0.6

y suponga que se realiza un muestro aleatorio de tamaño 2, con orden y con reemplazo. Obtenga la distribución muestral de S^2 y calcule su varianza.

29. El departamento académico de Estadística del ITAM está organizando una conferencia magistral que dará cierto ponente en el mes de abril. Dependiendo del número de asistentes, la conferencia se puede llevar a acabo en el auditorio o en la SC. Si el número de asistentes es mayor a 82 individuos se opta por el auditorio, de lo contrario se elige la SC.

    Por lo que se ha observado en los últimos años, asisten en promedio 75 personas a este tipo de eventos. Suponga que el número de asistentes a esta conferencia es una variable aleatoria Poisson. Calcule la probabilidad de que el departamento opte por reservar el auditorio.

30. Sea X el tiempo de duración de una batería de respaldo para computadoras (*no-break*) y sea $f(x)$ la función de densidad correspondiente,

$$f(x) = 2\left(\frac{1}{\theta} - \frac{x}{\theta^2}\right) \quad \text{si } 0 \leq x \leq \theta \quad \text{con } \theta > 0.$$

Determine la distribución muestral de $\bar{X}$. Argumente con claridad cada uno de los supuestos que utilice.

31. En una sucursal bancaria se desea estimar el tiempo mínimo que tarda un ejecutivo de cuenta en atender a los clientes con un cambio de domicilio. Se sabe que en un día dado la probabilidad de que llegue algún cliente para realizar un cambio de domicilio es 0.7768, y que el número de clientes que llegan a esa sucursal para realizar dicho trámite sigue una distribución de Poisson con parámetro λ.

    Definimos a la variable aleatoria X_i como el tiempo que tarda el ejecutivo de cuenta en atender al i-ésimo cliente en realizar el cambio de domicilio.

    a) Determine el modelo de probabilidad adecuado para X_i

    b) Calcule λ

    c) Considere una muestra aleatoria de tamaño n y sea $Y = \text{mín}\{X_1, \ldots, X_n\}$. Obtenga la función de densidad de Y

32. Para la temporada regular de la NFL 2017 se espera que el número de boletos comprados en taquilla para ver cierto partido es una variable aleatoria con media 2.4 y varianza 4. Suponga que unas horas antes de comenzar el partido, 100 aficionados desean adquirir boletos en taquilla. Si solamente quedan 250 boletos, ¿cuál es la probabilidad de que los 100 aficionados consigan entrar al estadio?

33. La estatura de un grupo de hombres de 18 años tiene media 1.72cm y desviación estándar 2.5cm. Si se toma una muestra aleatoria de tamaño 900, ¿cuál es la probabilidad de que la estatura promedio se encuentre entre 1.70cm y 1.75cm?

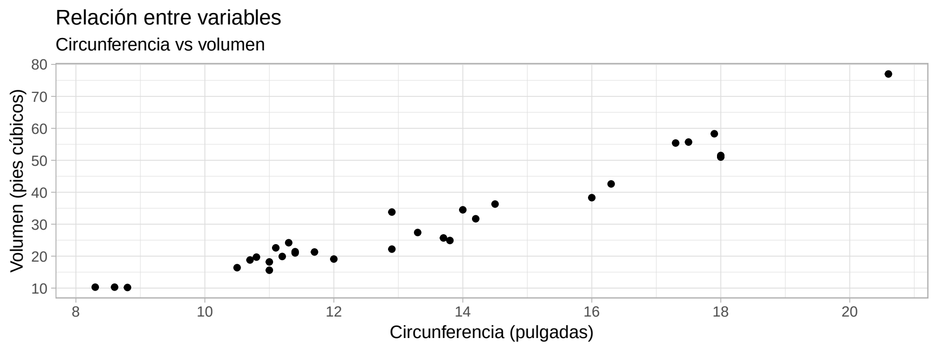
34. La tarea de inferencia consta de 40 ejercicios diferentes los cuales se estima se resuelven en 5 minutos con una desviación estándar de 2 minutos. Calcule la probabilidad de terminar la tarea en menos de 3 horas y 10 minutos.

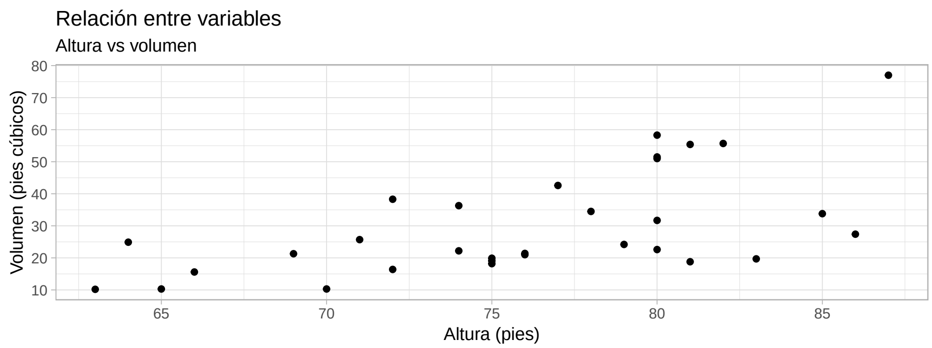
35. Uno de los retos que enfrentan dendrólogos y guardabosques es el cálculo del volumen de algunas especies. Resulta complejo obtener el volumen ya que los métodos actuales requieren equipamiento especializado, costoso y además son métodos destructivos pues conducen a la tala y desmantelamiento del árbol.

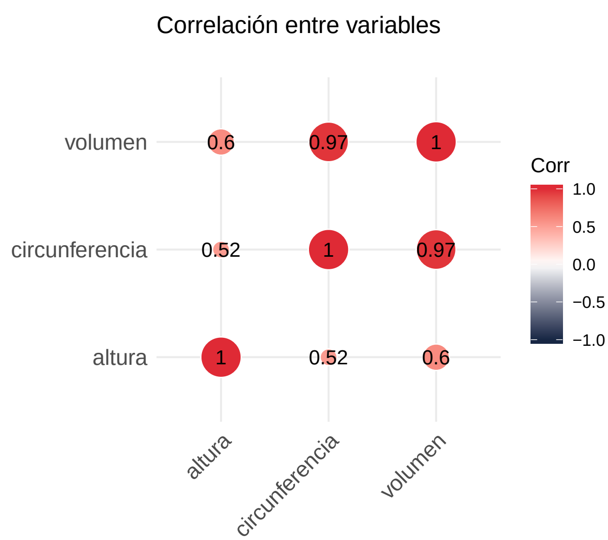
    Por otro lado, calcular la altura y la circunferencia del tronco es considerablemente más sencillo y no requiere equipo especial. Por lo tanto, resulta interesante obtener una herramienta que tenga la capacidad de estimar apropiadamente el volumen de los árboles a partir de la altura y/o la circunferencia del mismo.

    Consideremos una muestra aleatoria de tamaño $n = 31$ de diferentes especímenes. A partir de dichos datos se obtiene la siguiente información:

	Circunferencia	Altura	Volumen
Mín.	8	63	10
1er Cuartil	11	72	9
Mediana	13	76	24
Media	13	76	30
3er Cuartil	15	80	37
Máx.	21	87	77







Con base en toda la información provista, si se debiera utilizar alguna de las dos variables (altura o circunferencia) para determinar el volumen de un árbol, ¿cuál de las dos variables utilizaría?. Argumente y justifique su respuesta desarrollando la mayor cantidad de puntos posibles utilizando únicamente la información anterior.

36. Un ingeniero cuenta con un juego de 40 baterías las cuales pueden ser de dos diferentes tipos: A y B. Las baterías tipo A tienen una duración promedio de vida de 50 meses con una desviación de 15 meses, mientras que las del tipo B tienen una duración promedio de 30 meses con una desviación de 6 meses. El ingeniero sabe que es 3 más veces probable que haya baterías del tipo A que del tipo B. ¿Qué tan probable es que las 40 baterías superen los 1700 meses de vida?

37. Sea $X_1, \ldots, X_n$ una muestra aleatoria de normales estándar y considere el estadístico $W = \sqrt{\frac{1-\tau}{\tau} \sum_{i=1}^{n} X_i^2}$, con $\tau \in \mathbb{R}^+$. Calcule $Var(W)$.

38. Sea X_n una sucesión de variables aleatorias tal que

$$f_{X_n}(x) = \frac{1}{2}\mathbb{I}_{\{-1,1\}}(x).$$

Definimos

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Demuestre que $S_n \xrightarrow{d} 0$.

Hint: Utilice la desigualdad de Chebyshev para resolver el problema:

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

39. Una constructora mexicana ha solicitado un préstamo al banco para financiar la construcción de un complejo en la ciudad de Querétaro. Debido al historial crediticio de la constructora, un analista de la institución financiera debe estimar el tiempo que le tomaría a la constructora caer en *default*.

Dicho analista considera que el tiempo (en meses) que tarda una empresa, con características similares a la constructora, en caer en *default* tiene función de densidad

$$f_X(x;\eta,\alpha) = \frac{\eta^{-\alpha}}{(\alpha-1)!}x^{\alpha-1}\exp\left\{-\frac{x}{\eta}\right\}\mathbb{I}_{(0,\infty)}(x),$$

con $\alpha = 1$ y $\eta > 0$.

