Ejercicios adicionales de Estadística Inferencial

Giovanni Sanabria Brenes & Félix Núñez Vanegas

1 Estimación puntual y verosimilitud

Ejemplo 1 *Sea X y Y dos variables aleatorias independientes tales que*

$$X \sim B\left(10\theta, \frac{1}{2}\right), \quad Y \sim B\left(20\theta, \frac{1}{5}\right)$$

Considere los siguientes estimadores de un parámetro $\theta > 0$ de una determina población:

$$\widehat{\theta}_1 = X - Y, \qquad \widehat{\theta}_2 = \frac{X+Y}{9}$$

¿Cuál de estos es el mejor estimador del parámetro θ? Justifique su respuesta.

1. ***¿Cuál estimador es insesgado?*** *Recuerde que si estádístico es insesgado y el otro no, es mejor estimador, aquel que sea insesgado. En este caso, usando las propiedades de esperanza vistas en probabilidad, se tiene que:*

$$E\left(\widehat{\theta}_1\right) = E(X-Y) = E(X) - E(Y) = 10\theta \cdot \frac{1}{2} - 20\theta \cdot \frac{1}{5} = 5\theta - 4\theta = \theta$$

Por lo tanto, $\widehat{\theta}_1$ es un estimador insesgado de θ. Note que el cálculo anterior se utilizó la esperanza de la Binomial ($W \sim B(n,p) \Longrightarrow E(W) = np$).

Por otro lado, note que

$$E\left(\widehat{\theta}_2\right) = \frac{1}{9}E(X+Y) = \frac{1}{9}\left[E(X) + E(Y)\right] = \frac{1}{9}\left[5\theta + 4\theta\right] = \theta$$

Así, $\widehat{\theta}_1$ es un estimador insesgado de θ. Ambos son estimadores insesgados, pero ¿cuál es el mejor?

2. ***¿Cuál estimador varía menos?*** *En este caso, usando las propiedades de varianza vistas en probabilidad, se tiene que:*

$$\begin{aligned} Var\left(\widehat{\theta}_1\right) &= Var(X-Y) = Var(X + -Y) = Var(X) + (-1)^2 Var(Y) \\ &= Var(X) + Var(Y) = 10\theta \cdot \frac{1}{2} \cdot \frac{1}{2} + 20\theta \cdot \frac{1}{5} \cdot \frac{4}{5} = \frac{57}{10}\theta \end{aligned}$$

Similarmente:

$$Var\left(\widehat{\theta}_2\right) = \frac{1}{81} Var\left(X+Y\right) = \frac{1}{81}\left[Var\left(X\right) + Var\left(Y\right)\right] = \frac{1}{81}\cdot\frac{57}{10}\theta$$

Entonces, $\widehat{\theta}_2$ *tiene menor variancia. Por lo tanto,* $\widehat{\theta}_2$ *es el mejor estimador de* θ.

Ejemplo 2 *La variable aleatoria* X *tiene densidad de probabilidad dada por*

$$f\left(x\right) = \frac{x}{a^2 e^{x/a}},\ \text{para } x \geq 0$$

Tres observaciones de X *son* $x_1 = 12; x_2 = 16$ *y* $x_3 = 15$. *Encuentre la estimación de máxima verosimilitud del parámetro* a.

1. ***Determinación de la función de verosimilitud*** g. *En este caso la muestra observada es* $(12, 16, 15)$ *por lo tanto la función de verosimitud es*

$$g\left(a\right) = f\left(12\right) f\left(16\right) f\left(15\right) = \frac{12}{a^2 e^{12/a}} \frac{16}{a^2 e^{16/a}} \frac{15}{a^2 e^{15/a}} = \frac{2880}{a^6 e^{43/a}}$$

2. ***Simplificación de*** $\ln g\left(a\right)$. *Se busca el valor de* a *que maximice* g. *Para ello lo mejor es primeramente aplicar logaritmo a* g :

$$\ln g\left(a\right) = \ln 2880 - 6\ln a - \frac{43}{a}$$

3. ***Calcular*** $g'\left(a\right)$. *Derivando la expresión anterior se tiene que*

$$\frac{g'\left(a\right)}{g\left(a\right)} = -\frac{6}{a} + \frac{43}{a^2} \qquad \Longrightarrow \qquad g'\left(a\right) = \left(-\frac{6}{a} + \frac{43}{a^2}\right) g\left(a\right)$$

4. ***Estimación de*** a. *Para que* $g'\left(a\right) = 0$ *basta que* $-\frac{6}{a} + \frac{43}{a^2} = 0$ *o que* $g\left(a\right) = 0$, *dado que se busca el valor máximo de* $g\left(a\right)$ *y que* $g\left(a\right) \geq 0$, *pues es una multiplicación de probabilidades, se descarta la ecuación* $g\left(a\right) = 0$, *ya que está busca el valor de* a *que hace mínimo a* g. *Así,*

$$-\frac{6}{a} + \frac{43}{a^2} = 0 \Longrightarrow a = \frac{43}{6}$$

Por lo tanto, la estimación de máxima verosimilitud del parámetro a *es* $\frac{43}{6}$. *El lector puede verificar que en este valor la función* g *alcanza su máximo.*

Realice los siguientes ejercicios:

1. Suponga que $F \sim f(6,17)$. Acote lo mejor posible, utilizando las tablas, la probabilidad $P(F > 6)$ $R/ \quad]0, 0.01[$

2. Considere la población dada por la variable aleatoria X con media poblacional μ y varianza σ^2. Dada una muestra aleatoria de esta población de tamaño $n > 1 : (X_1, X_2, X_3, ..., X_n)$, considere los siguientes estadísticos de σ^2

$$S_1^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1} \quad y \quad S_2^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}$$

   (a) Exprese S_2^2 en términos de S_1^2 y de n.

$$R/ \quad S_2^2 = \frac{n-1}{n} S_1^2$$

   (b) Si S_1^2 es un estadístico insesgado de σ^2, ¿es S_2^2 un estimador insesgado de σ^2? $R/ \quad No$

3. Una variable aleatoria X tiene una distribución con densidad $f(x) = \dfrac{\lambda^2 (x-5)}{e^{\lambda(x-5)}}$ para $x \geq 5$, con λ un parámetro. Dadas las observaciones $x_1 = 6$, $x_2 = 8$ y $x_3 = 11$, encuentre las estimaciones de máxima verosimilitud de λ. $R/ \quad \frac{3}{5}$

4. Una variable aleatoria X tiene una distribución geométrica de parámetro p (la densidad es $f(x) = p(1-p)^x$ para $x = 0, 1, 2, ...$). Dadas las observaciones $x_1 = 6$, $x_2 = 8$, $x_3 = 8$ y $x_4 = 9$, encuentre las estimaciones de máxima verosimilitud de p. $R/ \quad \frac{4}{35}$

5. El tiempo que un estudiante tarda en cierto examen es una variable aleatoria T con función de densidad $f(t) = (\theta + 1) t^\theta$ para $0 \leq t \leq 1$ con $\theta > -1$. Cuatro estudiantes seleccionados al azar tardan $t_1 = 0.79$, $t_2 = 0.65$, $t_3 = 0.47$ y $t_4 = 0.97$. Encuentre la estimación de máxima verosimilitud de θ. $R/ \quad 1.754\,85$

6. Una variable aleatoria X tiene una distribución Gamma$(2, \beta)$ con β un parámetro (la densidad es $f(x) = \dfrac{xe^{-x/\beta}}{\beta^2}$ para $x > 0$). Dadas las observaciones $x_1 = 1$, $x_2 = 2$, y $x_3 = 7$, encuentre la estimación de máxima verosimilitud de β. $R/ \quad \beta = \frac{5}{3}$

7. Una variable aleatoria X tiene una distribución con densidad $f(x) = \dfrac{k}{7}\left(\dfrac{x}{7}\right)^{k-1}$ para $0 \leq x \leq 7$, con k constante.

   (a) Dadas las observaciones $x_1 = 2.8$, $x_2 = 3.5$ y $x_3 = 1.4$, encuentre la estimación de máxima verosimilitud de k. $R/ \quad 0.932\,002$

   (b) Dadas las observaciones $x_1 = 1.4$, $x_2 = 0.7$ y $x_3 = 3.5$, encuentre la estimación de máxima verosimilitud de k. $R/ \quad 0.651\,442$

8. Sea X una variable aleatoria tal que $X \sim Exp(\lambda)$, donde λ es un parámetro desconocido. Es decir
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$
   (a) Dada una muestra aleatoria de $X : (X_1, X_2, ..., X_n)$, determine el estadístico $\widehat{\lambda}$ que brinda el método de máxima verosimilitud para estimar λ. $R/ \quad \widehat{\lambda} = \dfrac{1}{\overline{X}}$
   (b) Pruebe que $\dfrac{1}{\widehat{\lambda}}$ es un estimador insesgado de $\dfrac{1}{\lambda}$.Sugerencia: recuerde que si $X \sim Exp(\lambda)$ entonces $E(X) = \dfrac{1}{\lambda}$

9. Sea X una variable aleatoria tal que $X \sim G(p)$, donde p es un parámetro desconocido. Es decir
$$f_X(k) = (1-p)^k p, \text{ para } k = 0, 1, ...$$
   (a) Dada una muestra aleatoria de $X : (X_1, X_2, ..., X_n)$, determine el estadístico $\widehat{P_1}$ que brinda el método de máxima verosimilitud para estimar p. $R/ \quad \widehat{P_1} = \dfrac{1}{1+\overline{X}}$
   (b) Pruebe que $\dfrac{1}{\widehat{P_1}}$ es un estimador insesgado de $\dfrac{1}{p}$. Sugerencia: recuerde que si $Y \sim G(p)$ entonces $E(Y) = \dfrac{1-p}{p}$

10. Sea X una variable aleatoria tal que $X \sim U[0, a]$, donde a es un parámetro desconocido. Es decir
$$f_X(k) = \begin{cases} \dfrac{1}{a} & si \quad x \in [0, a] \\ 0 & \text{en otro caso} \end{cases}$$
Dada una muestra aleatoria de $X : (X_1, X_2, ..., X_n)$, demuestre que $A = 2\overline{X}$ es un estimador insesgado de a.

11. Sea X una variable aleatoria tal que $X \sim Gamma(3, \beta)$ es decir
$$f_X(x) = \begin{cases} \dfrac{x^2 e^{-x/\beta}}{2\beta^3} & si \quad x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$
donde β es un parámetro desconocido.
   (a) Dada una muestra aleatoria de $X : (X_1, X_2, ..., X_n)$, determine el estadístico $\widehat{\beta}$ que brinda el método de máxima verosimilitud para estimar β. $R/ \quad \widehat{\beta} = \dfrac{1}{3}\overline{X}$
   (b) Pruebe que $\widehat{\beta}$ es un estimador insesgado de β. Sugerencia: recuerde que si $X \sim Gamma(\alpha, \beta)$ entonces $E(X) = \alpha\beta$.

12. Sean $a, b \in \mathbb{R}$. Sea X una variable aleatoria tal que $P(a \leq X \leq b) = p$. ¿Cómo estimar p utilizando máxima verosimilitud? Para ello realice lo siguiente:

   (a) Sea $Y = \begin{cases} 1 & si \quad X \in [a,b] \\ 0 & si \quad no \end{cases}$. Determine la distribución de probabilidad para Y.

   $$R/ \quad f_Y(k) = \begin{cases} p & si \quad k = 1 \\ 1-p & si \quad k = 0 \end{cases}$$

   (b) Suponga que se toma una muestra aleatoria de $X : (x_1, x_2, ..., x_n)$ y que de estas observaciones hay m que se encuentran en $[a, b]$. Esta muestra genera n observaciones de $Y : (y_1, y_2, ..., y_n)$, utilice la distribución de Y para hallar la estimación de máxima verosimilitud de p.

   (c) Compare la estimación de máxima verosimilitud de p con la estimación dada por el estadístico $\widehat{P}$.

13. Sea X una variable aleatoria tal que $X \sim N(0, \sigma^2)$ es decir

$$f_X(x) = \frac{e^{-x^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

donde σ es un parámetro desconocido. Considere una muestra aleatoria de $X : (X_1, X_2, ..., X_n)$.

   (a) Pruebe que el estadístico D^2 que brinda el método de máxima verosimilitud para estimar σ^2 es .

   $$D^2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

   (b) Pruebe que $E(X_i^2) = \sigma^2$ para $i = 1, 2, 3, ..., n$.