Si se sabe que la probabilidad de que la empresa caiga en *default* después de 3 años es 0.05, demuestre que si se toma una muestra de tamaño 6 entonces

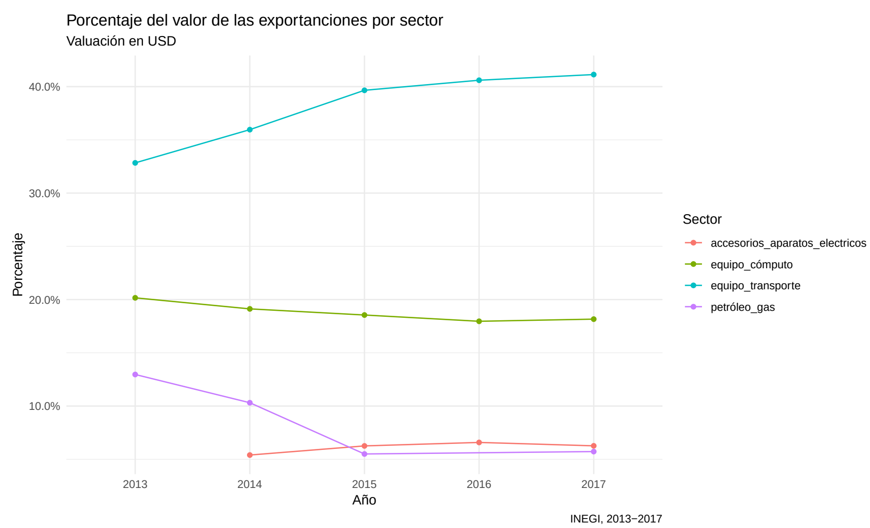
$$\bar{X} \sim \mathcal{X}^2_{(12)}.$$

Hint: Recuerde que si $X \sim Gamma(\alpha, \beta)$, entonces $\frac{2X}{\beta} \sim \mathcal{X}_{(2\alpha)}$.

40. Con base en los datos publicados por el INEGI, se obtuvo la información de las exportaciones totales anuales que hizo México en el periodo de 2013 a 2017 en miles de dólares estadounidenses. Asimismo, se calculó la aportación en porcentaje de cada Entidad Federativa al total de importaciones. A continuación se muestran algunos resúmenes informativos de estos datos.

Año	Exportaciones totales
2013	329,582,726.00
2014	347,790,214.00
2015	336,998,168.00
2016	324,882,422.00
2017	349,840,638.00

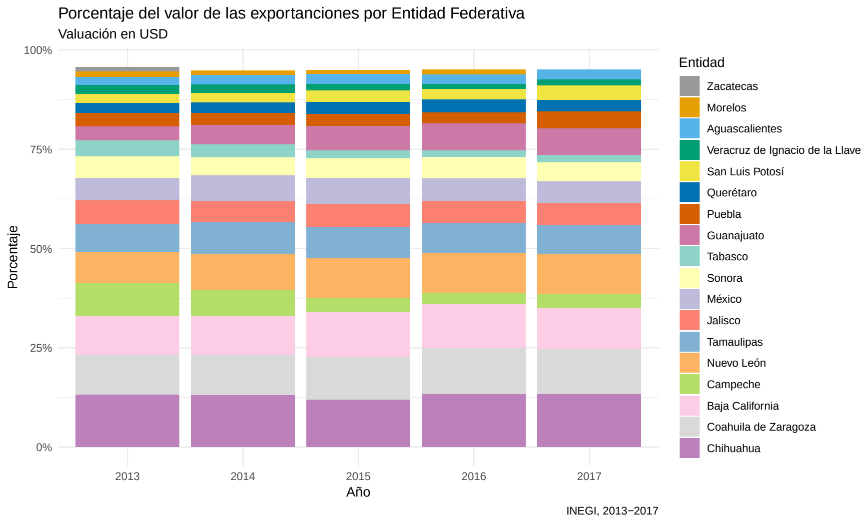
Valores en miles de USD



a) Describa el comportamiento (tendencia) de cada sector de importanción en los últimos años.

b) Calcule, aproximadamente, el crecimiento real para cada sector. ¿Cuál de todos los sectores ha mantenido un crecimiento en los últimos años?

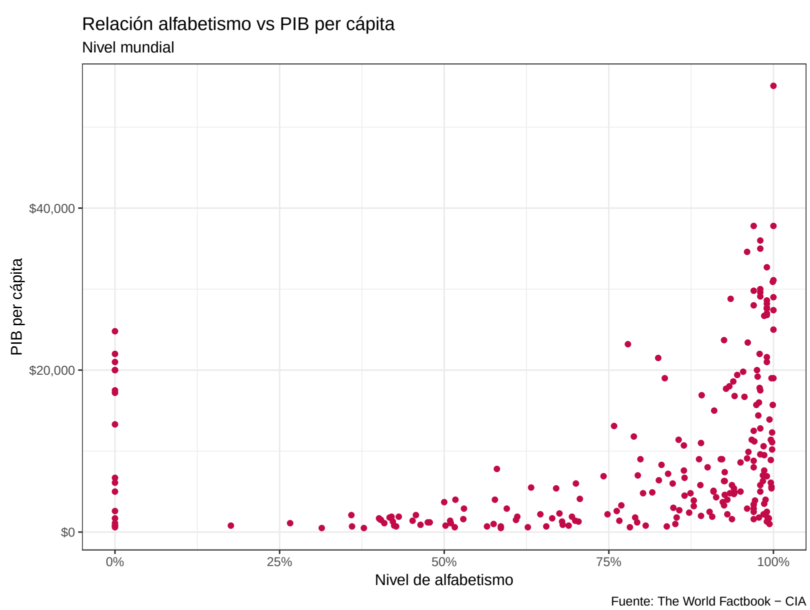
c) ¿Qué podemos decir acerca del sector de gas y petróleo respecto a los otros sectores?



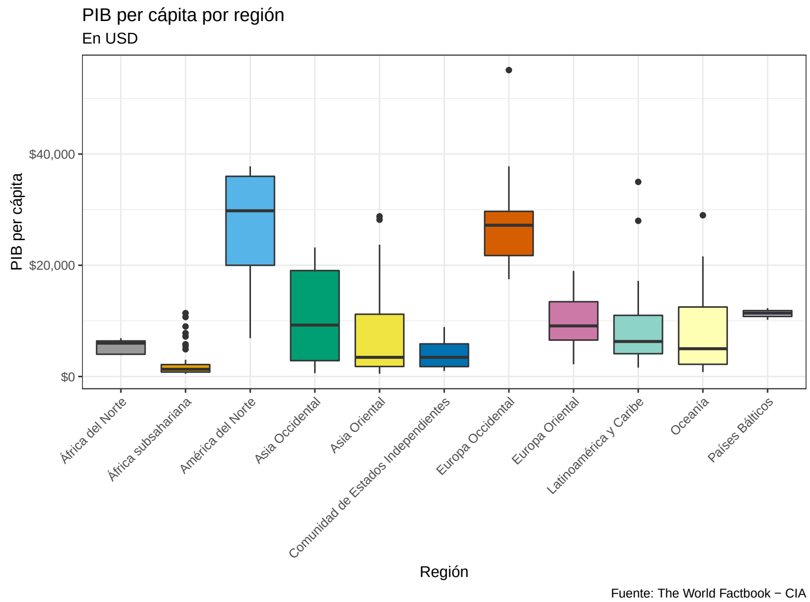
d) ¿Podemos decir que el estado de Campeche ha disminuido su constribución puesto que el sector de equipo de cómputo ha presentado una caída?

e) ¿Existe alguna zona/región del país que represente una gran proporción del total de exportaciones?

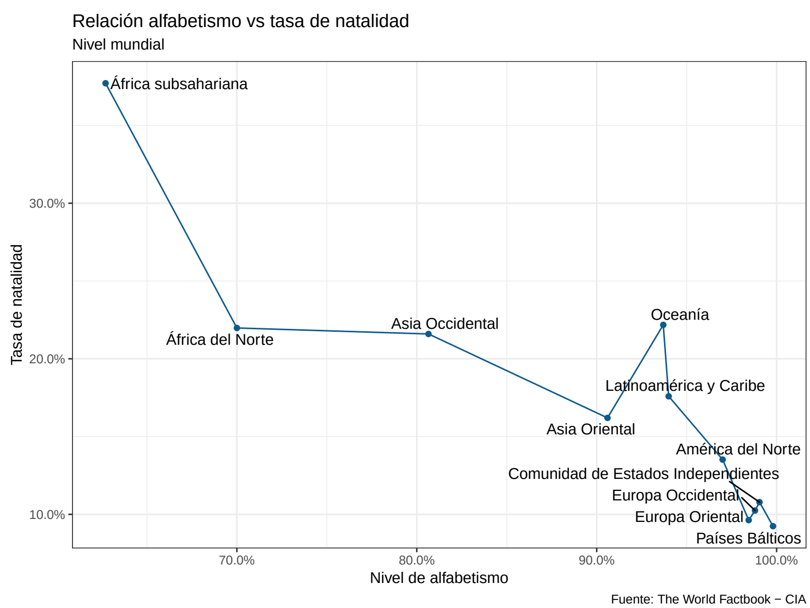
41. *The World Factbook* es una fuente de información pública del gobierno de Estados Unidos que provee datos acerca de población, gobierno, economía, energía, geografía y más sobre 267 países del mundo. Con base en la información de este sitio se obtuvieron algunos datos sobre alfabetismo y PIB per cápita:



a) Describa la relación entre el PIB per cápita y el porcentaje de alfabetismo, ¿qué patrón de comportamiento sugieren los datos? ¿qué conclusiones se pueden hacer?



b) Describa el comportamiento del PIB per cápita por región. Realice un análisis comparativo entre regiones y concluya.



c) La gráfica muestra la relación entre el alfabetismo y la tasa de natalidad. Describa el comportamiento observado, ¿qué conclusiones se pueden establecer?. Con base en toda la información mostrada **elabore una conclusión acerca de la situación de los países** argumentando **únicamente** con los datos provistos.

42. El tiempo en minutos que tarda un sistema de cómputo en procesar una solicitud de estado financiero es una variable aleatoria X_i con función de densidad

$$f(x_i;\theta) = \exp\left\{-(x_i-\theta)\right\} \mathbb{I}_{(\theta,\infty)}(x_i), \quad \theta \in \mathbb{R}$$

Suponga que se cuenta con una muestra aleatoria $X_1, \ldots, X_n$ de los tiempos de procesamiento en un día dado.

a) Defina un estadístico para el tiempo total destinado al procesamiento de las solicitudes de dicho día.

b) Obtenga una distribución aproximada para el estadístico propuesto en el inciso anterior. **Hint:** $\int xe^{-x}\,dx = -e^{-x}(x+1)+c$ y $\int x^2e^{-x}\,dx = -e^{-x}(x^2+2x+2)+c$

43. Sea $X_1, \ldots, X_n$ una muestra aleatoria de $U(0,\theta)$ y sea

$$X_{(n)} = \text{máx}\{X_1, \ldots, X_n\}.$$

a) Obtenga la función de densidad de $X_{(n)}$

b) Calcule $E[X_{(n)}]$

c) Demuestre que $X_{(n)} \xrightarrow{P} \theta$

Inferencia Estadística

Tarea 2

Fecha de entrega: examen segundo parcial

Opción múltiple

1. Considera las siguientes afirmaciones:

   A La función de verosimilitud es una función de densidad de probabilidad.

   B Si $f(x;\theta)$ es una función que cumple con las condiciones de regularidad, entonces cualquier estimador $\tilde{\theta}$ es insesgado.

   a) A y B son ambas verdaderas.
   b) A es verdadera y B es falsa.
   c) A es falsa y B es verdadera.
   d) A y B son ambas falsas.

2. Un intervalo de confianza para la media con varianza desconocida de una población de Normales a un nivel de 68 % de confianza

   a) Se construye a partir de una cantidad pivotal con distribución Normal
   b) Es un intervalo asimétrico alrededor de la media
   c) Es un intervalo más corto que uno equivalente a un 95 % de confianza
   d) Es un intervalo más largo que uno equivalente a un 95 % de confianza

3. Los trenes de la línea 7 del metro de la CDMX (Barranca del Muerto - El Rosario) circulan en un horario de 5am - 12am. Definimos a X como el tiempo que espera una persona que llega a cierta estación a cualquier hora del día durante el horario de servicio hasta que aparece un tren. Suponga que $X \sim U(0,\theta)$ y que se emplea como cantidad pivotal

   $$Q(\underline{X},\theta) = \frac{X_{(n)}}{\theta} \qquad \text{donde} \qquad X_{(n)} = \max\{X_1,\ldots,X_n\}.$$

   Si una persona ha observado que sus tiempos de espera en la última semana han sido $\underline{x} = \{2,5,2,7,3,8,3,5,5,3,9,8,2,3,5,6,6,7,8,9,10,3\}$, entonces el intervalo de confianza para θ a un nivel de confianza de 95 % está dado por

   a) $(9.92, 10.31)$

b) $(10.12, 11.88)$
c) $(10.02, 11.45)$
d) $(10.52, 12.03)$

4. Considera las siguientes afirmaciones:

   A Si el tamaño de muestra es grande, es preferible un estimador insesgado que un estimador eficiente

   B Todos los estimadores máximo verosímiles alcanzan la cota inferior de Crámer Rao

   a) A y B son ambas verdaderas.
   b) A es verdadera y B es falsa.
   c) A es falsa y B es verdadera.
   d) A y B son ambas falsas.

5. Una función de los datos es cantidad pivotal si

   a) La distribución asociada depende del parámetro de interés
   b) La distribución asociada es simétrica
   c) Es función del parámetro de interés pero su distribución no depende de dicho parámetro
   d) No es función del parámetro de interés pero su distribución depende de dicho parámetro

6. Si se sabe que $n = 35, \sum x_i = 8575$ y $\sum x_i^2 = 2,104,445$, entonces un intervalo al 95 % de confianza para la media está dado por

   a) (225.20, 264.79)
   b) (228.33, 261.66)
   c) (224.47, 265.52)
   d) (219.42, 270.57)

7. Sea $X_1, \ldots, X_{100} \sim Po(\lambda)$. Un intervalo asintótico para λ está dado por

   a) $\bar{X} \pm \frac{|z_{\alpha/2}|}{n}\sqrt{\sum X_i}$
   b) $\bar{X} \pm \frac{|z_{\alpha/2}|}{n}\bar{X}^2$
   c) $\bar{X} \pm |z_{\alpha/2}|\sqrt{\bar{X}}$
   d) $\bar{X} \pm \frac{|z_{\alpha/2}|}{\sqrt{n}}\sqrt{\sum X_i^2}$

8. Sea $X_1, \ldots, X_{23}$ una muestra aleatoria con función de densidad $f(x;\theta)$ donde θ es el parámetro de interés y sea $\tilde{\theta}$ el EMV. ¿Cuál de las siguientes afirmaciones es correcta?

   a) $\tilde{\theta}$ es insesgado
   b) $\tilde{\theta}$ es consistente

c) $\tilde{\theta}$ es eficiente

d) Ninguna de las anteriores

9. Considera las siguientes afirmaciones:

A Todos los EMV cumplen con las condiciones de regularidad

B Si se cumple que $\tilde{\theta} \xrightarrow{d} \theta$ pero no se cumple que $\tilde{\theta} \xrightarrow{2} \theta$, entonces $\tilde{\theta}$ no es consistente

a) A y B son ambas verdaderas.

b) A es verdadera y B es falsa.

c) A es falsa y B es verdadera.

d) A y B son ambas falsas.

10. Sea $X_1, \ldots, X_{20}$ una muestra aleatoria de Exponenciales tal que $E[X_i] = \theta$ y considere la cantidad pivotal $Q(\underline{X}; \theta) = \frac{2\sum X_i}{\theta}$. Si $\sum x_i = 45$, entonces un intervalo de confianza para θ al 95 % está dado por

a) (1.51, 3.68)

b) (1.62, 3.39)

c) (1.75, 3.09)

d) (1.41, 4.01)

11. Considera las siguientes afirmaciones:

A Si la función de densidad $f(x; \theta)$ no cumple las condiciones de regularidad no es posible obtener la cota inferior de Crámer-Rao

B Considere que $T_1(\underline{X})$ y $T_2(\underline{X})$ son dos estimadores de θ. Se sabe que $Var(T_1) > Var(T_2)$, $Bias(T_1) = 0$ y
$Bias(T_2) = \frac{3n^2\theta}{(n^2-3)(n-1)}$. Por lo tanto si el tamaño de muestra es grande entonces T_1 es mejor que T_2.

a) A y B son ambas verdaderas.

b) A es verdadera y B es falsa.

c) A es falsa y B es verdadera.

d) A y B son ambas falsas.

12. Se va a realizar una estimación con un tamaño de muestra $n < 30$ ¿cuál de los siguientes escenarios es preferible para dicho estimador?

a) $T(\underline{X})$ sea eficiente pero sesgado

b) $T(\underline{X})$ sea insesgado

c) $T(\underline{X})$ sea ineficiente y sesgado

d) $T(\underline{X})$ sea consistente

13. Considera las siguientes afirmaciones:

   A Sea (t_1, t_2) un intervalo de confianza para θ a un nivel $(1-\alpha)$, entonces la probabilidad de que θ se encuentre entre esos valores es de $(1-\alpha)$.

   B El estimador máximo verosimil se distribuye asintóticamente Normal con varianza igual al inverso de n veces la información de Fisher

   a) A y B son ambas verdaderas.

   b) A es verdadera y B es falsa.

   c) A es falsa y B es verdadera.

   d) A y B son ambas falsas.

14. Se desea estimar las ventas mensuales totales de todos los competidores en la industria de botanas saladas. Únicamente se cuenta con la información sobre las ventas propias pero se sabe que dichas ventas representan aproximadamente un 80 % de las ventas totales.

   a) Es posible estimar las ventas totales considerando que el otro 20 % de las ventas sigue un comportamiento similar al observado

   b) La varianza del estimador empleado en este caso es alta

   c) El estimador es eficiente

   d) No es posible estimar las ventas totales

15. Considera las siguientes afirmaciones:

   A La función de verosimilitud es una función de densidad de probabilidad.

   B Por las propiedad del EMV, sabemos que converge en media cuadrática al verdadero valor del parámetro.

   a) A y B son ambas verdaderas.

   b) A es verdadera y B es falsa.

   c) A es falsa y B es verdadera.

   d) A y B son ambas falsas.

16. Una empresa dedicada a la venta de botanas saladas tiene 3 diferentes productos en su portafolio. Con base en una estimación de las ventas promedio (en miles de unidades) para el próximo trimestre, determine cuál de los siguientes argumentos es correcto:

Producto	Intervalo de confianza al 95 %
A	(3.4, 4.7)
B	(2.5, 3.5)
C	(4.2, 4.7)

   a) El producto A producirá más ventas (en promedio) que el producto C con una probabilidad de 95 %

b) La estimación para el producto B es menos variable que para el producto C

c) La distribución asociada a la cantidad pivotal empleada para estimar el intervalo es t-Student

d) La distribución asociada a la cantidad pivotal empleada para estimar el intervalo es Normal

17. Sea $\tilde{\theta} = T(\underline{X})$ un estimador para θ, entonces por propiedades de los estimadores sabemos que

    a) $\tilde{\theta}$ maximiza la probabilidad de haber generado la muestra observada

    b) $\tilde{\theta}$ es un estadístico

    c) $\tilde{\theta}$ es consistente

    d) $\tilde{\theta}$ es asintóticamente eficiente

18. El estimador de máxima verosimilitud es aquel estimador que por propiedades

    a) Máximiza la probabilidad de haber generado la muestra observada y es insesgado

    b) Máximiza la probabilidad de haber generado la muestra observada y es eficiente

    c) Se distribuye asintóticamente Normal con varianza igual a la información de Fisher

    d) Alcanza la cota inferior de Cramer-Rao

19. Si $L(\hat{\theta}_1; \underline{x}) < L(\hat{\theta}_2; \underline{x})$ entonces se cumple que

    a) $\hat{\theta}_1$ es el estimador máximo verosimil

    b) $\hat{\theta}_2$ es el estimador máximo verosimil

    c) Es más probable haber generado la muestra observada con $\hat{\theta}_1$

    d) Es más probable haber generado la muestra observada con $\hat{\theta}_2$

20. Los empleados que trabajan para una compañía de mudanzas llevan las cajas al sitio, las ensamblan y empacan. La Compañía quiere una estimación del tiempo requerido para ensamblar un tipo especial de caja. Se eligen aleatoriamente 16 situaciones en las cuales se toma el tiempo requerido para ensamblar un tipo especial de cajas. Los resultados son $\sum X_i = 336$ segundos y $\sum X_i^2 = 7116$. El IC al 99 % para el tiempo medio de ensamblaje es

    a) $(19.6988, 22.3012)$

    b) $(20.1235, 21.8765)$

    c) $(19.5267, 22.4733)$

    d) $(19.9343, 22.0657)$

Preguntas abiertas

1. Considere una muestra aleatoria de tamaño 3 proveniente de una distribución exponencial con media θ. Sean

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3} \quad \text{y} \quad \hat{\theta}_4 = \bar{Y}$$

estimadores de θ. Demuestre que todos estos estimadores son estimadores insesgados de θ y calcule la eficiencia relativa entre $\hat{\theta}_1$ y $\hat{\theta}_4$, entre $\hat{\theta}_4$ y $\hat{\theta}_2$ y entre $\hat{\theta}_3$ y $\hat{\theta}_4$.