   (c) ¿Es D^2 un estimador insesgado de σ^2? $R/$ Si

14. Utilice el app Probability Distribution para estimar los siguientes valores:

   (a) $\chi^2_{0.04,30}$ R/ 17.9082

   (b) $t_{0.13,15}$ R/ -1.170656

   (c) $1 - z_{0.8}$ R/ 0.158379

   (d) $f_{0.975,4,14}$ R/ 3.891914

15. Determine las probabilidades solicitadas (Use el app Probability Distributions)

   (a) Si $\chi^2 \sim \chi^2(22)$, estime $P(\chi^2 > 14.86)$ R/ 0.868176

   (b) Si $t \sim t(15)$, estime $P(t < -1.14)$ R/ 0.136082

   (c) Si $z \sim N(0,1)$, estime $P(z > 0.16 \vee z < -0.16)$ R/ 0.872881

   (d) Si $f \sim f(14, 18)$, estime $P(f < 2.16 \vee f > 3.14)$ R/ 0.94937

(e) Si $\chi^2 \sim \chi^2(30)$, estime $P\left(\chi^2 < 14.86 \vee \chi^2 > 16.15\right)$ R/ 0.990931

16. Considere $\widehat{P}$ y $\widehat{Q}$ dos estimadores insesgados independientes de un cierto parámetro θ con varianzas $a\sigma^2$ y $b\sigma^2$, con a y b constantes reales positivas, respectivamente. Considere el estimador $\widehat{R} = t\ \widehat{P} + (1-t)\,\widehat{Q}$,con $t \in [0,1]$.

    (a) Demuestre que $\widehat{R}$ es un estimador insesgado del parámetro θ, $\forall t \in [0,1]$.

    (b) Calcule el valor de t de modo que $\widehat{R}$ tenga la menor varianza.

    (c) Sin hacer más cálculos, si $a = 1$ y $b = 5$, ¿Cuál de los siguientes estimadores insesgados de θ es mejor: $\widehat{R} = \frac{5}{6}\widehat{P} + \frac{1}{6}\widehat{Q}$ ó $\widehat{T} = \frac{2}{3}\widehat{P} + \frac{1}{3}\widehat{Q}$? R/ $\widehat{R}$.

2 Intervalos de Confianza para una población

1. Actualmente las personas debido a sus ocupaciones dedican poco tiempo a bañarse antes de ir al trabajo. Una médico señala que una persona se baña un tiempo saludable si tarda al menos 15 minutos en esta actividad. Seguidamente se presentan el tiempo en minutos que tarda 15 personas en bañarse antes de ir al trabajo, tomadas al azar de la ciudad C

$$10, 6, 8, 7, 10, 11, 15, 20, 17, 17, 16, 5, 12, 8, 9$$

Sea X : tiempo que tarda en bañarse una persona elegida al azar de la ciudad C que labora. Suponga que $X \sim N$.

   (a) Determine un IC del 90% para el promedio en minutos del tiempo que tardan en bañarse las personas de la ciudad C que laboran. $R/ \quad]9.31741, 13.4826[$

   (b) ¿Considera que las personas de la ciudad C que laboran, en promedio, tardan un tiempo saludable en bañarse? Justifique. $R/ \quad No$

   (c) Determine un intervalo de confianza de 95% para **la proporción p** de personas de la ciudad C que tardan un tiempo saludable en bañarse antes de ir al trabajo. $R/$ $]0.152077 , 0.647923[$

   (d) ¿Considera que el IC hallado es muy grande? Determine el tamaño de la muestra si se quiere un IC del 95% para p con un radio menor a la cuarta parte del radio del IC hallado en (c), sin importar el verdadero valor de p. $R/ \quad 250$

2. Una persona denuncia ante la Oficina del Consumidor que el peso de una bolsa de azúcar Dulce, con peso marcado de 2 kg, suele ser muy variable. Ante esto, un Inspector de la Oficina del Consumidor desea determinar si la desviación estándar de los pesos de las bolsas es superior a $50g$, si esto sucede sancionará a la empresa Dulce S.A. En una muestra al azar de 20 bolsas de azúcar Blanco se observó una desviacíon estándar de 0.105, a partir de estos el inspector obtuvo un IC para la variancia de los pesos de la bolsa de azúcar Blanco en kilogramos.

$$]0.00578806, 0.0274443[$$

Suponga que los pesos de las bolsas de arroz Blanco siguen una distribución normal.

   (a) ¿Es confiable el IC determinado por el inspector. Determine aproximadamente su nivel de confianza. $R/ \quad 98\%$

   (b) ¿Aceptaría que la empresa Blanco reciba una sanción de parte de la Oficina del Consumidor? Justifique su respuesta. $R/ \quad]0.0760793, 0.1165663[$

3. Una línea aérea afirma que el tiempo de retraso en la salida de sus vuelos tiene un promedio menor que 10 minutos. En una muestra de siete vuelos se registran los siguientes retrasos, en minutos: 11, 7, 13, 9, 17, 8 y 12. Suponga que los retrasos en la salida de los vuelos siguen una distribución normal.

   (a) Encuentre un intervalo de confianza del 90% para el tiempo de retraso promedio. $R/ \quad]8.49137, 13.5086[$

(b) ¿Ese intervalo es evidencia en contra de la afirmación de la empresa? ¿Por qué? $R/\quad No$

4. Se desea estimar el porcentaje de habitantes de una ciudad C que no se lavan los dientes regularmente. Con una muestra de tamaño 80 se observó que menos de la mitad no se lavan los dientes regularmente y se obtuvo un IC del 95% cuyo radio fue de 0.095. Determine aproximadamente cuántas personas en la muestra no se lavan los dientes regularmente. $R/\quad \frac{1}{4}$

5. Para controlar el buen embolsado de sus productos, un productor de fertilizantes toma una muestra de 15 bolsas del mismo, obteniendo una desviación estándar de 0.50 kg. Suponga que los pesos de las bolsas siguen una distribución normal.

   (a) Determine un intervalo de confianza del 98% para la variancia de los pesos de las bolsas de fertilizante. $R/\quad]0.120\,105\,, 0.751\,004[$

   (b) ¿Es razonable suponer que la desviación estándar de los pesos de las bolsas de fertilizante es de 300 gramos? $R/\quad]0.346\,562\,, 0.866\,605\,[$, no

6. Taxis Tiquicia planea comprar una flota de taxis. La decisión de comprar cierta marca depende de si los autos de dicha marca rinden por lo menos 43 km por galón de gasolina, en promedio. Se alquilan 36 carros de la marca Odoronly, estos reportan una media de 40.2 km por galón, con una desviación estándar de 5.5 km por galón. A un nivel de confianza de 98%, ¿aconsejaría usted a Taxis Tiquicia comprar autos de esta marca? $R/\quad]38.064\,2, 42.335\,8[$, No

7. Considere la siguiente muestra, tomada de un estudio de acceso a Internet en hogares de cierta comunidad:

$$\begin{aligned} x_1 &= 1, x_2 = 0, x_3 = 1, x_4 = 1, x_5 = 0, x_6 = 1, x_7 = 0, x_8 = 1, \\ x_9 &= 0, x_{10} = 0, x_{11} = 1, x_{12} = 0, x_{13} = 0, x_{14} = 1, x_{15} = 1 \end{aligned}$$

   donde $x_i = 1$ si el hogar i-ésimo tiene acceso a Internet, o $x_i = 0$ si no lo tiene.

   (a) Encuentre un intervalo de confianza de 95% para p, la proporción de hogares en esta comunidad que tienen acceso a Internet. $R/\quad]0.280\,861, 0.785\,805[$

   (b) ¿Cuántos hogares debe incluir la muestra si queremos un intervalo de confianza de 95% con radio menor que 0.02, considerando el valor observado de p? $R/\quad 160$

8. Para estimar una desviavión estándar de una variable aleatoria X se halló el siguiente IC del 98% utilizando una muestra aleatoria específica de tamaño 15 :

$$]0.120\,105\,, 0.751\,004[$$

   Para hallar el IC se supuso que X sigue una distribución normal. Detemine la variación muestral observada que se utiliza en el calculo del IC. $R/\quad 0.5$

9. Una muestra aleatoria específica de 10 alturas de álboles de cierto tipo con 30 años de edad dio una media 4.38 metros y una desviación estándar de 0.06 metros. Hallar un intervalo de confianza del 95% para la altura promedio de este tipo de árboles con 30 años de edad. $R/\quad]4.334\,76, 4.425\,24[$

10. Si $]30, 46[$ es el intervalo de confianza del 95 % para la media de una variable aleatoria normalmente distribuida con variancia desconocida, basado en una muestra aleatoria específica de tamaño 16, halle el valor de la variancia muestral observada. $R/ \quad 225.399$

11. Considera la variable aleatoria normal X con media μ y variancia σ^2, ambas desconocidas. Se tomaron dos muestras aleatorias específicas, ambas de tamaño n : en la primera se obtuvo una desviación estándar $s_1 = 4$ y en la segunda se obtuvo una desviación estándar $s_2 = 5$. Si, utilizando la primer muestra, se obtuvo un intervalo de confianza del 90% para σ^2:

$$]10.5451\ ,\ 27.7288[\,.$$

Determine, utilizando la segunda muestra, un intervalo de confianza del 90% para σ^2.
$R/ \quad]16.4767 \quad , \quad 43.3263[$

12. En una muestra de 8 niños de 6 años de la ciudad A se obtuvieron los siguientes pesos en kilogramos:

$$18, 14.5, 19.2, 20, 22.3, 23.4, 27, 28$$

Suponga que los pesos de los niños de 6 años de A siguen una distribución normal.

(a) Encuentre un intervalo de confianza de 95% para la desviación estándar de los pesos de los niños de 6 años de la ciudad A. $R/ \quad]3.01459, \quad 9.27972[$

(b) Un Doctor indica que el peso de los niños de 6 años de la ciudad A varían demasiado. Él considera que la desviación estándar de los pesos de los niños de 6 años de la ciudad A es superior a 2.5 kilogramos. ¿Considera aceptable la afirmación del Doctor? $R/$ Si

13. Considera la variable aleatoria normal X con media μ y variancia σ^2, ambas desconocidas. En una muestra aleatoria de tamaño $n = 25$ se obtuvo una desviación estándar $s = 4$. Si, a partir de estos datos, se obtuvo un intervalo de confianza para σ^2:

$$]10.5451\ ,\ 27.7288[\,.$$

Determine, aproximadamente, el nivel de confianza del IC hallado. $R/ \quad \alpha = 0.1$

14. Un estudio ha determinado que la desviación estándar de las estaturas de los niños de 5 años del país C es de 7 cm. En una muestra de 15 niños de 5 años se obtuvieron las siguientes estaturas en centímetros:

$$88, 97, 102, 105, 110, 113, 93, 85, 105, 108, 118, 99, 93, 92, 86$$

Suponga que las estaturas de los niños de 5 años del país C se distribuyen normalmente.

(a) Encuentre un intervalo de confianza de 95% para la estatura promedio de los niños de 5 años del país C. $R/ \quad]96.0575, \quad 103.142.[$

(b) De acuerdo a la OMS, la estatura promedio ideal de los niños de 5 años es de 105cm. ¿Considera que los niños de 5 años del país C tienen en promedio una estatura ideal?Justifique su respuesta. $R/ \quad NO$

(c) ¿Qué tamaño de muestra mímimo debe tomarse para estimar la estatura promedio de los niños de 5 años de C con una confiabilidad del 99% y un error de estimación máximo de 2 cm? $R/\quad 82$

15. El Ministerio de Salud de cierto país ha indicado que menos de la mitad aprueban el examen de manipulación y preparación de alimentos. En una muestra de 200 personas, que realizaron dicho examen, se obtuvo que 92 lo aprobaron. Utilizando un nivel de confianza del 90%, ¿aceptaría la afirmación del Ministerio de Salud? $R/\quad]0.402\,027\ , 0.517\,973\ [$, no

16. Considera la variable aleatoria normal X con media μ y variancia σ^2, ambas desconocidas. En una muestra aleatoria de tamaño $n = 20$ se obtuvo una desviación estándar s. Si, a partir de estos datos, un intervalo de confianza del 90% para σ es

$$]5.\,119\,76\,, 15.254\,3[\,,$$

determine un intervalo de confianza al 90% para μ. $R/\quad]11.\,296\,2, 19.\,903\,8[$

17. En un concurso se desea elegir a las personas que arman rápidamente el cubo Rubik. Una persona se considera elegible para el concurso si tarda en promedio a lo sumo 30 segundos armándolo. Ana, en 30 veces que armó el cubo Rubik, tardó en promedio 27 segundos con una desviación estándar de 2 segundos.

(a) Encuentre un intervalo de confianza de 95% para la duración promedio que tarda Ana en armar el cubo. $R/\quad]26.\,284\,3,\quad 27.\,715\,7[$

(b) ¿Considera que Ana debe ser elegible? Justifique su respuesta. $R/\quad$ Si .

18. ¿Es correcto para una persona negarse a ponerse la vacuna contra el Covid? Suponga que se realizó un estudio en una cierta ciudad y se obtuvo que, de 690 encuestados, 390 dijeron que sí, en tanto que el resto respondió que no estaba obligada a vacunarse.