2. Sea $X_1, \ldots, X_n$ una muestra aleatoria de $U(\theta, \theta + 1)$. Sean

$$\hat{\theta}_1 = \bar{X} - \frac{1}{2} \quad \text{y} \quad \hat{\theta}_2 = X_{(n)} - \frac{n}{n+1}.$$

Determine cuál es el mejor estimador para θ.

3. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ muestras aleatorias independientes provenientes de poblaciones con media μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Demuestre que $\bar{X} - \bar{Y}$ es un estimador consistente de $\mu_1 - \mu_2$.

4. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad dada por

$$f(x;\theta) = \left(\frac{1}{\theta}\right) r x^{r-1} \exp\left\{-\frac{x}{\theta}\right\} \mathbb{I}_{(0,\infty)}(x), \quad \theta > 0.$$

Encuentre el EMV de θ.

5. $X_1, \ldots, X_n$ una muestra aleatoria de una Uniforme con función de densidad

$$f(x;\theta) = \frac{1}{2\theta + 1} \mathbb{I}_{(0,\, 2\theta+1)}(x), \quad \theta > 0.$$

Encuentre el EMV para $Var\left(\sum_{i=1}^{n} X_i\right)$.

6. Considere la siguiente función de densidad

$$f(x;\alpha,\theta) = \left(\frac{1}{\Gamma(\alpha)\theta^\alpha}\right) x^{\alpha-1} e^{-x/\theta} \mathbb{I}_{(0,\infty)}(x)$$

con $\alpha > 0$ conocida.

   a) Encuentre el EMV $\hat{\theta}$ de θ.
   b) Calcule el sesgo de $\hat{\theta}$.
   c) Demuestre que $\hat{\theta}$ es consistente.
   d) Obtenga la cota inferior de Cramér-Rao para $\hat{\theta}$.
   e) Calcule $E[\hat{\theta}]$ y $Var(\hat{\theta})$.

f) Determine la distribución de $\hat{\theta}$.

7. Sea $X_1, \ldots, X_n$ una muestra aleatoria de $U(0, \theta)$. Considere a los estimadores

$$\hat{\theta}_1 = 2\bar{X} \quad \text{y} \quad \hat{\theta}_2 = \left(\frac{n+1}{n}\right) X_{(n)}.$$

Calcule el ECM para cada estimador y determine cúal de los dos es mejor estimador para θ.

8. Calcule el estimador máximo verosímil de θ si

$$f(x; \theta) = \left(\frac{2x}{\theta}\right) \exp\left\{-\frac{x^2}{\theta}\right\} \mathbb{I}_{(0,\infty)}(x).$$

y determine la distribución de $\hat{\theta}$.

9. Sean $X_1, \ldots, X_n$ variables aleatorias con distribución de Poisson con parámetro λ. Determine el EMV de $P(X_i = 0) = e^{-\lambda}$.

10. Sea $X \sim N(0, \sigma^2)$ y considere la cantidad pivotal X^2/σ^2. Obtenga un intervalo de confianza para σ^2 al 95 % de confianza.

11. Un instrumento de precisión tiene como garantía leer con un error máximo de una unidad. Una muestra de cuatro lecturas del mismo objeto dio como mediciones 353, 351, 351 y 355. Calcule un intervalo de confianza al 90 % para la varianza de la población. ¿Qué supuestos deben establecerse? ¿Es adecuada la garantía?

12. En un proceso de llenado de bolsas de detergente se ha verificado que la media del proceso es de 1005 gramos. Se desea estimar ahora la varianza σ^2 del proceso mediante un intervalo del 95 % de confianza. Para esto el departamento de ingeniería solicitó una muestra de 20 bolsas seleccionadas al azar y se obtuvo que el estadístico $s^2 = 2.48$. Obtenga un intervalo de confianza para σ^2.

13. La profesora de cierto grupo asegura que sus alumnos obtienen, *en promedio* mejores calificaciones en el segundo parcial comparado con el primero. El coordinador del curso seleccionó al azar 10 alumnos y registró las siguientes calificaciones (sobre 100 puntos)

Alumno	1	2	3	4	5	6	7	8	9	10
Parcial 1	82	66	48	65	41	79	52	90	55	53
Parcial 2	98	65	39	68	56	49	69	100	57	51

Construya un intervalo al 97 % de confianza y determine si la profesora tiene razón.

14. El ITAM considera que más del 40 % de sus alumnos son foráneos. Para corroborar sus sospechas, se realiza una encuesta a 1000 alumnos y resulta que 378 de ellos provienen del interior de la República. Calcule un intervalo al 95 % de confianza para la proporción $\hat{p}$ de alumnos foráneos en el ITAM y determine si las sospechas del ITAM son ciertas.

15. Sea X_1, X_2, X_3 una muestra aleatoria de una población definida por

$$f(x;\theta) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \mathbb{I}_{\mathbb{R}}(x) & \text{si } \theta = 1 \\ \frac{1}{\pi(1+x^2)} \mathbb{I}_{\mathbb{R}}(x) & \text{si } \theta = 2 \end{cases}$$

Si el valor observado de la muestra es (-3,0,1), encuentre el estimador máximo verosímil de θ.

16. Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con función de densidad

$$f(x;\theta) = \exp\left\{-(x-\theta)\right\} \mathbb{I}_{(\theta,\infty)}(x), \quad \theta \in \mathbb{R}$$

   a) Encuentre el estimador de máxima verosimilitud para θ.
   b) ¿Es posible determinar la distribución asintótica del estimador de máxima verosimilitud de $\hat{\theta}$? Si es así calcule dicha distribución, si no, argumente porqué no es posible determinarla.
   c) Encuentre el estimador de máxima verosimilitud del tercer cuartil de X.

17. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población definida por

$$f(x;\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\lambda \frac{(x-\mu)^2}{2\mu^2 x}\right\} \mathbb{I}_{(0,\infty)}(x)$$

Demuestre que los EMVs de λ y μ son

$$\hat{\mu} = \bar{X} \quad \text{y} \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)}.$$

18. La distribución de Pareto es útil para modelar el ingreso. Si $X \sim Pareto(\alpha,\beta)$ entonces

$$f(x;\alpha,\beta) = \alpha\beta^{\alpha} x^{-(\alpha+1)} \mathbb{I}_{[\beta,\infty)}(x)$$

donde β corresponde al salario mínimo. Sea $X_1, \ldots, X_n$ una muestra aleatoria de $Pareto(\alpha,\beta)$.

   a) Obtenga el EMV de α y β
   b) Obtenga el EMV de la mediana de X
   c) Determine la distribución asintótica de $\hat{\alpha}$

19. Se observa una muestra de tamaño $n = 1$ de la variable aleatoria X, con $X \sim Exp(2/\theta)$. Utilice la cantidad pivotal

$$Q(\underline{X};\theta) = \frac{\theta X}{2}$$

para obtener un intervalo de confianza para θ al 90 %.

20. Sean $y_1, ..., y_n$ variables aleatorias tales que

$$y_i = \beta x_i + \epsilon_i, \quad i = 1, \dots, n$$

donde $x_1, \dots, x_n$ son constantes fijas y $\epsilon_1, \dots, \epsilon_n$ son variables aleatorias independientes e idénticamente distribuidas $N(0, \sigma^2)$, con σ^2 conocida.

   a) Encuentre el estimador máximo verosímil de β
   Hint. Considere que $\epsilon_1 = y_i - \beta x_i$ y que $\epsilon_i \sim N(0, \sigma^2)$

   b) Demuestre que $\hat{\beta}$ es insesgado

21. Sean $X_1, \dots, X_n$ variables aleatorias independientes e idénticamente distribuidas $U(0, \theta)$ y sea

$$Q(\underline{X}; \theta) = \frac{1}{\theta} X_{(n)}$$

donde $X_{(n)} = \text{máx}\{X_1, \dots, X_n\}$.

   a) Demuestre que Q es cantidad pivotal

   b) Obtenga un intervalo de confianza para θ al 95 %.

22. Sean $X_1, \dots, X_n$ variables aleatorias con función de densidad

$$f_X(x) = \frac{2\theta^2}{x^3} \mathbb{I}_{(\theta, \infty)}(x)$$

   a) Encuentre el estimador máximo verosímil de θ

   b) Obtenga la distribución del estimador máximo verosímil

   c) Obtenga la cota inferior de Cramér-Rao

23. Sea X una variable aleatoria con función de densidad

$$f_X(x) = \frac{2(\theta - x)}{\theta^2} \mathbb{I}_{(0,\theta)}(x) \quad \text{y sea} \quad Q(X; \theta) = \frac{X}{\theta}.$$

   a) Demuestre que Q es cantidad pivotal

   b) Obtenga un intervalo de confianza para θ al 95 %.

24. Dos economistas están haciendo un estudio sobre el nivel de ingreso de la población económicamente activa en México para determinar si se debe subir el salario minímo. Para ello, los economistas cuentan con una muestra X_i de tamaño $n = 1000$ donde X_i representa el ingreso mensual del i-ésimo individuo.

    Adicionalmente los economistas suponen que la función de densidad correspondiente a las X_i está dado por

$$f(x_i; \alpha, \beta) = \alpha \beta^\alpha x_i^{-(\alpha+1)} \mathbb{I}_{[\beta, \infty)}(x_i)$$

    donde β es el parámetro que representa al salario mínimo.

a) El economista *A* propone emplear el EMV $\hat{\theta}$ para estimar el salario mínimo. Obtenga dicho estimador.

b) El economista *B* propone emplear al siguiente estimador $\tilde{\theta} = \bar{X}$ para proximar el salario mínimo. ¿Cuál de los dos estimadores es mejor?

c) Con base en los datos (misma muestra) se observa que $\hat{\theta} = 2500$ mientras que $\tilde{\theta} = 5700$, ¿cuál de las dos estimaciones es más apropiada para reportar en el estudio? Justifique su respuesta.

25. El número de vuelos que llegan en una hora dada al aeropuerto de París Charles de Gaulle, denotado por X_i, es una variable aleatoria Poisson con parámetro λ. Con base en información de 1 día se obtiene que

$$\sum_{i=1}^{n} X_i^2 = 298,383 \quad \text{y} \quad s^2 = 68.1721$$

a) Calcule el EMV de λ.

b) Calcule el estimador de la probabilidad de que llegue algún vuelo al aeropuerto.

c) Obtenga un intervalo de confianza al 95 % para el parámetro λ. Para ello, suponga que el tamaño de muestra es suficientemente grande.

26. Los empleados que trabajan para una compañía de mudanzas llevan las cajas al sitio, las ensamblan y empacan. La Compañía quiere una estimación del tiempo requerido para ensamblar un tipo especial de caja. Se eligen aleatoriamente 16 situaciones en las cuales se toma el tiempo requerido para ensamblar un tipo especial de cajas. Los resultados son $\sum X_i = 336$ segundos y $\sum X_i^2 = 7116$. Obtenga un IC al 99 % para el tiempo medio de ensamblaje

27. Una casa de encuestas realizó un estudio para determinar la intención de voto para el candidato independiente en las elecciones presidenciales en México en 2018. Para lograr lo anterior, levanta una encuesta en un punto de afluencia y comienza a preguntar a los individuos por el candidato que votará hasta que encuentra a la primera persona que declara que votará por el candidato independiente.

El proceso anterior se modela con una función de densidad está dada por

$$f(x_i;\theta) = \theta(1-\theta)^{x_i-1}\,\mathbb{I}_{\{0,1,2,\ldots\}}(x_i), \quad \text{con } 0 < \theta < 1$$

y además se sabe que $E[X] = \frac{1-\theta}{\theta}$.

a) Obtenga el EMV $\hat{\theta}$ de la intención de voto para el candidato independiente θ.

b) Obtenga el EMV del promedio de personas que se tienen que entrevistar antes de encontrar al primero que votará por el candidato independiente, i.e., $E[X]$.

c) Argumente una interpretación para el estimador del inciso (b)

28. El tiempo en minutos que tarda un sistema de cómputo en procesar una solicitud de estado financiero es una variable aleatoria X con función de densidad

$$f(x;\lambda,\beta) = \frac{1}{\lambda}\exp\left\{-\frac{(x-\beta)}{\lambda}\right\}\,\mathbb{I}_{(\beta,\infty)}(x)$$

Suponga que se cuenta con una muestra aleatoria $X_1,\ldots,X_n$ de dicha variable aleatoria.

a) Obtenga el EMV de λ y β.

b) Determine si el EMV $\hat{\lambda}$ es insesgado. **Hint:** $\int xe^{-x}\,dx = -e^{-x}(x+1)+c$

c) Determine si $\hat{\lambda}$ es consistente

29. Sea $X_1,\ldots,X_n$ una muestra aleatoria con función de densidad

$$f(x;\theta) = \frac{1}{\theta}x^{\frac{1}{\theta}-1}\,\mathbb{I}_{(0,1)}(x), \quad \theta > 0.$$

a) Obtenga el EMV de θ.

b) Calcule la distribución de $-\sum_{i=1}^{n}\log X_i$.

Hint: Primero calcule la distribución de $Y = -\log X$ y luego obtenga la distribución de la suma.

c) Considere la siguiente función

$$Q(\underline{X};\theta) = -\frac{2}{\theta}\sum_{i=1}^{n}\log X_i$$

Demuestre que $Q(\underline{X};\theta)$ es cantidad pivotal.

d) Obtenga un intervalo de confianza a un nivel $(1-\alpha)$ para θ.

30. Sea X una variable aleatoria discreta tal que

X	0	1	2	3
$P(X=x)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

con $0\leq\theta\leq 1$. Suponga que se observa una muestra $\underline{x} = \{3,0,2,1,3,2,1,0,2,1\}$. Obtenga el estimador máximo verosimil de θ.

31. Una empresa dedicada a la venta de bebidas frutales ha decidido modificar la fórmula de sus productos cambiando los azúcares por endulzantes artificiales de bajo contenido calórico. Lo anterior con el objetivo de disminuir los niveles de azúcares en sus productos y aumentar la utilidad, ya que resulta más rentable emplear endulzantes artificiales que endulcorantes naturales.

Para identificar si existe una respuesta favorable o no favorable por parte del consumidor a la modificación, se realiza una prueba de producto con 350 consumidores regulares. Dicha prueba consiste en administrar la nueva versión del jugo sin hacer mención del cambio en el mismo (prueba ciega). El estudio consiste en preguntarle al individuo la disposición de compra: sí lo compraría o no lo compraría. La empresa determina que si la disposición de compra es mayor al 90 %, entonces se lanza la nueva versión del producto.

Suponga que después de realizar el estudio se encontró que la relación entre individuos que declararon disposición de compra favorable vs disposición de compra no favorable fue de 19 a 1. Determine si la empresa debería lanzar el nuevo producto con base en un intervalo de confianza a un nivel de 95 %.

Hint: Se dice que una apuesta paga 2 a 1 cuando la probabilidad de éxito es aproximadamente de 67 %.

32. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad dada por

$$f(x;\lambda) = \frac{1}{\lambda}\exp\left\{-\frac{x}{\lambda}\right\}\ \mathbb{I}_{(0,\infty)}(x), \qquad \lambda > 0$$

y sean

$$T(\underline{x}) = \text{mín}\{X_1, \ldots, X_n\} \qquad \text{y} \qquad S(\underline{x}) = Var\left(\sum_{i=1}^{n} X_i\right)$$

dos estimadores para λ.

a) Obtenga la distribución de $T(\underline{x})$

b) Calcule el error cuadrático medio de $T(\underline{x})$

c) Obtenga el estimador máximo verosímil de $S(\underline{x})$

33. Una agencia de investigación de mercados realizó un estudio sobre consumo de aguas naturales embotelladas. El objetivo del estudio es tener un mejor entendimiento de los hábitos de consumo de dicha categoría de producto. El estudio contempló un total de 1000 entrevistas a consumidores frecuentes de agua embotellada en presentación de 2 litros o menos. Algunas de las preguntas que se realizaron fueron las siguientes:

Utilizando una escala del 1 al 7 donde 1 es nada importante y 7 muy importante, cuando decides comprar una marca de agua embotellada natural en presentaciones de menos de 2 litros, ¿qué tan importante o no es para ti que... (ATRIBUTO)?

Los mercadológos encargados de llevar a cabo el análisis de este estudio deciden utilizar como medida el porcentaje de individuos que eligieron la opción más alta de la escala definida para la evaluación, la cual denominan *Top Box*. En nuestro caso definimos el *top box* del individuo i para el atributo j como

Atributo	Versió en corto
Sea una marca que amas	amor_marca
Te ayude a compensar todo lo malo que comes	compensa_comes_mal
Te ayude a cuidar tu peso	cuida_peso
Te ayude a eliminar todo lo malo	elimina_lo_malo
La encuentres fácilmente	fácil_encontrar
Tenga una imagen saludable	imagen_saludable
Esté más a la vista cuando la vas a comprar	más_visible
Tenga la mejor publicidad	mejor_publicidad
Te ayude a mejorar tu salud	mejora_salud
Sea para todos los días	para_diario
Sea para cuando vas a hacer ejercicio o actividad deportiva	para_ejercicio
Tenga la publicidad más original	publicidad_original
Te quite la sed	quita_sed
Te ayude a sentirte ligero después de tomarla	sentir_ligero
No tenga sodio	sin_sodio
Valga lo que cuesta	vale_lo_que_cuesta

$$top\_box_{ij} = \begin{cases} 1 & \text{si la calificación para el atributo } j \text{ es igual a 7} \\ 0 & \text{si la calificación para el atributo } j \text{ es distinto de 7} \end{cases}$$

a) Obtenga la distribución de muestreo del *top box* correspondiente al j-ésimo atributo considerando toda la muestra.

b) Obtenga el estimador máximo verosímil para el parámetro de interés del inciso anterior.

c) Calcule un intervalo de confianza asintótico para el parámetro de interés del inciso anterior.

34. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad

$$f(x;\theta) = \frac{x}{\theta}\mathbb{I}_{(0,\sqrt{2\theta})}(x) \qquad \theta > 0.$$

a) Obtenga el EMV de θ, es decir, $\hat{\theta}$.

b) Obtenga $E[\hat{\theta}]$.

c) Encuentre una constante c tal que el EMV sea insesgado.

35. Sean $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad

$$f(x;\theta) = \theta(1+x)^{-\theta-1}\mathbb{I}_{(0,\infty)}(x), \qquad \theta > 0.$$

a) Obtenga el EMV de θ

b) Obtenga la distribución asintótica del EMV
Hint: $\log(1+X) \sim Exp(\theta)$

c) Obtenga un IC de confianza para θ a partir de la distribución obtenida en el inciso anterior. Suponga que se observa la muestra $\underline{x} = \{0.85, 1.08, 0.35, 3.28, 1.24, 2.58, 0.02, 0.13, 0.22, 0.52\}$

Inferencia Estadística

Tarea 3

Opción múltiple

1. En una población $N(\mu, \sigma^2)$ la potencia de la prueba

$$H_0 : \mu \geq 5 \text{ vs. } H_1 : \mu < 5$$

   cuando $\mu = 3$ corresponde a

   a) La probabilidad de rechazar H_0 cuando $\mu = 5$
   b) La probabilidad de rechazar H_0 cuando $\mu = 3$
   c) La probabilidad de no rechazar H_0 cuando $\mu = 5$
   d) La probabilidad de no rechazar H_0 cuando $\mu = 3$

2. En un contexto de contraste de hipótesis considere las siguientes afirmaciones

   A La región de rechazo de una prueba de hipótesis se determina en función del valor-p.
   B Si $\alpha \geq 2\beta$ y valor-$p < \alpha$ entonces sabemos con certeza que se rechaza H_0.

   a) A y B son ambas verdaderas.
   b) A es verdadera y B es falsa.
   c) A es falsa y B es verdadera.
   d) A y B son ambas falsas.