(a) Encuentre un intervalo de confianza del 90% para la proporción de personas que están a favor de la obligatoriedad de vacunarse. R/$]0.53427, 0.596\,17[$

(b) De acuerdo con la parte a, ¿hay evidencia para afirmar que es mayor la proporción de los que están a favor que los que están en contra de la obligatoriedad de vacunarse? R/ Sí

(c) ¿Cuántas personas debería encuestarse para que el IC de 90% tenga radio menor a 0.02? R/Al menos 1653

19. Se construyó un IC de 95% para Se sospecha que una máquina de una cierta fábrica está fallando. Se sospecha que la proporción de artículos defectusos que construye la máquina es del 12%. En una muestra de 140 artículos, se encontró que 10 estaban defectuos.

(a) Construya un IC de 95% para la proporción de articulos defectuosos que fabrica la máquina R/$]0.028767, 0.114\,09[$

(b) Hay evidencia, según la parte a, que apoye la sospecha. R/No

(c) ¿Cuánto debe ser el tamaño de la muestra si se quiere un IC de 95% con radio menor que 0.01? R/2548

20. Si un IC de 95% para la proporción de jóvenes de una cierta ciudad que no piensa tener hijos es $]0.1608, 0.2392[$, ¿cuál fue el tamaño de la muestra que se usó?

21. En un determinado país se encuentra que, en la ciudad A, de 1500 personas encuestadas se encuentra que 450 están a favor de un cierto proyecto de ley, mientras que en la ciudad B, de 1650, 650 están a favor de dicho proyecto de ley.

    (a) Construya un IC de 90% para la diferencia de proporciones entre las personas de ambas ciudades que está a favor de tal proyecto de ley. R/$]-0.121\,61, -0.066268[$ para $p_A - p_B$.

    (b) ¿Puede decirse que la proporción es mayor en la ciudad B? R/ Sí

    (c) Calcule el tamaño de las muestra que deben tomarse en ambas ciudades para que el IC de 90% tenga radio menor que 0.04. R/Al menos 755 de cada una.

22. En un estudio se encontró que 375 profesionales de 460 dijeron estar felices con el ejercicio de su profesión, mientras que 195 técnicos de 260 dijeron estar felices con el ejercicio de su carrera técnica.

    (a) Construya un IC de 99% para la diferencia de proporciones entre profesionales y técnicos felices con su trabajo. R/$]-0.014639, 0.152\,77[$

    (b) ¿Puede decirse que la proporción es mayor en los profesionales? R/ No

    (c) Calcule el tamaño de las muestra que deben tomarse en poblaciones para que el IC de 90% tenga radio menor que 0.1. R/Al menos 226 de cada una.

3 Intervalos de Confianza para dos poblaciones

1. Considere la población X_1 con media poblacional μ_1 y variancia poblacional σ_1^2. Considere la población X_2 con media poblacional μ_2 y variancia poblacional σ_2^2. Suponga que $\overline{X}_1$ y $\overline{X}_2$ siguen una distribución normal para muestras de tamaño n y $3n$ respectivamente, y se conocen σ_1 y σ_2. Demuestre que para encontrar un intervalo de confianza del $100\,(1-\alpha)\,\%$ para $\mu_1-\mu_2$ con un radio menor o igual r, se debe cumplir que

$$n \geq \left(\frac{z_{\alpha/2}\cdot\sqrt{\sigma_1^2+\dfrac{\sigma_2^2}{3}}}{r}\right)^2.$$

2. Se quiere analizar la vida útil en años de las marcas de computadoras A y H. Un vendedor indica que las computadoras H son más caras porque la vida útil promedio de las A es inferior en al menos 2 años a la vida útil promedio de las computadoras H. Se determanaron duraciones de ambos tipos de computadoras, la información se resume en la siguiente tabla

Computadora	tamaño de muestra	$\overline{x}$	s
A	17	2.6 años	0.9 años
H	11	4.7 años	1.1 años

Juan, un estudiante de estadística, a partir de estos datos determinó un IC del 95% para la diferencia de las vidas útiles promedio de las computadoras entre el tipo A y el tipo H, suponiendo que las variancias poblacionales son iguales.

(a) Determine el IC que halló Juan. $R/\quad]-2.880\,88, -1.319\,12[$

(b) ¿Es aceptable la suposición que hace Juan para hallar el IC de suponer que las variacias son iguales? Utilice un nivel de significancia del 5%. $R/\quad]0.598\,968\ , 4.224\,54[$, Si

(c) ¿Se comprueba la afirmación del vendedor? $R/\quad No$

3. Zapatos Únicos desea abrir una tienda exclusiva de su calzado en una de las ciudades A ó B, las cuales tiene cantidad de habitantes similares. El gerente asegura que la tienda se debe abrir en A pues el porcentaje de habitantes que utilizan este calzado es mayor, con respecto al porcentaje en la ciudad B. En un encuesta se reuniéron los siguientes datos:

Ciudad A:	23 habitantes utilizan el canzado de 120
Ciudad B:	19 habitantes utilizan el canzado de 130

(a) Encuentre un intervalo de confianza del 96% para la diferencia de proporciones de habitantes que utilizan el calzado entre las ciudades A y B. $R/\quad]-0.0517495, 0.0.142775[$

(b) ¿Los datos apoyan la afirmación que realiza el gerente? $R/\quad No$

4. Considere la población A con media poblacional μ_1 desconocida y con variancia poblacional 8.1^2. Considere también la población B con media poblacional μ_2 desconocidad y variancia poblacional 6.7^2. Suponga que $\overline{X}_1$ y $\overline{X}_2$ siguen una distribución normal para muestras de tamaños respectivos n y $3n$. ¿De qué tamaño deben ser las muestras para encontrar un intervalo de confianza del 96% para $\mu_1-\mu_2$ con un radio menor o igual a 2? $R/\quad 85$

5. Se quiere estimar la diferencia entre la proporción de personas albinas de Europa y la proporción de personas albinas de América. ¿De qué tamaños deben ser las muestras para obtener un intervalo de confianza del 95% con radio no mayor que 0.03? $R/\quad 2135$

6. Suponga que los pesos de los estudiantes hombres del Tec siguen una distribución normal al igual que los pesos de las estudiantes. Una muestra de estudiantes del Tec contiene 61 hombres y 31 mujeres. En la muestra, los pesos de los hombres tienen una desviación estándar de 8.79 kg, y los de las mujeres una desviación estándar de 8.52 kg.

   (a) Encuentre un intervalo de confianza del 95% para el cociente de las desviaciones estándar. $R/\quad IC$ para $\dfrac{\sigma_h}{\sigma_m}:]0.740\,71, 1.\,389\,91[$

   (b) ¿Es razonable suponer que las desviaciones estándar poblacionales son iguales? $R/\quad$ Si

7. Seguidamente se resumen los resultados obtenidos en un examen de Bachillerato de Matemáticas para una muestra de estudiantes de las provincias A y B:

Provincia	tamaño de muestra	Nota promedio observada $(\overline{x})$	Desviación estándar (s)
A	31	56.3 puntos	4.3 puntos
B	31	61.2 puntos	5.7 puntos

   (a) Encuentre un intervalo de confianza de 90% para el cociente de la varianza de las notas obtenidas por estudiantes de B y la varianza de las notas obtenidas por los estudiantes de A. $R/\quad]0.954\,463 \quad , \quad 3.\,234\,94[$

   (b) ¿Es razonable suponer que las varianzas poblacionales son iguales?. $R/\quad$ Si, pues $1 \in]0.954\,463 \quad , \quad 3.\,234\,94[$

   (c) Encuentre un intervalo de confianza de 90% para la diferencia entre los promedios de las notas de ambos tipos de estudiantes.. $R/\quad]-7.\,009\,53, \quad -2.\,790\,47[$

   (d) ¿Considera aceptable indicar que la nota promedio de los estudiantes de B es superior en más de 2 puntos a la nota promedio de los estudiantes de A? Justifique. $R/\quad$ Si, pues IC es menor a 2

8. El Rector de la Universidad Futuro Garantizado ha indicado que la edad con se graduan los estudiantes de Computadores (C) es mucha más variable que la edad con que se graduan los estudiantes de Tecnologías de Información (TI). Para contrastar esta afirmación se tomó una muestra de estudiantes recién graduados de ambas carreras:

$$\begin{array}{lcccccccccccccc} \text{C} & : & 32 & 28 & 26 & 34 & 26 & 28 & 26 & 22 & 23 & 27 & 35 & 28 & 25 \\ TI & : & 33 & 28 & 26 & 28 & 29 & 32 & 27 & 31 & 28 & 26 & 28 & & \end{array}$$

   Asuma que las edades a las que se graduan los estudiantes siguen una districión normal, para ambas carreras.

   (a) Encuentre un intervalo de confianza de 80% para el cociente de las varianzas de las edades de estudiantes de ambos tipos de carreras. $R/\quad]0.162\,586 \quad , \quad 0.812\,508[$

   (b) ¿Considera aceptable la afirmación del rector?. $R/\quad$ Es aceptable la afirmación.

9. Actualmente existe un proyecto de ley sobre ajuste tributario. En una encuesta, en la ciudad A se encuentra que 164 de 250 ciudadanos están a favor del proyecto, y en la ciudad B, que x de 240 lo apoyan. Con estos datos, se obtuvo un IC para la diferencia de proporciones de apoyo al proyecto entre las dos ciudades (la proporción en A menos la proporción en B), este es

$$]\,0.0709767, 0.216023[$$

   (a) Determine el valor de x. $R/$ 123

   (b) Determine el nivel de confianza del IC dado. $R/$ $\alpha = 0.1$

10. Un profesor considera que el rendimiento promedio (nota promedio) de los estudiantes de Computación en el curso de Matemática Elemental es superior en al menos 9 puntos al rendimiento promedio de los estudiantes de otras carreras. Para analizar esto se tomó una muestra de estudiantes que cursaron el curso el año pasado, obteniendo los siguientes datoss:

Estudiantes	tamaño de muestra	Redimiento promedio observado $(\overline{x})$	Desviación estándar (s)
De computación:	19	78 puntos	4.3 puntos
De otras carreras:	17	65 puntos	4.7 puntos

Suponga que el redimiento promedio en el curso de Matemática Elemental, tanto en Computación como en otras carreras, se distribuye normalmente.

   (a) Encuentre un intervalo de confianza de 90% para el cociente de las varianzas de las notas de ambos tipos de estudiantes. $R/$ $]0.530\,978 \quad , \quad 2.\,750\,20[$

   (b) Justifique por qué es razonable suponer que las varianzas poblacionales son iguales.$R/$ Si, pues $1 \in]0.530\,978 \quad , \quad 2.\,750\,20[$

   (c) Encuentre un intervalo de confianza de 95% para la diferencia entre los promedios de las notas de ambos tipos de estudiantes.. $R/$ Un IC del 95% para $\mu_I - \mu_D$ es$]9.\,951\,9, \quad 16.0481[$.

   (d) ¿Considera aceptable la afirmación del profesor? Justifique. $R/$ Si, pues IC es mayor a 9

11. En una reciente encuesta sobre intención de voto presidencial se hizo una separación entre encuestados, los que votarán por primera vez y los que ya lo han hecho antes. Entre los resultados de la encuesta se tiene que 46 de 120 de los votantes NO nuevos consultados apoyarían a la candidata María José Figueroa; mientras que de los nuevos votantes 22 de 64 encuestados mostraron apoyo a esta candidata. ¿Sugieren los datos a un nivel de confianza del 99% que el apoyo hacia María José Figueroa es menor en votantes nuevos? $R/$ $I =]-0.151\,662, 0.230\,829[$. Si M es menor en votantes nuevos entonces $p_1 - p_2 > 0$ y los datos NO sugieren esto de acuerdo al IC

12. Un equipo de futbol S de primera división ha mostrado por lo general un mejor rendimiento histórico que un equipo L. Un entrenador ha indicado que lo anterior se debe a que el equipo S ha tenido jugadores que en promedio son más jóvenes que los jugadores de L. Para investigar esta afirmación se tomaron algunos jugadores al azar de cada equípo de los últimas 6 temporadas, y se registro su edad:

L	:	32	28	26	24	26	28	26	20	21	27	26	28	25	
S	:	30	24	26	23	25	23	24	31	23	22	21	20	22	21

Suponga que las edades de los jugadores que ha estado en cada equipo se distribuyen normalmente.

(a) Si se supone que las varianzas de las edades de los equipos son iguales, construya un intervalo de confianza del 95% para la diferencia de los promedios de edades en los equipos. $R/ \quad]-0.513\,244\,, \quad 4.\,502\,24[$

(b) Con base en lo anterior, ¿Apoyaría la afirmación del entrenador? Justifique. $R/ \quad No.$Pues en el IC se tiene también valores negativos

4 Pruebas de hipótesis de una población

1. Una muestra aleatoria de doce estudiantes de secundaria dio las siguientes cifras en horas para el tiempo semanal que dedican a jugar video juegos (Play Station, Nintendo,..):

$$25, 53, 39, 30, 15, 39, 29, 32, 55, 39, 53, 38.$$

El promedio de los datos recolectados es $\frac{149}{4}$ horas con una desviación estándar de 12.114 horas. Un psicólogo señala que el tiempo semanal que dedica un estudiante a jugar video juegos es perjudicial si es de por lo menos 40 horas.

(a) Un investigador afirma que en promedio el tiempo semanal que dedica un estudiante a jugar video juegos es perjudicial. Pruebe esta afirmación con un nivel de significancia de 0.05, determinando las regiones de aceptación y rechazo.$R/$ Valor $P = 0.22413$, $\mu_c = 33.7198$. No hay evidencia en contra de la afirmación

(b) Por otro lado, un sociólogo indica que hoy en día más de la tercera parte de los estudiantes de secundaria dedican un tiempo semanal **perjudicial** a jugar video juegos. ¿Existe evidencia a favor de esta afirmación? $R/$ Valor $P \approx 0.441\,654$, No.

(c) ¿De qué tamaño debería ser una muestra para probar la hipótesis en (b) con un nivel de significancia de 5% y un potencia de 90% cuando realmente una cuarta parte de los estudiantes de de secundaria dedican un tiempo semanal perjudicial a jugar video juegos? $R/$ 255

2. Un ingeniero de computación, se ha quejado ante la Oficina de Trabajo por las grandes diferencias salariales que se le paga a un ingeniero en computación de un empresa a otra. Ante esto, el presidente de la Cámara de Empresas ha señalado que la desviación estandár del salario mensual de un ingenierio en computación es menor a 100 dólares. Para analizar esta afirmación, la Oficina de Trabajo en una muestra de 40 ingenieros observó un salario promedio de 2000 dólares con una desviación estándar de 90 dólares. Se quiere contrastar la afirmación a un nivel de significancia del 10%.

(a) Determine las regiones de aceptación y rechazo de la prueba para el estadístico S^2.$R/$ $A =]7229.69, +\infty[\,, \quad R =]0, 7229.69[\,.$

(b) ¿Es razonable la afirmación del presidente de la Cámara de Empresas? $R/$ No, no hay evidencia significativa a su favor.

(c) Acote la probabilidad del error tipo II de la prueba si $\sigma = 95$ dólares. $R/$ $]0.8, 0.9[\,.$

3. Un médico afirma que la edad promedio de personas con sobrepeso en cierta ciudad es de 35 años. Para investigar dicha afirmación se tomó una muestra de 20 personas con sobrepeso de dicha ciudad y se observó un edad promedio de 36.7 años con una desviación de 4.5. Al realizar contraste de hipótesis se obtuvo que uno de los promedios críticos es $u_c = 33.1389$.

(a) Determine aproximadamente el nivel de significancia utilizado en el contraste de hipótesis realizado. $R/$ 0.08.

(b) ¿Existe evidencia en contra de la afirmación del médico? $R/$ No, se acepta la afirmación del médico.

4. Un profesor de la Universidad Futuro Garantizado cree que más de la mitad de los estudiantes matriculados se retiran de al menos una materia. En una muestra de 80 estudiantes, se observó que 45 se habían retirado de al menos una materia.

   (a) Determine el valor P de la prueba de hipótesis para contrastar la creencia del profesor. $R/$ 0.131357.

   (b) ¿Considere aceptable la creencia del profesor? $R/$ No, no hay evidencia a favor de la afirmación del profesor.

   (c) ¿De qué tamaño debe ser una muestra para que la prueba tenga una significancia de 2% y una potencia de 90% cuando el verdadero porcentaje de estudiantes que se retiran de al menos una materia es del 60% ? $R/$ 273.

5. La compañía SERVICIO COMPLETO ofrecerá servicios de telecomunicaciones en nuestro país a penas le aprueben su proyecto, y afirma que más del 30% de los hogares del país pasarán a ser sus clientes. De 80 hogares encuestados, 25 mencionaron que efectivamente se suscribirán a la compañía

   (a) ¿Cuál es la probabilidad de error tipo II si la verdadera proporción es de 40%, el tamaño de la muestra es 120 y el nivel de significancia de 0.025? $R/$ 0.345.

   (b) Según el valor obtenido en la parte (b), ¿cuál es el valor de la potencia y qué significa? $R/$ 0.655.

6. Se estima que el tiempo que llenado de embases de la compañía BUEN SABOR se distribuye normalmente con una desviación estándar de 5 minutos. Si en una muestra de 25 embases dio una desviación estándar de 3 minutos, ¿hay evidencia de que en realidad la desviación es menor que 5 minutos? Utilice $\alpha = 10\%$. $R/$ Si.

7. Un modelo de automóvil ha registrado durante varios años un rendimiento de combustible promedio de 15 km por litro de gasolina y una desviación estándar de 3 km. Para el modelo de este año se mantuvo la desviación estándar y se quiere probar si el rendimiento promedio cambió, usando un nivel de significancia de 4% y una potencia de 90% si el nuevo promedio es 17 km/L.

   (a) ¿De qué tamaño debe ser la muestra a tomar? $R/$ 30.

   (b) Si se toma una muestra del tamaño escogido en el punto anterior, ¿cuál debe ser la región de aceptación para $\overline{X}$? $R/$ $\mu_{c1} = 13.8772$, $\mu_{c2} = 16.1228$.

8. La resistencia a la rotura de los cables producidos por un fabricante sigue una distribución normal con una media de 1800 libras y una desviación estándar de 100 libras. Mediante una nueva técnica en el proceso de fabricación, la desviación estándar se mantiene y se espera que la resistencia promedio se incremente. Para ensayar esta aspiración el Departamento de Ingeniería toma una muestra de cables fabricados utilizando la nueva técnica y encuentra que su resistencia media es de 1850 libras. Para determinar efectivamente si hay un aumento de resistencia, dicho departamento realizó una pueba de hipótesis con un nivel de significancia de 0.01. Al normalizar y estandarizar la media observada obtuvieron un valor observado de $z = 3.5$. Con estos datos,

   (a) Enuncie las hipótesis nula y alternativa que debió escribir el Departamento. $R/\quad H_1 : u > 1800$.

   (b) ¿De qué tamaño fue la muestra que tomó el Departamento de Ingeniería? $R/\quad 49$.

   (c) ¿Existe evidencia a favor de que la resistencia se incrementó? $R/\quad$ Si .

9. En una fábrica de tornillos, se estima que al menos 98% de las unidades producidas son satisfactorias. En una muestra de 50 unidades se encuentran dos defectuosas. ¿Es esto evidencia en contra de la estimación? $R/\quad$ No hay evidencia en contra de la afirmación.

10. Una empacadora de papas tostadas, cuando está bien ajustada, llena las bolsas con un contenido promedio de 500 gramos y una desviación estándar de no más de 5 gramos. Se toma una muestra de 21 bolsas para probar, con un nivel de significancia de 5%, si la máquina está bien ajustada. ¿Cuáles son los extremos del intervalo de aceptación para la desviación estándar muestral? $R/\quad A = [0, 6.26602[$.

11. Se han registrado los siguientes tiempos de un proceso: 3.3 s, 2.6s, 2.7s, 3.5s, 3.2s y 2.7s. ¿Hay evidencia de que la desviación estándar de los tiempos es menor que un segundo? $R/\quad$ Si .

12. Una pequeña tienda para su funcionamiento requiere de una venta promedio diaria de por lo menos 50 mil colones para cubrir sus gastos. En 90 días de estudio se obtuvo una venta promedio de 49 500 colones, con una desviación de 5700 colones.

   (a) Plantee las hipótesis nula y alternativa para esta prueba y halle las regiones de aceptación y rechazo con un nivel de significancia de 0.05. $R/\quad u_c = 49.2344$.

   (b) Determine si la tienda debería dejar de funcionar con un nivel de significancia de 0.05.

13. Una soda del país C requiere para su funcionamiento una venta promedio diaria de por lo menos 70 mil colones para cubrir sus gastos. Desde que inicio una huelga el país C, se cree que la soda ha sido afectada. Un muestra de 12 días posterior a la huelga se obtuvo una venta promedio diaria de 68120 colones, con una desviación de 6000 colones. Se quiere analizar si la soda debe dejar de funcionar. Suponga las ventas diarias siguen una disitribución normal.

(a) Plantee las hipótesis nula y alternativa para esta prueba y halle las regiones de aceptación y rechazo para $\overline{X}$ con un nivel de significancia de 0.05. $R/$ Región de rechazo: $]-\infty, 66889.4[$. Región de aceptación: $]66889.4, +\infty[$

(b) Determine si la soda debería dejar de funcionar con un nivel de significancia de 0.05.$R/$ $\overline{x} = 68120 \in A$,no hay evidencia en contra de que cierre.

(c) Acote la probabilidad del error tipo II si el verdadero promedio diario es de 65000 colones. $R/$ $\beta \in]0.1, 0.2[$, con $v = 11$:

14. Debido a la utilización de TV por cable, los habítos de sueño de las familias ha cambiado. Se cree que menos de la decima parte de los niños, se duermen antes de las 8:00pm. En una muestra de 40 niños, solo dos de ellos se duermen antes de las 8:00pm. ¿Es aceptable la creencia?$R/$ Valor $P \approx 0.222\,808 > 0.1$, se rechaza la afirmación.

15. Un cliente se queja que las baterías de las computadoras portátiles CDF tiene una duración promedio inferior a las 3 horas que indica el manual. Las duraciones de ocho de esas baterías en minutos, cargadas completamente, son

181	173	185	164	170	177	186	178

Suponga que la duración de las baterías, cargadas completamente, se distribuye normalmente.

(a) Plantee la hipótesis nula y alternativa para analizar la queja del cliente. Acote el valor P de la prueba. $R/$ $H_1 : u < 180$

(b) A un nivel de significancia del 7%, ¿contradicen los datos recolectados la afirmación del cliente? Justifique su respuesta. $R/$ Valor $P > 0.07$, se rechaza la afirmación

16. En una muestra aleatoria de 15 estudiantes del TEC solo uno de ellos apoya al Club Sport Cartaginés. ¿Es esto evidencia fuerte de que menos de una quinta parte de los estudiantes del TEC apoya al club? $R/$ Valor $P > 0.1$, no

17. Un médico afirma que el peso de niñas de tres años de cierta provincia son muy similares, especificamente señala que la desviación estándar de los pesos no llega a ser de 3 kg. Para analizar esta afirmación se toma el peso de 10 niñas, obteniendo los siguientes datos en kg:

14.5	11.6	12.8	15.1	14.2	13.7	12.9	13.8	14.1	11.9

Suponga que el peso de niñas de 3 años sigue una distribución normal. Utilice $\alpha = 0.1$.

(a) Determine las regiones de rechazo y de aceptación para el estadístico S que permiten realizar el contraste de hipótesis sobre la afirmación del médico. $R/$ $A =]2.041\,61, \quad +\infty[$

(b) ¿Hay evidencia a favor de la afirmación indicada por el doctor? Justifique su respuesta. $R/$ $s = 1.38303 \in R$, se acepta la afirmación del director

(c) Acote la probabilidad del error tipo II sabiendo que la desviación estándar de los pesos de los niñas de 3 años es en realidad $\sigma = 1.5$. $R/$ $\beta \in]0.05, 0.1[$

18. Cuando el promedio de ventas, por establecimiento autorizado, de una marca de relojes caen por debajo de las 170 000 unidades mensuales, se considera razón suficiente para lanzar una campaña publicitaria que active las ventas de esta marca. Para conocer la evolución de las ventas, el departamento de mercadeo realiza una encuesta a 51 establecimientos autorizados, seleccionados aleatoriamente, que facilitan la cifra de ventas del último mes en relojes de esta marca. A partir de estas cifras se obtienen los siguientes resultados: $\bar{x} = 169\ 411,8$ unidades, $s = 32\ 827,5$ unidades. Suponiendo que las ventas mensuales por establecimiento se distribuyen normalmente; con un nivel de significación del 5 % y en vista a la situación reflejada en los datos. ¿Se considera oportuno lanzar una nueva campaña publicitaria?$R/\quad u_c = 162438$, no es oportuno una nueva publicidad

19. Una empresa está interesada en lanzar un nuevo producto al mercado. Tras realizar una campaña publicitaria, se toma una muestra de 1000 habitantes, de los cuales 25 no conocían el producto. ¿Apoya el estudio la hipótesis de que menos del 3% de la población no conoce el nuevo producto? $R/$ Valor $P = 0.176186$. No se apoya el estudio

20. Dada una población normal con varianza σ^2 desconocida, un profesor afirma que $\sigma^2 \leq 20$. Para analizar esta afirmación se tomó una muestra de tamaño 30. Un estudiante realiza correctamente el contraste de hipótesis y obtiene que el valor P de la prueba se encuentra en el intervalo $]0.05, 0.1[$.