3. En un contraste de hipótesis simple contra compuesto, si la región de rechazo aumenta (se hace más grande) entonces

   a) Disminuyen los errores tipo I y tipo ||
   b) Disminuye el error tipo I
   c) Disminuye el error tipo II
   d) Ninguna de las anteriores

4. Se tiene una muestra aleatoria fija $\{x_1, \ldots, x_{20}\}$ proveniente de una variable aleatoria Poisson con parámetro λ tal que $\bar{x} = 1.7$ y suponga el siguiente contraste de hipótesis

$$H_0 : \lambda = 1 \text{ vs. } H_1 : \lambda = 2.$$

Si la región de rechazo está dada por $RR = \left\{ \underline{x} : \sum_{i=1}^{20} X_i \geq 15 \right\}$, entonces el valor-$p$ asociado a los datos observados es

a) $\mathbb{P}\left(\bar{X} \geq 15\right)$
b) $\mathbb{P}\left(\bar{X} < 0.75\right)$
c) $\mathbb{P}\left(\sum X_i \geq 34\right)$
d) $\mathbb{P}\left(\sum X_i < 1.7\right)$

5. Un instructor hace un cuestionario formado por 10 preguntas falso-verdadero. Para probar la hipótesis de que el estudiante acierta por casualidad, adopta la siguiente regla de decisión: si el estudiante acierta 6 o más respuestas concluirá que no las atina por casualidad, en caso contrario, concluirá que está adivinando. En este caso

a) $\alpha \approx 0.05$
b) $\alpha \approx 0.10$
c) $\alpha \approx 0.17$
d) $\alpha \approx 0.25$

6. Encuentre la afirmación correcta

a) Un contraste de hipótesis de dos colas realizado a un nivel de significancia α es equivalente a un intervalo de confianza con un nivel (1-α)
b) Si H_1 es rechazada, entonces seguramente el valor-$p < 0.05$
c) La función potencia de una prueba permite identificar al mínimo valor a partir del cual se rechaza H_0
d) El valor-p es la probabilidad del valor del estadístico de prueba bajo H_1 para los datos observados

7. En un contexto de contraste de hipótesis considere las siguientes afirmaciones

A La región de rechazo de una prueba de hipótesis se determina en función del valor-p.
B Si $\alpha \geq 2\beta$ y valor-$p < \alpha$ entonces sabemos con certeza que se rechaza H_0.

a) A y B son ambas verdaderas.
b) A es verdadera y B es falsa.
c) A es falsa y B es verdadera.
d) A y B son ambas falsas.

8. El intervalo (4.2,8.1) representa un intervalo de 95 % de confianza para la varianza σ^2 de una población normal. Considere las siguientes afirmaciones:

   A La hipótesis $H_0 : \sigma = 2$ se rechazaría con una significancia del 10 %

   B El parámetro σ^2 está contenido en el intervalo con una probabilidad de 0.95

   a) A y B son ambas verdaderas.

   b) A es verdadera y B es falsa.

   c) A es falsa y B es verdadera.

   d) A y B son ambas falsas.

9. En un contexto de contraste de hipótesis considere las siguientes afirmaciones

   A Si el tamaño de la muestra n aumenta, la potencia de la prueba disminuye.

   B El valor-p aumenta cuando la significancia α aumenta.

   a) A y B son ambas verdaderas.

   b) A es verdadera y B es falsa.

   c) A es falsa y B es verdadera.

   d) A y B son ambas falsas.

10. Con base en una muestra aleatoria de tamaño 11 de una población normal de media μ y varianza $\sigma^2 = 5$, se desea contrastar $H_0 : \mu = 880$ vs. $H_1 : \mu \neq 880$. Si el valor del estadístico de prueba resulta -2.15, el valor-p aproximado asociado es:

    a) $p = 0.0158$

    b) $p = 0.0285$

    c) $p = 0.0316$

    d) Ninguno de los anteriores.

11. Suponga que $X \sim N(\mu, \sigma^2)$ con $\sigma^2 = 9$ y que se desea probar $H_0 : \mu \leq 10$ vs. $H_1 : \mu > 10$. Si se toma una muestra aleatoria de tamaño 30 y se observa que $\bar{x} = 11.3$, entonces es cierto que

    a) El estadístico de prueba correspondiente se distribuye asintóticamente Normal

    b) La región de rechazo está dada por $\mathcal{RR} = \left\{ \underline{x} : T(\underline{x}) > t_{(n-1),1-\alpha} \right\}$

    c) Si definimos $\alpha = 0.05$, entonces se tiene que el valor-$p \leq \alpha$

    d) Se rechaza la hipótesis nula a un nivel de significancia $\alpha \leq 0.01$

12. En un contexto de contraste de hipótesis considere las siguientes afirmaciones

   A La región de rechazo depende del valor-p

   B El valor-p aumenta cuando la significancia α aumenta.

   a) A y B son ambas verdaderas.
   b) A es verdadera y B es falsa.
   c) A es falsa y B es verdadera.
   d) A y B son ambas falsas.

13. Indique cuál de las afirmaciones siguientes es la correcta:

   a) El cociente de verosimilitudes nos permite encontrar la región de rechazo correspondiente a la prueba más potente
   b) La potencia de la prueba es una función del parámetro de interés y está dada por la probabilidad de rechazar H_0
   c) La probabilidad α (probabilidad del error tipo I) siempre es mayor que la probabilidad β (probabilidad del error tipo II)
   d) La región de rechazo aumenta conforme el nivel de significancia disminuye

14. Considere una población distribuida normal con media μ y desviación estándar $\sigma = 7$, para contrastar

$$H_0 : \mu \geq 500 \text{ vs. } H_1 : \mu < 500$$

   Se toma una muestra aleatoria de tamaño 10 y el estadístico de prueba observado fue $z_{obs} = -1.76$. Considere las siguientes aseveraciones:

   A La prueba se rechaza con una significancia del 5 %

   B El estadístico de prueba sigue una distribución t-Student

   a) A y B son ambas verdaderas.
   b) A es verdadera y B es falsa.
   c) A es falsa y B es verdadera.
   d) A y B son ambas falsas.

15. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad $f(x;\theta)$ con $\theta > 0$. Si se desea probar $H_0 : \theta = 1$ vs. $H_1 : \theta = 2$ a un nivel $\alpha = 0.05$, entonces la potencia de la prueba se define como

    a) $\mathbb{P}(\text{Rechazar } H_0 \mid \theta = 1)$
    b) $\mathbb{P}(\text{No rechazar } H_0 \mid \theta = 2)$
    c) $\mathbb{P}(\text{Rechazar } H_0 \mid \theta = 2)$
    d) 1- $\mathbb{P}(\text{Rechazar } H_0 \mid \theta = 1)$

16. Encuentre la afirmación correcta

    a) Un contraste de hipótesis de dos colas realizado a un nivel de significancia α es equivalente a un intervalo de confianza con un nivel (1-$\alpha/2$)
    b) Si H_1 es rechazada, entonces seguramente el valor-$p < 0.05$
    c) La función potencia de una prueba permite identificar al mínimo valor a partir del cual se rechaza H_0
    d) El valor-p es la probabilidad del valor del estadístico de prueba para los datos observados

17. Encuentre la afirmación correcta

    a) Una prueba de hipótesis de dos colas es equivalente a un intervalo de confianza de nivel $(1 - \alpha)$.
    b) En un contraste de hipótesis se estudia el estadístico de prueba bajo H_1.
    c) La región de rechazo no depende del nivel de significancia de la prueba.
    d) El valor-p corresponde a un valor de β específico.

18. Suponga que $X \sim N(\mu, \sigma^2)$ con $\sigma^2 = 9$ y que se desea probar $H_0 : \mu = 10$ vs. $H_1 : \mu \neq 10$. Si se toma una muestra aleatoria de tamaño 30 y el estadístico de prueba observado tiene un valor-p asociado del 2 %, entonces el valor de $\bar{x}$ observado fue aproximadamente de

    a) $\bar{x} \approx 11.27$
    b) $\bar{x} \approx 11.12$
    c) $\bar{x} \approx 8.88$
    d) Ninguna de las anteriores.

19. El salario por hora (MXN) que reciben 25 trabajadores de cierta empresa es una variable aleatoria Y tal que $Y \sim N(\mu, \sigma^2)$. Se plantean las siguientes hipótesis $H_0 : \sigma = 10$ vs. $H_1 : \sigma < 10$ y se establece que la región de rechazo está dada por $\mathcal{RR} = \{\underline{x} : s \leq 8.08\}$. ¿Qué nivel de significancia se está utilizando en la prueba?

    a) 0.001
    b) 0.025
    c) 0.05
    d) 0.1

20. En una población $N(\mu, \sigma^2)$ la potencia de la prueba

$$H_0 : \mu \geq 5 \text{ vs. } H_1 : \mu < 5$$

cuando $\mu = 3$ corresponde a

a) La probabilidad de rechazar H_0 cuando $\mu = 5$

b) La probabilidad de rechazar H_0 cuando $\mu = 3$

c) La probabilidad de no rechazar H_0 cuando $\mu = 5$

d) La probabilidad de no rechazar H_0 cuando $\mu = 3$

Preguntas abiertas

1. Un investigador afirma que un medicamento provocará sueño en, por lo menos, el 80 % de las personas que padecen insomnio. Después de un análisis de éste, alguien considera que la afirmación es exagerada y para refutarla se administra la medicina a 20 personas elegidas al azar que padecen insomnio. Sea Y el número de personas que logran dormir con las medicina y la región de rechazo $RR = \{y : Y \leq 12\}$.

   a) Calcule α

   b) Calcule β para $p = 0.6$ y $p = 0.4$

   c) Si la región de rechazo $RR = \{y : Y \leq c\}$, encuentre c tal que $\alpha = 0.01$

   d) Para la región de rechazo $RR = \{y : Y \leq 11\}$, encuentre β si $p = 0.6$

2. En un estudio estomatológico, 68 de 160 niños considerados en la muestra recibieron un tratamiento de fluoruro y tuvieron caries. Otro tratamiento químico dio como resultado que 38 de 110 niños, seleccionados al azar, tuvieran caries. ¿Se puede afirmar, con un nivel de significancia de 0.05, que el tratamiento con fluoruro dio peor resultado que el tratamiento químico? ¿Cuál es el valor-p?.