    (a) ¿Existe evidencia en contra de la afirmación del profesor? $R/\quad Valor\ P > 0.05$. Se acepta la afirmación

    (b) Determine los posibles valores de la desviación estandar muestral de s utilizada por el estudiante en la prueba. $R/\quad 5.192 < s < 5.417\,53$

21. Se cree que la varianza del examen proveniente del banco de preguntas A tiene varianza menor a 16. Realice una prueba de hipótesis para investigar esta creencia. ¿Hay evidencia que la apoye? Use un nivel de significancia de 0.04. ¿Cuáles son las zonas aceptación y rechazo?ZA:$]0, 29.0310[$ y ZR:$]29.0310, \infty[$. No hay evidencia.

22. Suponga que se ha utilizado una muestra de tamaño 26 de una problación normal y un nivel de significancia de 4% para contrastar: $H_0 : \sigma^2 = 16$ contra $H_1 : \sigma^2 > 16$. ¿Para qué valores de la varianza muestral S^2 no se rechaza la hipótesis nula? R/$S^2 < 9.0226$

23. Se realiza un estudio sobre el lavado de manos en niños de 10 años de edad, de dos escuelas: A y B. La cantidad de lavado de manos diarios que realizan los niños en la siguiente tabla:

Cantidad de lavados de mano diarios	Cantidad de niños A	Cantidad de niños B
0	1	1
1	6	2
2	7	10
3	9	6
4	12	7
5	13	10
6	15	15

En algunas ocasiones, la persona que encuesta se olvidad de anotar la escuela de procedencia del niño, por lo que, utilizando la información anterior, plantea la siguiente prueba de hipótesis:
H_0: El niño asiste a la escuela A.
H_1: El niño asiste a la escuela B.
Para concluir, establece la siguiente regla de decisión: Rechazar H_0 si el niño realiza a lo sumo 2 lavados de mano.

(a) Calcule la probabilidad de error tipo I e interprete el resultado. R/0.2222

(b) Calcule la probabilidad de error tipo II e interprete el resultado . R/0.74510

24. Se cree que una proporción p de jóvenes menores de 25 años en una determinada ciudad piensan no tener hijos. Se cree que, por lo menos una cuarta parte de los jóvenes menores de 25 años de esta ciudad tienen esa posición. Se quiere probar esta hipótesis contando el número de jóvenes a favor en una muestra aleatoria de 400. Defina la región de aceptación y de rechazo, en términos de jóvenes a favor, para que:

    (a) la probabilidad de error tipo I no sea mayor al 4%. R/ ZA:$\widehat{\mathrm{P}} > 0.212\,11$, $\widehat{\mathrm{P}} \leq 0.212\,11$, o bien ZA:$\widehat{\mathrm{x}} \geq 85$, ZR:$\widehat{\mathrm{x}} < 85$

    (b) Calcule la potencia de la prueba para la alternativa específica $H_1' : p = 0.175$. Interprete. R/0.9973, La probabilidad de rechazar correctamente H_0 es 99.73.

25. Se afirma en una empresa que se tiene una producción media diaria mayor a 60 unidades. Se observarán 120 días y si se tiene una media mayor a 61.5 unidades producidas, se concluirá que en efecto la producción media es de 60 unidades ó más. Se sabe que la desviación estándar poblacional es de 10.

    (a) Determine las regiones de aceptación y rechazo. R/Aceptación $\overline{X} < 61.5$ y rechazo $\overline{X} \geq 61.5$

    (b) Calcule la probabilidad de error tipo I. R/0.05

    (c) Calcule la probabilidad de error tipo II para $H_1' : \mu = 64$.¿Cuál es la potencia de la prueba? R/$\beta = 0.003$, potencia:0.997

26. El consumo de café por día, en tazas, sigue una distribución aproximadamente normal Se piensa que en cierta ciudad se consumen, en promedio por día, 4.5 tazas de café con desviación estándar 10 tazas de café. Se desea contrastar $H_0 : \mu = 4.5$ contra $H_1 : \mu \neq 4.5$ con una muestra de $n = 120$ personas tomadoras de café, a las cuales se le preguntará por la cantidad de tazas de café que toman por día. Si se define la región de aceptación $2.5 \leq \overline{X} \leq 6.5$, determine:

    (a) La probabilidad de error tipo I. R/0.02845

    (b) La probabilidad de error tipo II si la verdadera media es de 1.25 tazas de café.¿Cuál es la potencia? R/$\beta = 0.085453$,Potencia:1-$\beta = 0.91455$

5 Pruebas de hipótesis de dos poblaciones

1. Suponga que las marcas de disco DVD para grabar con capacidad nominal de 4.7 gb son A y B. Los disco A son más baratos que los disco B, razón por la cuál un vendedor de accesorios para computadoras afirma que la capacidad promedio de los discos A es menor en por lo menos 70 mb a la capacidad promedio de los discos B. Se tomaron muestras de capacidades de ambos tipos de discos en mb, la información se resume en la siguiente tabla

Disco	tamaño de muestra	$\overline{x}$	s
tipo A	21	4698 mb	15 mb
tipo B	17	4752 mb	27 mb

Suponga que las capacidades de los discos, tanto los marca A como los marca B, siguen una distribución normal.

   (a) Un estudiante X del TEC, realizó la prueba de hipótesis para la afirmación del vendedor suponiendo que las desviaciones estándar de las capacidades, de ambos tipos de discos, son iguales. ¿Aceptó X la afirmación del vendedor? $R/$ No se halló evidencia significativa a favor de la afirmación.

   (b) El estudiante X considera que la suposición que las desviaciones estándar de las capacidades, de ambos tipos de discos, son iguales, es evidente.Con un nivel de significancia de 5%, ¿considera que la suposición de X es correcta? $R/$ $No.$

   (c) Considere la afirmación: el estudiante X reprobó Estadística, ¿Aceptaría esta afirmación? Respoda según su intuición. $R/$ Se acepta.

2. El pueblo C tiene problemas con el fumado en adolescentes. En una muestra de 60 hombres adolescentes se observó que 40 fuman, y en una muestra de 45 mujeres adolescentes se observó que 39 fuman. El sacerdote del pueblo afirma que en esto tiempos son más las mujeres adolescentes que fuman que hombres adolescentes. ¿Los datos apoyan la afirmación del sacerdote? $R/$ Si.

3. Las carreras de Electrónica y Computación tiene gran demanda laboral, esto hace que muchos estudiantes ingresen al mundo laboral antes de graduarse. Sin embargo, un profesor afirma que, para estudiantes de último de carrera, es mayor la proporción de estudiantes de Computación que laboran que la de estudiantes de Electrónica. En una pequeña encuesta se obtuvo la siguiente información

Estudiantes de último año	# de encuestados	Laboran
Computación	20	12
Electrónica	30	14

   (a) Determine el Valor P para contrastar la afirmación del profesor. $R/$ 0.178786.

   (b) ¿Hay evidencia que respalde la afirmación realizada por el profesor ? $R/$ No hay evidencia significativa a favor de la afirmación del profesor.

4. Una muestra de jugadores de fútbol de la primera división actual dieron los siguientes resultados, donde n es el número jugadores, $\overline{x}$ el promedio de los años de experiencia y s la desviación estándar de los años de experiencia:

Equipos	n	$\overline{x}$	s
A	15	3.2	2.85
B	21	5.7	1.94

Suponga que los años de experiencia se distribuyen normalmente, tanto para los jugares del equipo A como para los jugadores del equipo B.

   (a) Con un nivel de significancia de 5%, pruebe la hipótesis de que las varianzas entre los años de experiencia de ambos equipos son iguales $R/$ se puede suponer que las varianzas son iguales.

   (b) ¿Considera que el promedio de los años de experiencia en los jugadores de B es superior al promedio de los años de experiencia en los jugadores de A, en más de 9 meses?

5. Se quiere analizar la capacidad de dos tipos de tarjetas de memorias SD con capacidad nominal de 32 gb, pues un estudiante del TEC afirma que la capacidad promedio de las tarjetas A es mayor en por lo menos 0.25 gb a la capacidad promedio de las tarjetas B. Se tomaron muestras de capacidades de ambos tipos de tarjetas, la información se resume en la siguiente tabla

Tarjeta	tamaño de muestra	$\overline{x}$	s
tipo A	31	15.8 gb	0.25 gb
tipo B	31	15.63 gb	0.17 gb

Suponga que las capacidades de las tarjetas de memorias SD, tanto las tipo A como las tipo B, siguen una distribución normal.

   (a) Escoja un nivel de significacia apropiado para que se puede aceptar que las varianzas poblacionales son iguales. $R/$ $\alpha = 0.02$ o menor.

   (b) Con un nivel de significancia de 5%, ¿aceptaría la afirmación del estudiante del TEC?$R/$ Si .

6. Se quiere investigar si la duración promedio de los bombillos A excede a la duración promedio de los bombillos B en al menos 150 horas. Se supone que ambas poblaciones se distribuyen normalmente con desviaciones estándar respectivas $\sigma_A = 60$ horas y $\sigma_B = 50$ horas. ¿De qué tamaño deben ser las muestras para un nivel de significancia de 10% y una potencia de 85% cuando la verdadera diferencia es de 120 horas? $R/$ 37.

7. En una encuesta sobre práctica de deportes se encontró que, de 120 personas solteras encuestadas, 21 practican ciclismo al menos una vez por semana, y que de 80 personas casadas, 12 lo practican. ¿Hay evidencia significativa para creer que la proporción de personas que practican ciclismo varía entre los solteros y los casados? $R/$ No, $Valor\ P = 0.638356$.

8. Una gran empresa tiene dentro de sus instalaciones dos restaurantes: A y B, para sus empleados. El gerente de la empresa ha afirmado que la edad promedio de los empleados que utilizan el restaurante A es mayor en por lo menos 10 años a la edad promedio de los que utilizan el restaurante B. Para analizar esta afirmación se tomaron muestras de edades de los empleados que utilzan ambos restaurantes:

Restaurante	tamaño de muestra	$\overline{x}$	s
A	21	33.1	3.1
B	31	24.2	2.5

Sean X_1 : las edades de los empleados que utilizan el restaurante A, X_2 : las edades de los empleados que utilizan el restaurante B. Suponga que X_1 y X_2 se distribuyen normalmente.

   (a) Utilizando regiones, a un nivel de significancia del 0.05, pruebe la hipótesis de que la varianzas de X_1 y X_2 son iguales. $R/$ $f_{0bs} = 1.5376 \in A$, se puede suponer variancias iguales.

   (b) Utilizando Valor P, ¿existe evidencia en contra de la afirmación del gerente? $R/$ Valor $P = P(T < t_{obs}) \in]0.05, 0.1[$,Se acepta H_0 (leve) y se acepta la afirmación.

9. De una muestra de 100 acciones de la bolsa de valores A, 32 tuvieron ganancia el martes pasado. Una muestra de 100 acciones de la bolsa de valores B indica que 27 obtuvieron ganancia ese mismo día. Ante estos resultados, una persona afirma que hay una mayor proporción de acciones que obtuvieron ganancia en la bolsa A, con respecto a la otra bolsa, el martes pasado. Utilice $\alpha = 0.04$.

   (a) ¿Hay evidencia que respalde la afirmación de esta persona? $R/$ Valor $p = 0.223627$. No hay evidencia a favor de la afirmación

   (b) ¿De qué tamaño deberán haberse tomado las muestras si se desea hacer el contraste de hipótesis con un nivel de significancia de $\alpha = 0.04$ y una potencia de 80% cuando el 30% de las personas obtuvo ganancia de la bolsa A y el 25% en B? $R/$ $n \geq 1068.85$

10. Considere las siguientes variables aleatorias independientes

$$X_1 : \text{variable normal con una desviación estándar de } \sigma_1 = 5.2$$
$$X_2 : \text{variable normal con una desviación estándar de } \sigma_2 = 3.4$$

De X_1 se toma una muestra aleatoria de tamaño $n_1 = 25$ y se observa una media de $\overline{x}_1 = 81$. De X_2 se toma una segunda muestra de tamaño $n_2 = 36$ y se observa una media de $\overline{x}_2 = 78$. Pruebe la hipótesis, a un nivel de significancia del 0.05, que $\mu_1 = \mu_2$ contra la alternativa $\mu_1 \neq \mu_2$. $R/$ $z_{obs} = 2.533\,01, z_{c1} = -1.96,$ $z_{c2} = 1.96$, $Valor\ P = 0.011\,406$.Hay evidencia en contra de que las medias sean iguales

11. En la universidad A de 200 docentes se tiene que 131 toman al menos 3 tazas de café por día. En la universidad B de 300 docentes se tiene que 215 toman al menos 3 tazas de café por día. A un nivel de significancia del 5%, ¿existe evidencia de que la proporción de docentes que toman al menos 3 tazas de café por día en la Universidad A es menor la proporción de docentes que toman al menos 3 tazas de café por día en la Universidad B?. $R/$ Valor $p = P(Z < -146) = 0.072145$. No hay evidencia a favor...