3. Se sabe que el peso de los contenedores marítimos que se apegan a las prácticas internacionales se distribuye como una Normal con desviación estándar de 5 toneladas. A partir de una muestra aleatoria grande de contenedores se quiere probar $H_0 : \mu = 30$ vs. $H_1 : \mu > 30$. ¿Qué tamaño de la muestra garantiza $\alpha = \beta(32) = 0.01$?

4. Con base en una muestra aleatoria de tamaño 11 de una población normal de media μ y varianza $\sigma^2 = 5$, se desea contrastar $H_0 : \mu = 880$ vs. $H_1 : \mu \neq 880$. Si el valor del estadístico de prueba resulta -2.15, ¿cuál es el valor-p?

5. Considere una población distribuida normal con media μ y desviación estándar $\sigma = 7$, para contrastar

   $$H_0 : \mu \geq 500 \text{ vs. } H_1 : \mu < 500.$$

   Calcule el tamaño de muestra n si el estadístico de prueba observado fue $z_{obs} = -1.76$.

6. Sea $X_1, \ldots, X_n$ una muestra aleatoria de $N(\mu, \sigma^2)$ y considere el siguiente contraste de hipótesis

   $$H_0 : \sigma^2 = \sigma_0^2 \text{ vs. } H_1 : \sigma^2 = \sigma_1^2 \qquad \text{con } \sigma_1^2 > \sigma_0^2.$$

   a) Determine la región de rechazo correspondiente a la prueba más pontente para un nivel de significancia α.

   b) Considere que $n = 20$, $\sigma_0^2 = 1$ y $\alpha = 5\,\%$ para encontrar la constante k del inciso anterior.

7. Una empresa de investigación de mercados está realizando un estudio para evaluar el desempeño de una campaña publicitaria. El objetivo de la campaña es impulsar el volumen de ventas de cierto producto de la empresa A. Los analistas consideran que una forma sencilla de identificar de manera directa un posible impacto es mediante el incremental en ventas después del lanzamiento, es decir,

$$D_i = Y_i - X_i, \quad i = 1, \dots, 101$$

donde Y_i son las ventas del producto en la i-ésima tienda **después** del lanzamiento y X_i son las ventas del producto en la i-ésima tienda **antes** del lanzamiento. Para el estudio se están considerando 101 sucursales y los expertos consideran que la campaña es exitosa si el incremental en ventas es superior a 3 millones MXN.

   a) Determine las hipótesis adecuadas para este problema. Explique su razonamiento y especifique los supuestos necesarios.
   b) Encuentre el estadístico de prueba adecuado bajo H_0.
   c) Encuentre la región de rechazo correspondiente si se realiza una prueba a un nivel de significancia $\alpha = 0.05$.
   d) Si la media muestral del incremental en ventas fue de 5.3 millones MXN y la varianza muestral fue de 144, determine el valor-p y concluya: ¿la campaña logró su objetivo?.

8. En un estudio se desea comparar la opinión entre hombres y mujeres respecto a la ley antitabaco a casi 10 años de que entró en vigor dicha ley. Para lograr lo anterior se realizaron entrevistas cara a cara, de manera independiente, a 150 hombres y 200 mujeres en la Ciudad de México. De manera general, se reportó que 27 hombres y 128 mujeres se pronunciaron en contra.

   a) Determine las hipótesis adecuadas para este problema. Explique su razonamiento y especifique los supuestos necesarios.
   b) Encuentre el estadístico de prueba adecuado bajo H_0.
   c) Encuentre la región de rechazo correspondiente si se realiza una prueba a un nivel de significancia $\alpha = 0.01$.
   d) Utilice los resultados reportados en el estudio para calcular el valor-p y elabore una conclusión.

9. Una casa de encuestas desea realizar un estudio para determinar la intención de voto para las elecciones presidenciales en Estados Unidos en 2020. Para lograr lo anterior, levanta una encuesta en un punto de afluencia y comienza a preguntar a los individuos por el partido que votará hasta que encuentra a la primera persona que declara que votará por el partido demócrata.

   El proceso anterior se modela con una distribución geométrica cuya función que está dada por

$$f(x_i;\theta) = \theta(1-\theta)^{x_i}\, \mathbb{I}_{\{0,1,2,\dots\}}(x_i), \quad \text{con } 0 < \theta < 1.$$

   Uno de los analistas de la casa encuestadora considera que el partido demócrata tiene más posibilidades de ganar por lo cual se plantea el siguiente contraste de hipótesis

$$H_0 : \theta = \frac{1}{2} \text{ vs. } H_1 : \theta = \frac{3}{4}.$$

   Determine la región de rechazo correspondiente a la prueba más pontente a un nivel de significancia $\alpha = 0.05$ suponiendo que el tamaño de muestra fue $n = 100$.

   Hint Si $X_i \sim Geom(p)$ entonces $\sum^{n} X_i \sim BinNeg(n,p)$

10. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población definida por

$$f(x_i;\theta) = \theta x_i^{\theta-1}\mathbb{I}_{(0,1)}(x_i), \quad \theta > 0.$$

   Se desea probar $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, con $\theta_1 > \theta_0$.

   a) Mediante el Lema de Neyman-Pearson, demuestre que la región de rechazo de la correspondiente prueba uniformemente más potente es de la forma

$$RR = \left\{ -\sum_{i=1}^{n} \log(X_i) \leq k^*_\alpha \right\}.$$

   b) Demuestre que para una prueba de tamaño α, el valor crítico de la región de rechazo del inciso anterior es

$$k = \frac{\mathcal{X}^2_{(2n),\alpha}}{2\theta_0}.$$

   Hint Primero demuestre que $-\log(X_i) \sim \text{Exp}(\frac{1}{\theta})$ y luego recuerde que si $S \sim \text{Gamma}(\alpha, \beta)$ entonces $\frac{2S}{\beta} \sim \mathcal{X}^2_{(2\alpha)}$.

   c) Suponga que X_i denota el porcentaje de daño que sufre un hotel en caso de un tsunami. A partir de una muestra aleatoria de 5 hoteles del sureste asiático afectados por tsunamis, se observó que los daños fueron de 60 %, 75 %, 68 %, 80 % y 92 %.

   Un reportero observa estos datos y publica una nota diciendo que «la media poblacional del daño que sufren los hoteles en caso de tsunami es mayor a 75 %». Realice una prueba de hipótesis de tamaño 5 % para determinar si la afirmación del reportero es correcta.

   Hint Note que $E[X_i] = \frac{\theta}{\theta+1}$ para plantear adecuadamente su hipótesis.

11. Un psicólogo piensa que la edad influye en el coeficiente de inteligencia (IQ). Se toma una muestra aleatoria de 100 personas de mediana edad, de quienes se conoce su IQ a la edad de 16 años y actualmente, De restar los coeficientes en su juventud y en la actualidad, se obtuvo una diferencia promedio de 6 puntos, con una desviación estándar muestral de 7 puntos. Utilice $\alpha = 0.01$ para probar la hipótesis de que el IQ aumenta con la edad.

12. La profesora de cierto grupo asegura que sus alumnos obtienen, *en promedio*, mejores calificaciones en el segundo parcial comparado con el primero. El coordinador del curso seleccionó al azar 10 alumnos y registró las siguientes calificaciones (sobre 100 puntos)

Alumno	1	2	3	4	5	6	7	8	9	10
Parcial 1	82	66	48	65	41	79	52	90	55	53
Parcial 2	98	65	39	68	56	49	69	100	57	51

   a) ¿La información anterior presenta evidencia de lo que aseveró la profesora?. Realice una prueba de hipótesis con un nivel de significancia de $\alpha = 0.05$ y concluya.

   b) Si para el coordinador los resultados únicamente son válidos si el valor-$p \leq 0.01$ ¿El coordinador aceptará la afirmación de la profesora? Argumente.

13. Un estudio sobre las deficiencias de hierro en la sangre de los bebés comparó un grupo de bebés cuyas madres eligieron alimentarlos exclusivamente con leche materna, con otro grupo en el que fueron alimentados exclusivamente con fórmula láctea. La información sobre los niveles de hemoglobina en la sangre de los bebés es la siguiente

grupo	n	media	varianza
Leche materna	26	13.3	1.7
Fórmula láctea	31	12.4	3.3

   Los pediatras que llevaron a cabo el estudio sospechan que hay mayor variabilidad en los contenidos de hemoglobina de los bebés alimentados con fórmula. Pruebe esta afirmación utilizando un contraste de hipótesis con un nivel de significancia de 0.05.

14. La agencia de Mazda Picacho desea monitorear el proceso de pintado de cofres de la flotilla de Mazda 6 que acaba de llegar a la agencia. Sea X el número de defectos de pintura detectados en el cofre por el inspector y suponga que $X \sim Po(\lambda)$

   a) Si se tiene una muestra aleatoria $X_1, \ldots, X_n$, obtenga la prueba más potente para probar

$$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda = \lambda_1 \qquad \text{con} \qquad \lambda_0 < \lambda_1$$

   b) En el proceso productivo, se toma una muestra aleatoria de tamaño 2 cada hora. Obtenga la región de rechazo con un nivel de significancia lo más cercano a 5 % si $\lambda_0 = 1.6$ defectos por cofre.

   c) En la producción de 10:00 a 11:00 se observaron 2 y 3 defectos de pintura respectivamente. ¿Existe evidencia para concluir λ aumentó?