12. Se desea comparar los melones de dos regiones A y B en cuanto al peso. Se tomaron dos muestras de cada región y se obtuvo la siguiente información:

Región A	Región B
$n_1 = 26$	$n_2 = 28$
$\overline{x}_1 = 1.5kg$	$\overline{x}_2 = 1.9kg$
$s_1 = 1.6kg$	$s_2 = 2.1kg$

Asuma que la diferencia entre las medias de los pesos de los melones entre amabas regiones se distribuye normalmente.

(a) Con un nivel de significancia de 0.02, ¿se puede considerar que las varianzas son iguales?.Indique cuáles son los intervalos de aceptación y rechazo
R/Sí./ Aceptación:$]0.435391, 2.5360022[$, Rechazo:$]0, 0.435391[\,;\,]2.5360022, +\infty[$

(b) ¿Se puede afirmar que hay diferencia entre las medias de los pesos de los melones de ambas regiones? Calcule el valor P de la prueba para concluir. R/No. $P = 0.427425$

13. Se desea comparar el tiempo que tardan en hacer efecto dos tipos de analgésicos, A y B. Una muestra de 20 enfermos reportó un tiempo promedio de efecto de 35 minutos con deviación estándar de 4,1 minutos del analgésico A, mientras que otra muestra de 24 personas enfermas, reportó un tiempo promedio de efecto 40 minutos del analgésico B, con desviación estándar de 2,2 minutos Suponga que la diferencia entre las medias de tiempo de alivio de los analgésicos sigue una distribución normal.

(a) Con un nivel de significancia de 0.05, ¿se puede considerar que las varianzas son iguales? Indique cuáles son los intervalos de aceptación y rechazo R/No. Zona Aceptación:$]0.470973, 2.060754[$,Zona Rechazo:$]0, 0.470973[\,;\,]2.060754, +\infty[$

(b) ¿Se puede afirmar que el tiempo promedio de efecto del analgésico B aventaja en más de 2 minutos al del analgésico A? Use el valor P. R/$P \approx 0.000019$. Sí.

14. Se desea comparar la proporción de personas mayores de edad que están a favor de un proyecto de ley entre dos ciudades A y B. En la ciudad A se encontró que 130 de 200 están a favor, mientras que en la ciudad B, 100 de 250 encuestados dijeron estar a favor.

(a) Con un nivel de significancia de 0.02, ¿hay evidencia para afirmar que en la ciudad A hay una mayor proporción de personas a favor del proyecto que en la ciudad B? R/ Sí hay evidencia.

(b) ¿De qué tamaño deberían ser las muestras para un nivel de significancia de 2% y una potencia de 92%, cuando 40% de los habitantes de la ciudad A y 25% de los ciudad B están a favor de dicho proyecto de ley? R/ De al menos 163 de cada cidudad.

15. En una cierta universidad, se diseñaron dos tipos de pruebas de un parcial, formados de con dos bancos de preguntas distintos, A y B. Se sospecha que la proporción de estudiantes que aprueba el parcial construido con el banco A supera a la proporción construido con el banco

B en al menos 3%. Para verificar esto, la universidad aplicó ambas pruebas a un grupo de 84 estudiantes, obteniendo los siguientes resultados:

Parcial	n	aprobados	s
Banco A	45	32	3.2
Banco B	39	27	1.5

Con un nivel de significancia del 5%, ¿hay evidencia que apoye la sospecha?R/ Sí hay evidencia.

6 Otras pruebas de hipótesis: Bondad de ajuste, Independencia y ANOVA

1. A continuación se presentan los datos obtenidos en una investigación realizada entre estudiantes universitarios, quienes evaluaron el desempeño de alguno de sus profesores. Cada estudiante seleccionó al profesor por evaluar. Se trata de un total de 780 estudiantes:

	Alumnos	Alumnas
Profesores	269	275
Profesoras	59	177

   Con base en estos datos, ¿considera que el sexo del estudiante y el sexo del profesor evaluado son dependientes? $R/$ Valor $P < 0.01$, Si

2. Se desea estudiar la dependencia entre el grado académico y la aceptación a una reforma tributaria en la ciudad C, para ello se tomaron los datos de 365 individuos, los cuales se resumen en la siguiente tabla:

	Primaria	Secundaria	Universidad
Aceptan la reforma	50	65	70
No aceptan la reforma	75	55	50

   Puede concluirse con estos datos que la aceptación de la reforma es independiente del grado académico. $R/$ Se rechaza H_0, hay evidencia en contra de la independencia.

3. En una encuesta se encontró la siguiente información sobre estudiantes de una universidad de distintas áreas:

	Salud	Ciencias básicas	Ciencias sociales
$Hace$ ejercicio	21	12	7
No hacen ejercicio	99	68	48

   ¿hay evidencia de que la proporción de personas que realizan ejercicio varía entre las áreas de estudio indicadas? $R/$ No

4. En una encuesta sobre la soda comedor EL COMELON, se les preguntó a 200 clientes su opinión sobre la variedad de los alimentos y su nivel de ingreso. Los resultados se resumen en la siguiente tabla de contingencia.

		Nivel de ingreso		
		Bajo	Medio	Alto
Variedad	Poco	3	10	27
	Regular	15	20	50
	Mucha	21	40	14

   ¿Existe evidencia de que la opinión que tiene un cliente sobre la variedad de los alimentos dependiente de su nivel de ingreso? $R/$ Si, valor $P < 0.05$.

5. Un curso es impartido normalmente por tres profesores. Un estudiante considera que la nota en el curso depende del profesor que lo imparta. Ante esto la Oficina de Registro seleccionó al azar el promedio de 11 estudiantes que aprobaron el curso, los cuales se muestran a continuación

Profesor A	Profesor B	Profesor C
77	96	74
88	84	84
81	75	75
82	80	
78		

Con base a estos datos, ¿Puede afirmarse que la nota promedio en el curso no varía según el profesor que lo imparte? $R/$ Si, $f_{obs} \approx 0.758\,938$

6. Se extraen tres cartas de una baraja ordinaria, con reemplazo, y se registra el número Y de espadas. Después de repetir el experimento 64 veces se registran los siguientes resultados

y	0	1	2	3
Frecuencia	21	31	12	0

¿existe evidencia en contra, con un nivel de significancia de 0.01, de que los datos se pueden ajustar a una distribución binomial $P(y) = C(n, y)\left(\frac{1}{4}\right)^{y}\left(\frac{3}{4}\right)^{n-y}$ para $y = 0, 1, 2, 3$?.$R/\quad No$, $X_{obs} \approx 2.\,325\,9$

7. Actualmente, en el Ministerio de Transportes del país C, se analiza la posibilidad de la apertura en el examen teórico para obtener la licencia de conducir. Existen tres compañías en el extranjero interesadas, con experiencia en la elaboración y aplicación de este tipo de pruebas, las cuales tienen una escala de 0 a 100. Se compararon los resultados de ciudadanos en las pruebas elaboradas por estas compañías:

Compañía		
C1	C2	C3
60	95	74
40	80	84
79	30	53
		47

A un nivel de significación del 5%, ¿hay evidencia en contra de que los promedios en las pruebas de las tres compañías son iguales? $R/\quad No$

8. Una partida ESCUDO consiste en una sucesión de lanzamientos hasta obtener un escudo, sea X el número de lazamientos realizados en un partida. Se registraron los valores de X en 80 partidas ESCUDO para una moneda particular, los resultados se resumen en la siguiente tabla:

X :	1	2	3	4
# de familias:	35	20	18	7

A un nivel de significancia del 5%, ¿Existe evidencia en contra de que X sigua la distribución $f_X(k) = \frac{1}{2^k}$? $R/\quad \chi^2_{obs} = 12.825$, se halla evidencia en contra de que $X \sim f_X$

9. Seguidamente se presentan los datos obtenidos en una investigación realizada entre estudiantes universitarios, quienes indicaron cual de los dos sabores A y B de helados preferían. Cada estudiante seleccionó el sabor preferido y en total se obtuvo la opinión de 200 estudiantes:

	Alumnos	Alumnas
Sabor A	73	76
Sabor B	31	20

Con base en estos datos, ¿considera que el sexo del estudiante y el sabor de helado son dependientes? $R/$ $\chi^2_{obs} = 2,170940171$, valor $P \in]0.1, 0.2[$Se acepta que son independientes.

10. En un test de memoria, una lista de10 palabras es presentada en la pantalla del ordenador al sujeto, luego se le pide al sujeto que recuerde las palabras presentadas. La nota obtenida será el número de palabras que logró recordar. Cada test selecciona las palabras utilizando uno de tres procedimientos diferentes. Seguidamente se presenta la nota obtenida de una muestra de 10 personas elegidas al azar:

$ProcedimientoI$	$ProcedimientoII$	$ProcedimientoIII$
5	9	7
7	6	6
6	8	9
		5

¿Puede afirmarse que las notas del test no varían según el tipo de procedimiento utilizado para formularlo? $R/$ El valor P es $P(F > 0.9497) \in]0.1, 0.9[$. Se acepta que las notas del test no varían según el tipo de procedimiento.

11. Sea X el número de piezas defectuosas fabricadas por día en una compañía. Se registraron los valores de X en 140 días, los resultados se resumen en la siguiente tabla:

$X:$	0	1	2	3
# de familias:	55	39	30	16

A un nivel de significancia del 5%, ¿Existe evidencia en contra de que X sigua la distribución $f_X(k) = \dfrac{5-k}{14}$ para $k = 0, 1, 2, 3$? $R/$ $\chi^2_{obs} = 7.81473$, no se halla evidencia en contra de que $X \sim f_X$

12. Sea X el número de niños con sobrepeso que tiene una familia con 3 hijos. Se registraron los valores de X en 100 familias, los resultados se resumen en la siguiente tabla:

$X:$	0	1	2	3
# de familias:	35	41	16	8

A un nivel de significancia del 5%, ¿Existe evidencia en contra de que X sigua la distribución $f_X(k) = C(3,k)(0.3)^k(0.7)^{3-k}$? $R/$ $\chi^2_{obs} = 0,498866213$. No se halla evidencia en contra de que $X \sim f_X$

13. Cada año, los miembros del equipo de atletismo de una universidad se dividen al azar en tres grupos que entrenan con métodos diferentes. El primer grupo realiza largos recorridos a ritmo pausado, el segundo grupo realiza series cortas de alta intensidad y el tercero trabaja en el gimnasio con pesas y en bicicleta estacionaria con pedaleo de alta frecuencia. Después de un mes de entrenamiento se realiza un test de rendimiento de la prueba de 110m con vallas. Seguidamente se muestran los tiempos obtenidos en el test en una muestra de 11 miembros:

$MetodoI$	$MetodoII$	$MetodoIII$
15	14	13
16	13	12
14	15	14
15	16	

   A un nivel de significancia del 5%, ¿Puede considerarse que los tres métodos producen resultados equivalentes? $R/\quad f_{obs} \approx 3.19, f_c = 4.459$, puede considerarse que los tres métodos producen resultados equivalentes.

14. Se realiza una investigación para determinar si las incidencias de ciertos tipos de enfermedades varían de un distrito a otro, en una ciudad grande. Los enfermedades particulares de interés son alergias, asma y rinitis La siguiente tabla muestra el número de personas enfermas en tres distritos de la ciudad durante el año pasado:

Distrito	Alergias	Asma	Rinitis
A	150	120	20
B	300	200	25
C	258	190	11

   Pruebe la hipóstesis, con un nivel de significancia del 5%, de que la ocurrencia de estos tipos de enfermedades es independiente del distrito.de la ciudad. R/(la ocurrencia de los tipos de enfermedades es dependiente del distrito

   Para ello:

   (a) Realice el procedimiento de pruebas de hipótesis y establezca sus conclusiones sobre si la ocurrencia de estos tipos de enfermedades es independiente del distrito.de la ciudad. R/la ocurrencia de los tipos de enfermedades es dependiente del distrito.

15. Ante la pandemia que se está viviendo, numerosas compañías en Costa Rica permiten que sus empleados hagan parte de su trabajo en casa. Las personas de una muestra aleatoria de 300 trabajadores se clasificaron de acuerdo al salario (en colones) y número de días de trabajo por semana que pasan en casa:

	Días de trabajo en casa por semana		
Salario	Menos de 1	Al menos uno, pero no todos	Todos en casa
Menos de 250000	38	16	14
De 250000 a 499000	54	26	12
500000 a 749999	35	22	9
Más de 750000	33	29	12

(a) ¿Presentan los datos suficiente evidencia para indicar que el salario depende del número de días de trabajo que pasan en casa? Pruebe usando $\alpha = 0.05$.R/ el salario es independiente del número de días de trabajo en casa.