15. Una marca americana de productos dermatológicos ha comenzado a vender en la Ciudad de México. Un fotoprotector en gel que venden se envasa directamente en el punto de venta. Actualmente hay dos sucursales: *Perisur* y *Antara*. Sea $X_1, \ldots, X_n$ una muestra aleatoria del proceso de llenado en la sucursal de *Antara*. Se desea contrastar la hipótesis

$$H_0 : \mu \geq 100 \text{ ml. } vs. \ H_1 : \mu < 100 \text{ ml.},$$

considerando la región de rechazo

$$RR = \left\{ \bar{X} \geq k \right\}.$$

Determine el tamaño de muestra n y el valor crítico k tales que, la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera sea menor que 0.05, y la probabilidad de no rechazar la hipótesis nula es a lo más de 0.10 cuando el proceso de llenado tiene en realidad $\mu = 98$ ml y $\sigma = 3$.

16. Sea $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ muestras aleatorias independientes tales que $X_i \sim Exp(\theta)$ y $Y_i \sim Exp(\mu)$. Encuentre la región de rechazo correspondiente a la prueba de cociente de verosimilitudes para

$$H_0 : \theta = \mu \ vs. \ H_1 : \theta \neq \mu$$

17. Una reconocida marca de joyería de lujo lanzará próximamente una nueva línea de anillos de compromiso y tiene que elegir entre dos modelos, Margot y Camille, para que representen la imagen de los nuevos productos en la campaña publicitaria. Para elegir a la modelo se deben considerar los sueldos que cobrarían. Una agencia francesa maneja a Camille mientras que una inglesa a Margot. En general, se puede suponer que los sueldos de las modelos (en miles de dólares) se distribuyen Normal con la misma variabilidad. Se tomaron muestras aleatorias de los sueldos de las modelos de ambas agencias y se obtuvo que

agencia	n	media	desviación estándar
francesa	11	14.7	3.9
inglesa	8	12.4	3.1

¿Se puede considerar que las agencias tienen sueldos medios iguales?. Argumente su respuesta en términos del valor-p.

18. Los fabricantes de una bebida refrescante embotellada decidieron cambiar la fórmula de su producto. Varios consumidores se mostraron en contra del cambio y la compañía convino que si más de la mitad de ellos estaban en desacuerdo, volverían a la formula anterior. Se entrevistaron a 937 personas y 531 se manifestaron en contra. Realice un contraste de hipótesis a un nivel de significancia de 0.05 y determine qué debería hacer dicha compañía.

19. Una empresa de tiendas de conveniencia, con 417 establecimientos, desea evaluar el impacto de una campaña publicitaria en las ventas de una línea de productos. Para ello obtuvo una muestra de manera aleatoria, con reemplazo y probabilidades iguales de tamaño 10. De manera piloto, únicamente se hizo el despliegue de la campaña publicitaria en esas 10 sucursales. Para medir el impacto, se midieron las ventas del producto en el mes previo y posterior al despliegue de la campaña. Además, se eligieron meses en los cuales las ventas de los productos son similares, para evitar el efecto temporal de las ventas.

    a) Considere el siguiente procedimiento para evaluar el impacto de la campaña: si la diferencia entre las ventas posterior-previo para la sucursal i fue positiva, entonces $Y_i = 1$ y cero en otro caso. Sea $W = \sum Y_i$. Obtenga la distribución de probabilidad de W cuando no hay diferencia en las ventas debido a la campaña.

    b) Utilice W para probar la hipótesis sobre si la campaña tiene un impacto positivo sobre las ventas con un nivel de significancia lo más cercano a 5 %. Especifique la hipótesis nula, la hipótesis alternativa, la región de rechazo y concluya si $W = 8$.

20. La Secretaría de Seguridad Pública de la Ciudad de México desea probar un nuevo radar detector de velocidades. Para hacer esto, hace circular frente al radar 61 patrullas a una velocidad de 80km/h. De las 61 lecturas del radar se obtiene $s^2 = 1.69$. Según la SSP-CDMX el radar es aceptable si $\sigma < 2$km/h. Adopte el siguiente punto de vista: *El error más grave es comprar el radar cuando en realidad el error de medición está por arriba del límite*. Formule las hipótesis y realice su contraste considerando un nivel de significancia de 1 %.

21. El dueño de una empresa quiere donar dinero para la construcción de una guardería. Tiene en mente dos zonas de la Ciudad de México (A y B), por lo que entrevista de manera aleatoria a 200 madres de la zona A, de las cuales el 52 % trabajan tiempo completo, mientras que en la zona B, el 40 % de las 150 madres entrevistadas trabaja tiempo completo.

    a) Determine si existe diferencia significativa en la proporción de madres con trabajo de tiempo completo en las dos zonas, al nivel de significancia de 0.04. Justifique su respuesta.

    b) Si el dueño basa su decisión de construir la guardería en la proporción de madres que trabajan de tiempo completo, ¿cuál sería la zona elegida?

    c) En la entrevista a las madres que trabajan, también se les preguntó el salario por día. El salario promedio diario de las que trabajan tiempo completo en la zona A es de 79.50 pesos y en la zona B es de 82 pesos, con desviación estándar de 5.38 y 4.16, respectivamente. Note que los tamaños de muestra consideran únicamente a las madres que trabajan tiempo completo, es decir, 104 en la zona A y 60 en la zona B. ¿Se podría decir que la variación en los salarios es la misma?

    d) ¿Se podría decir que las diferencias en los salarios medios de las madres que trabajan tiempo completo se deben a causas aleatorias, o son mayores los salarios de las madres en la zona B que los de la zona A? Justifique su respuesta.

    e) Si el dueño basa su decisión en el salario medio de las madres que trabajan tiempo completo, ¿en dónde decidiría construir la guardería? ¿Concuerda esto con la decisión tomada en el inciso b)?

22. Sea $X_1, \ldots, X_n$ una muestra aleatoria de variables $Gamma(1, \beta)$ y suponga que

$$H_0 : \beta = 2 \text{ vs. } H_1 : \beta = 4$$

a) Determine la región de rechazo correspondiente a la prueba más pontente si $\alpha = 0.05$ y $n = 10$.

b) Considere la muestra $\underline{X} = \{1.5, 1.4, 1.9, 1.7, 2.1, 1.8, 1.6, 2.0, 1.6, 1.7\}$. Obtenga el valor-$p$ correspondiente y concluya.

23. Considere el caso donde se tienen dos clasificadores (falibles) de objetos, C_a y C_b. Cada objeto de una muestra aleatoria (i.i.d.) de tamaño m es sometida a clasificación con ambos clasificadores, C_a y C_b, simultáneamente, obteniendo los siguientes resultados:

$$\begin{aligned}
m_{00} &= \text{número de muestras mal clasificadas por } C_a \text{ y } C_b, \\
m_{10} &= \text{número de muestras mal clasificadas por } C_a \text{ pero no por } C_b, \\
m_{01} &= \text{número de muestras mal clasificadas por } C_b \text{ pero no por } C_a, \\
m_{11} &= \text{número de muestras clasificadas correctamente por } C_a \text{ y } C_b.
\end{aligned}$$

También observamos las tasas de error de clasificación de cada clasificador, dadas por

$$\begin{aligned}
p_a &= \frac{m_{00} + m_{10}}{m}, \\
p_b &= \frac{m_{00} + m_{01}}{m},
\end{aligned}$$

para C_a y C_b, respectivamente. Estos cocientes pueden interpretarse como los estimadores puntuales de las tasas error de los clasificadores C_a y C_b, respectivamente.

Considere el escenario donde se desea contrastar la hipótesis nula H_0 referente a que ambos clasificadores tengan la misma tasa de error. Para ésto, se propone el estadístico bajo H_0 dado por

$$t = \frac{p_a - p_b}{\left(2p(1-p)/m\right)^{1/2}},$$

donde $p = (p_a + p_b)/2$ es la tasa de error de clasificación promedio entre los dos clasificadores.

a) Especifique los supuestos que se emplearon para definir el estadístico de prueba t

b) ¿Son p_a y p_b independientes?

c) Determine una distribución asintótica para t bajo H_0

d) Obtenga la región de rechazo correspondiente

24. Una empresa mexicana dedicada a la venta de bebidas carbonatadas tiene un producto que representa más del 60 % de sus ventas totales, por lo que se le denomina producto estella. Dicha empresa realizó un estudio de mercado y encontró que los consumidores valoran más las bebidas con menor cantidad de azúcares. Por ello, decide lanzar una versión sin azúcar de su producto estrella.

    El nuevo prototipo, a pesar de contener menos azúcares, se produjo de manera que el sabor fuera similar a la versión original. Además, el empaque y las presentaciones disponibles son las mismas que para la versión original. Los directivos de la empresa quieren evaluar si el consumidor es capaz de percibir la diferencia entre los dos productos (original y prototipo).

    Para lograr lo anterior, realizaron una prueba ciega con 35 consumidores frecuentes y les hiceron algunas preguntas, entre ellas el gusto general y la calificación sobre el sabor de ambos productos. Estas preguntas se midieron en una escala del 1 al 10, donde 1 es no me gustó en lo absoluto y 10 es me gustó muchísimo.

    Suponga que se considera como métrica el promedio de las calificaciones. Tomando en cuenta únicamente el gusto en sabor, los analistas de la compañía deciden dar respuesta a la pregunta de los directivos planteando el problema mediante un contraste de hipótesis.

    a) Determine las hipótesis adecuadas para este problema. Explique su razonamiento.
    b) Encuentre el estadístico de prueba adecuado bajo H_0 y especifique los supuestos necesarios.
    c) Encuentre la región de rechazo correspondiente si se realiza una prueba a un nivel de significancia $\alpha = 0.075$.
    d) Los resultados del estudio para el producto original fueron $\bar{x} = 8.2$ y $s_X = 0.7$, mientras que para el prototipo fueron $\bar{y} = 7.8$ y $s_Y = 1.1$. Calcule el valor-p y con base en este resultado dé una recomendación para los directivos.

25. En 1000 lanzamientos de una moneda (águila y sol), 560 lanzamientos fueron soles. ¿Es razonable pensar que la moneda empleada fue una moneda justa? Justifique su respuesta realizando un contraste de hipótesis determinando todos los elementos y supuesto necesarios.