(b) Calcule el valor p de la prueba e indique si este valor confirma su conclusión obtenida en el inciso **a**. R/$P = 0.3737$

16. De una caja que contiene 9 lápiceros de la misma marca y estilo, cinco de color rojo y 4 de color azul, se seleccionan al azar 4 lapiceros, y después de registrar el número Y de lapiceros rojos, se devuelven a la caja, y el experimento se repite 150 veces. Los resultados obtenidos son los siguientes: (7 puntos)

Y	0	1	2	3	4
Frecuencia	3	27	60	40	20

Pruebe la hipótesis, con un nivel de significancia de 5%, de que los datos registrados se pueden ajustar mediante una distribución hipergeométrica: R/los datos sí se ajustan mediante una distribución. hipergeometrica.

$$P(Y=k) = \frac{\binom{n}{k} \cdot \binom{N-n}{4-k}}{\binom{N}{4}}, k = 0, 1, 2, 3, 4.$$

17. En una muestra aleatoria de 360 días, el número de llamadas que recibe el Call-Center de una cierta compañía por día queda reflejada en la siguiente tabla:

Número de llamadas	Número de días
$0-5$	24
$5-10$	70
$10-15$	96
$15-20$	90
$20-25$	64
$25-30$	16

(a) Pruebe la hipótesis de que los datos se ajustan a una distribución normal con $\mu = 15$ y $\sigma = 6$. Utilice un nivel de significancia de 5%. R/ no se ajaustan a la distribución normal con $\mu = 15$ y $\sigma = 6$

18. Suponga que tres profesores impartieron un mismo curso en una cierta una universidad en un mismo semestre en modalidad virtual. El profesor A dio el curso en forma sincrónica, el profesor B en forma asincrónica, y el profesor C dio el curso mediante una combinación de ambas. El resumen de la cantidad n de estudiantes por grupo, el promedio $\overline{x}$ de las notas obtenidas por profesor, así como la desviación estándar de las notas en cada caso se da a continuación

Distrito	A	B	C
n	100	105	90
$\overline{x}$	62.8	73.4	70.2
s	15.1	14.8	18.6

Con un nivel de significancia de 5%, pruebe la hipótesis de que los promedios poblacionales en las tres modalidades son distintos. Para ello:

(a) Plantee la hipótesis nula y alternativa correspondientes al problema así como las zonas de aceptación y rechazo.

(b) Compruebe que la suma total de todas las observaciones (notas) es $T = 20305$

(c) Compruebe que el promedio global de las notas es $\overline{y} = 68.831$.

(d) Calcule la suma de variaciones entre tratamientos SSA directamente de la fórmula de la defînción dada por $SSA = \sum_{i=1}^{3} n_i \left(\overline{y_i} - \overline{y}\right)^2$ y compruebe que esta suma es igual a $SSA = 5997.9$.

(e) Calcule la suma de errores dentro de tratamientos SSE directamente de la fórmula de la defînción dada por $SSE = \sum_{i=1}^{3} \sum_{j=1}^{n_i} \left(y_{ij} - \overline{y_i}\right)^2$ y compruebe que esta suma es igual a $SSE = 76144$

(f) Complete el procedimiento de la prueba de hipótesis correspondiente y establezca sus concluisones. R/Hay evidencia de que los promedios son distintos.

7 Regresión lineal simple

1. Seguidamente se presentan los datos de 7 bebés sobre el número de días que tienen de nacidos (X) y su peso en kilogramos (Y):

X	2	7	14	21	30	60	90
Y	3.3	3.7	4	5	5.5	8	10

Suponga que se cumplen las hipótesis de regresión lineal.

(a) Encuentre una ecuación de regresión lineal para el peso de un bebé en función del número de días de nacido. $R/\quad \hat{y} = 3.162621926 + 0.077507351x$

(b) ¿Aproximadamente cuánto es el peso promedio de un bebé recién nacido? $R/\quad 3.162621926$ kg

(c) ¿Aproximadamente qué porcentaje de variación en el peso de un bebé se debe a otros factores distintos al número de días de nacido? $R/\quad 0.4329112\%$

(d) Encuentre un intervalo de confianza de 90% para el promedio de pesos de los bebés de 150 días de nacidos. $R/\quad]14.22845786, 15.34899115[$

2. En una pequeña empresa se ha estimado que el aumento promedio en las ventas mensuales es de 5 dólares por cada dólar invertido en publicidad. Para mejorar este valor se implementaron nuevas técnicas de publicidad desde hace unos meses y se han registrado los siguientes datos para algunos de estos meses:

Gastos en publicidad en dólares (X)	50	55	60	64	70
Ventas en dolares (Y)	2040	2070	2093	2120	2171

Suponga que se cumplen las hipótesis de regresión lineal.

(a) Encuentre la ecuación de regresión lineal para las ventas en función de los gastos en publicidad, a partir de la implementación de las nuevas técnicas de publicidad . $R/\quad \hat{y} = 1716.656146 + 6.390365449x$

(b) A un nivel de significancia del 5%, ¿existe evidencia de que el aumento promedio en las ventas mensuales por cada dólar invertido en publicidad es mayor ahora con las nuevas técnicas en publicidad? $R/\quad$ Si . $t_{obs} = 3.058297412$

3. En una fábrica se tiene una política de que los empleados reciban un pequeño incentivo por cada pieza elaborada (cada pieza es elaborada de forma conjunta entre los empleados). Sin embargo, se cree que desde la vigencia de la política han aumentado el número de piezas defectuosas fabricadas. Para analizas este problema se tomaran los datos del número de piezas defectuosas que se fabricaron por día (X) y el número de piezas elaboradas diarias (Y) , durante 6 días. Los datos son los siguientes:

X	7	15	11	5	20	10
Y	200	250	220	195	270	225

Suponiendo que se cumplen las hipótesis de regresión.

(a) Encuentre la ecuación de regresión lineal para el número número total de piezas elaboradas por día como función del número de piezas defectuosas que se fabrican por día. $R/\quad \hat{y} = 167.34375 + 5.234375x$

(b) Aproximadamente, ¿cuánto es el aumento promedio en el número de piezas fabricadas diarias por cada pieza defectuosa elaborada? $R/$ 5.234375 piezas por cada pieza defectuosa

(c) Aproximadamente, ¿cuál es el número promedio de piezas elaboradas en un día si ninguna es defectuosa? $R/$ 167.34375 piezas en promedio

(d) Cada empleado de la fábrica recibe un salario diario fijo de 30 mil colones más 100 colones por cada pieza que se fabrique en el día. Es decir, el salario diario de un trabajador es $S = 30000 + 100Y$. Determine un intervalo de confianza del 95% para el salario diario promedio de un trabajador si ese día se fabricaron tres piezas defectuosas.$R/\quad]47247.6 \quad , \quad 49361.8[$

(e) Considera apropiada la política de la fábrica. Explique.

(f) Aproximadamente, ¿qué porcentaje de variación de en el número de piezas elaboradas en un día se debe a otros factores distintos al número de piezas defectuosas diarias? $R/$ Aproximadamente el 2.19%.

(g) Encuentre un intervalo de confianza de 90% para el aumento promedio en el número de piezas fabricadas diarias por cada pieza defectuosa elaborada.$R/\quad]4.398653526, 6.070096474[$

4. Considere los datos de la siguiente tabla:

$X:$	1	2	4	5
$Y:$	7	12	22	32

A partir de estos datos, estime el coeficiente β de la ecuación de regresión $y = \beta x$ utilizando el método de mínimos cuadrados. $R/\quad 6.065\,22$

5. La tabla siguiente presenta las notas de aprobación en matemática discreta y programación de 6 estudiantes elegidos al azar de Ingeniería en Computación.

Matemática discreta:	35	60	93	65	87	71
Programación:	50	57	95	73	91	80

(a) Encuentre la ecuación de regresión lineal para la nota de programación como función de los nota de matemática discreta. $R/\quad \widehat{y} = 0.833217351x + 17.25794479$

(b) ¿Aproximadamente, que porcentaje de variación en las notas de programación se debe a otros factores aparte de la nota obtenida en matemática discreta? $R/\quad 0.082677751$

(c) Determine un intervalo de predicción del 95% para la nota en programación de un estudiante que obtuvo un 80 en matemática discreta. $R/\quad]66.04931621, 101.7813495[$

6. Una fabrica recolectó la siguiente información de 8 de sus trabajadores sobre el número de minutos que llegaron tarde al trabajo (X) y el número de piezas defectuosas que fabricaron ese día (Y):

X	2	5	10	15	20	25	30	40
Y	2	4	9	12	15	20	24	30

Suponiendo que se cumplen las hipótesis de regresión.

(a) Encuentre la ecuación de regresión lineal para el número de piezas defectuosas que fabrica un empleado por día como función del número de minutos que llega tarde al trabajo. $R/\quad \hat{y} = a + bx = 0.732887615 + 0.749230606x$

(b) ¿Aproximadamente, ¿cuál es el número promedio de piezas defectuosas que realiza un empleado que llega puntual a su trabajo? $R/$ 0.732887615 piezas en promedio

(c) ¿Aproximadamente, ¿cuánto es el aumento promedio en el número de piezas defectuosas que realiza un empleado por cada minuto que llega tarde al trabajo? $R/$ 0.749230606 piezas por minuto

(d) Encuentre un intervalo de predicción de 95% para el número de piezas defectuosas que realiza un empleado en un día que llega 35 minutos tarde al trabajo$R/\quad]25.00546394, 28.90645371[$

(e) Si un trabajo realiza más de 20 piezas defectuosas en un día será sancionado. ¿Considera probable que si un trabajador llega 35 minutos tarde al trabajo será sancionado? Justifique su respuesta. $R/$ Si

(f) Aproximadamente, ¿qué porcentaje de variación de en el número de piezas defectuosas que elabora un empleado en un día se debe a otros factores distintos al número de minutos que llega tarde a trabajar? $R/$ Aproximadamente el 0,42%.

(g) Un trabajador tiene un salario base diario de 20 mil colones del cuál se le descuenta 200 colones por cada pieza defectuosa que realiza. Es decir, el salario diario de un trabajador es $S = 20000 - 200Y$. Determine un intervalo de confianza del 90% para el salario diario promedio de un trabajador que llega 13 minutos tarde a su trabajo.
$R/\quad]18008.15747 \quad , \quad 17802.68833[$

7. Considere la siguiente tabla de datos, donde x es el número total de años de educación (primaria, secundaria y técnica o universitaria) y y es el ingreso anual en dólares de varias personas.

x	4	4	6	6	8	8	10	12	12	14
y	8280	10516	9212	11744	12405	14664	15336	16908	18347	19512

Encuentre un intervalo de predicción de 95% para el salario anual de una persona con once años de educación $R/\quad]13578.1921, 19312.6079[$

8. Considere los datos de la siguiente tabla:

$X:$	0	6	12	24
$Y:$	200	270	365	670

A partir de estos datos, estime el coeficiente β de la ecuación de regresión $y = 200 \cdot \beta^x$. $R/$ 1. 051 60

9. En los registros de cursos de Estadística se tiene los datos sobre el tiempo en horas dedicado a estudiar y la nota obtenida en el tercer examen.

$Nota:$	55	60	65	80	90	50	75	60	85	85	55
$Estudio:$	1	2.5	1	3	4	0.5	3	1.5	3.5	4	2

Suponga que se cumplen las hipótesis de regresión lineal.

(a) Encuentre una ecuación de regresión lineal para nota en Estadística en función de las horas de estudio. $R/\quad \hat{y} = 44,59064327 + 10,36549708x$

(b) Aproximadamente, ¿cuál es la calificación promedio de un estudiante que no estudia para el curso? $R/\quad x = 0 \Longrightarrow \hat{y} = 44,59064327$

(c) ¿Aproximadamente qué porcentaje de variación en la nota de Estadística se debe a otros factores distintos al tiempo dedicado a estudiar?$R/$ Aproximadamente el 18.16%

(d) Construya un intervalo de 95% para la calificación en el tercer examen de un estudiante que ha dedicado 3.25 horas de estudio. $R/\quad 62,76850834 < \hat{y}_{x=3.25} < 93,7885092$

(e) ¿Considera que un estudiante que estudie 3.25 horas pueda sacarse un 50 en el tercer examen?. $R/\quad NO$

10. Con el cierre técnico de una entidad financiera se estima que el tiempo (en días) de devolución de dinero depende linealmente de la cantidad de dinero (millones de colones) que el cliente tiene en su cuenta. Los siguientes datos se refieren a información de 10 clientes sobre dinero en cuenta (x) y de tiempos de espera para la devolución (y):

$$\sum x = 51, \quad \sum y = 95, \quad \sum x^2 = 433, \quad \sum y^2 = 1545, \quad \sum xy = 808$$

Suponga que se cumplen las hipótesis de regresión lineal.

(a) Encuentre una ecuación de regresión lineal para el tiempo de espera para la devolución dinero en función de la cantidad de dinero en la cuenta. $R/\quad \hat{y} = -0,042220937 + 1,871023713x$

(b) Aproximadamente, ¿cuántos días más en promedio debe esperar un cliente por la devolución de su dinero por cada millón de colones que tiene en su cuenta? $R/$ Por cada millón de colones en su cuenta un cliente debe esperar en promedio 1.871 días para la devolución de su dinero.

(c) Los jerarcas de la entidad bancaria aseguran que los clientes esperaran menos de dos días por cada millón de colones para recibir su dinero de vuelta. Con una significancia del 2%, ¿respaldan los datos esta afirmación?. $R/$ Los datos no respaldan lo que dicen los jerarcas de la entidad.

11. Considere los datos de la siguiente tabla:

$X:$	1	2	5	10
$Y:$	4	6	7	14

A partir de estos datos, estime el coeficiente α de la ecuación de regresión $y = \alpha + x$ utilizando el método de mínimos cuadrados. $R/\quad a = \frac{13}{4}$

12. Considere los datos de la siguiente tabla:

$X:$	1	4	9	10
$Y:$	3	11	28	31

A partir de estos datos, estime el coeficiente α de la ecuación de regresión $y = \beta x$ utilizando el método de mínimos cuadrados. $R/\quad \frac{203}{66}$

13. Los datos que se dan en la siguiente tabla corresponden a siete países:

País	Superficie en km^2	Población
Argentina	2776661	32664000
Bolivia	1098582	7157000
Paraguay	406752	4799000
Uruguay	186925	3121000
Venezuela	912050	20189000
Chile	757102	19116201
Costa Rica	51100	5000000

(a) Denotando con x al área en km^2 y y a la población en miles, y asumiendo una relación entre las variables x y y, explique por qué en la ecuación $y = \alpha + \beta x$, α debe ser igual a 0.

(b) Utilizando el método de mínimos cuadrados, demuestre que si, para los datos (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ se supone una relación de la forma $y = \beta x$, la estimación $\widehat{b}$ de β viene dada por

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

Para ello, defina $f(b) = \sum_{i=1}^{n} (y_i - bx)^2$ y calcule el valor de b que minimiza esta función.R/

(c) Utilice la parte **b)** para dar el valor de la estimación de β en la parte **a).** R/12.754353

(d) Interprete el valor de esta estimación en el contexto del problema. R/

La población aumenta 12.75 personas en promedio por cada km^2 que aumente la superficie.

14. Las estaturas X de diez padres con las respectivas estaturas Y de sus hijos adultos se dan en la siguiente tabla:

Estatura (X) del padre en cm	165	158.5	155	180	166	$172,5$	175	190	185	167
Estatura (Y) del hijo en cm	167	$160,2$	$145,5$	$190,6$	170	175	$189,3$	$192,1$	185	173

(a) Encuentre la ecuación de regresión lineal para la estatura y de los hijos como función de la estatura de los padres x. R/ $y = -35,216 + 1,2251x$

(b) Calcule el porcentaje de variación en la estatura de los hijos que se debe a la variación de la estaura de los padres. R/84,9%

(c) En promedio, ¿cuánto aumenta la estatura de los hijos, en relación con la estatura de los padres? R/ $1,225cm$

(d) Encuentre un intervalo de confianza de 95% para la estatura esperada del hijo cuando la estatura del padre es de 162.9 cm. R/ $]150.18, 159.19[$

8 Regresión no lineal simple

1. Para los datos en la tabla, encuentre una buena ecuación que dé y como función de x.$R/\quad \widehat{y} = x/(0.0988085x - 0.510081)$

x	-10	2	5	8	11	18
y	6	-7	-48	38	21	14

2. Considere los datos de la siguiente tabla:

$X:$	2	4	5	7	10
$Y:$	110	40	20	10	2

(a) Determine una ecuación para el modelo de regresión exponencial $y = \alpha\beta^x$.
$R/\quad \widehat{y} = 280.9609532\,(0.610654856)^x$

(b) Determine un IC del 80% para α. $R/\quad]227.450358, 347.0605979[$

(c) Pruebe, a un nivel de significacia del 10%, si el porcentaje de disminución de Y, por cada unidad de X, es del 40%. $R/\quad$ No hay evidencia en contra de que $\beta = 0.6$.

3. Se sabe que las variables x, y están relacionadas por una ecuación $y = \dfrac{x}{(\alpha - 1)\,x + (2 - \beta)}$ donde α y β son constantes. Encuentre estimaciones de α y β a partir de la siguiente tabla. $R/\quad a = 1.52296046, \quad b = 3.028528529$

x	1	4	6	20
y	-2	4	3	2

4. Una investigación indica que cierta enfermedad tuvo su aparición en los primeros días del año 1998. Seguidamente se presenta algunos datos sobre el porcentaje de personas infectadas de la enfermedad en un país C, X años después del 1° de enero del 1998:

Años (X)	0.2	2	5	10	12
Porcentaje (Y)	3%	12.7%	13.4%	13.9%	14%

(a) Justifique por qué el modelo Hiperbólico es un buen modelo de regresión para explicar la relación entre las variables X,Y.

(b) Encuentre la ecuación de regresión del modelo Hiperbólico para Y como función de X. $R/\quad \widehat{y} = \dfrac{x}{0.062475762x + 0.053998736}$

(c) Encuentre un intervalo de confianza de 90% para α (uno de los parámetros del modelo hiperbólico, según la notación vista en el curso) $R/\quad]0.053623507, \quad 0.071328017[$

(d) Encuentre un intervalo de predicción del 95% para el porcentaje de personas del país C que tendrán la enfermedad el 1° de enero del 2011. $R/\quad]10.83451446, 24.41323233[$

5. En el año 2000 inicio el correo basura ("spam") que circula por las redes de un país C. Seguidamente se presenta algunos datos sobre el porcentaje Y de correos Spam que han circulado X años después del 1° de enero del 2000:

Años (X)	0.5	2	5	8	10
Porcentaje (Y)	1%	15%	29%	29.7%	29.9%

   (a) Seleccione y justifique un modelo de regresión del porcentaje de correos Spam en función del número de años después de 2000. $R/$ Hiperbólico

   (b) Encuentre la ecuación de regresión del modelo seleccionado en la parte (**a**). $R/ \quad y = \dfrac{x}{-0.0728x + 0.5238}$

   (c) Encuentre un intervalo de confianza de 90% para α y β, parámetros del modelo escogido en **a.**

	Extremos del IC para α		Extremos del IC para β	
$R/$	izquierdo:	-0.1840	izquierdo:	0.4041
	derecho:	0.0384	derecho:	0.6435

   (d) Encuentre un intervalo de confianza del 95% para el porcentaje esperado de correos Spam que han circulado hasta el 1° de enero del 2010. $R/ \quad]-3.1531, 3.6190[$

6. Se desea realizar un estudio que relacione el tiempo n en meses de entrenamiento de un estudiante que participará en la Olimpiada Internacional de Matemáticas (OIM) y el tiempo $t$en minutos que tardará resolviendo un problema de la OIM. Según un entrenador, el tiempo mínimo que tarda un estudiante muy entrenado en resolver este tipo de problemas es de 30 minutos. Se obtuvieron los siguientes datos de ambas variables:

Tiempo de entrenamiento (n)	2	3	5	10	20
Tiempo de resolución de un problema (t)	280	195	130	80	50

   (a) Justifique por qué el modelo Recíproco es un buen modelo de regresión para t con funcón de n.
   $R/$ Recíproco

   (b) Encuentre la ecuación de regresión del modelo Recíproco para t como función de n.$R/ \quad \widehat{y} = 27.29825291 + \dfrac{505.78203}{x}$

   (c) ¿Cuántos meses de entrenamiento se necesitan para esperar resolver un problema de este tipo en dos horas o menos? $R/$ Aproxidamente, al menos 5.46 meses.

   (d) Encuentre un intervalo de confianza de 90% para β (uno de los parámetros del modelo Recíproco, según la notación vista en el curso) $R/ \quad]491.6619096, \quad 519.9021503[$

   (e) Encuentre un intervalo de predicción 90% para el tiempo que tarda un estudiante resolviendo un problema de la IMO si tuvo cuatro meses de entrenamiento.
   $R/ \quad]148.0878732, 159.3996476[$

   (f) Una prueba de la IMO consta de 3 problemas y tiempo de resolución es de 4 horas 30 minutos. Si un estudiante tuvo cuatro meses de entrenamiento, ¿considera que pueda resolver un problema de la prueba?. $R/$ Si

7. Considere los datos de la siguiente tabla:

$X:$	2	3	4	5
$Y:$	3	12	28	45

Se sabe que las variables X y Y están relacionadas por la ecuación $y+5=2x^{\beta}$, donde β es una constante desconocida. Use los datos en el cuadro anterior para estimar β, transformando la ecuación en una lineal y usando el criterio de mínimos cuadrados. $R/$ 1.99675.

8. Considere los datos muestrales de dos variables X, Y :

$X:$	3.20	3.23	3.34	3.65
$Y:$	128	43	9	3.2

   (a) Determine una ecuación para el modelo de regresión hiperbólico $y=\dfrac{x}{\alpha x+\beta}$ $R/$ $y=\dfrac{x}{2.5034x-7.9960}$

   (b) ¿Cuál es el valor aproximado esperado de Y cuando $x=3.5$? $R/$ $y=4.5694$.

   (c) Determine un IC del 90% para β. Se asume la hipótesis de regresión $R/$ $]-8.3939,-7.5981[$

9. Considere los datos muestrales de dos variables X, Y :

x	0.4	1	1.5	2
y	0.5	4	21	80

   (a) Determine una ecuación para el modelo de regresión potencial $y=\alpha x^{\beta}$ $R/$ $y=6.6183x^{3.0735}$

   (b) Determine un IC del 90% para α. Se asume la hipótesis de regresión $R/$ $]3,3118050863 ::: 13,2258620588[$

10. La siguiente tabla muestra los tiempos, en segundos, registrados por los ganadoras de los 200 metros planos (lisos) en diferentes Olímpiadas:

Año	Tiempo (en segundos)
1948	24.4
1952	23.7
1956	23.55
1964	23
1976	22.37
1980	22.19
1984	21.81
1992	21,81
2008	21,74

   (a) Si sabe que histtóricamente los tiempos en femenino no han llegado a 17,5 segundos, explique por qué un modelo de regresión exponencial apropiado para el tiempo de las campeonas como función del año es adecuado. R/$y=17.5+a*b^{x}$

(b) Encuentre la ecuación de regresión para el modelo escogido en la parte **a**. R/$y = 17.5 + 120119783.7 * 0.9914^x$

(c) Si el modelo exponencial se mantiene, ¿cuánto sería el tiempo en en el 2016? (El teimpo en 2016 fue de 21,78 s) R/20. 792

(d) ¿A cuánto tiende el mejor tiempo si el modelo se mantiene? Segerencia: es un límite.R/17.5

9 Regresión lineal múltiple

1. Se tienen las siguientes observaciones:

$X_1:$	0	1	-1	2
$X_2:$	1	0	2	3
$Y:$	15	10	20	40

Estime la ecuación de regresión de Y como función lineal de X_1,X_2. $R/\quad y = \frac{25}{4} + \frac{15}{4}x_1 + \frac{35}{4}x_2$

2. Un profesor de Estudios Sociales considera que la nota obtenida en su examen final (Y) depende linealmente del número de horas que durme un alumno en la noche antes del examen (x_1) y del número de horas de estudio en la semana previa al examen (x_2). Seguidamente se presentan los datos obtenidos de 4 estudiantes

x_1	4	1	6	8
x_2	4	8	12	15
Y	45	60	80	96

(a) Estime los coeficientes β_0, β_1 y β_2 en la ecuación de regresión lineal múltiple $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. $R/\quad y = 24.533 + 0.770\,285\,x_1 + 4.313\,65x_2$

(b) De acuerdo a la ecuación estimada, ¿Qué tiene más peso en la nota: las horas de estudio ó las horas que duerme la noche antes del examen? $R/$ Las horas de estudio.

3. Considere los siguientes valores de las variables x_1, x_2, Y :

$x_1:$	2.1	1.1	0.9	1.6
$x_2:$	3	4	5	4
$Y:$	0.43	0.31	0.32	0.46

Estime los coeficientes β_0, β_1 y β_2 en la ecuación de regresión lineal múltiple $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
$R/\quad y = -0.333559 + 0.244068x_1 + 0.0914407x_2$

4. Un estudio reciente indica que la nota obtenida en el curso de Cálculo (Y) depende linealmente de la nota obtenida en Matemática General y de la nota obtenida en el examen de admisión a la universidad. Todas las notas fueron adecuadas a la escala de 1 a 100. Seguidamente se presentan los datos obtenidos de 4 estudiantes:

$x_1:$	90	70	30	40
$x_2:$	95	82	70	50
$Y:$	100	65	44	48

(a) Estime los coeficientes β_0, β_1 y β_2 en la ecuación de regresión lineal múltiple $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. $R/\quad y = 5.316\,27 + 0.771\,296\,x_1 + 0.196\,42x_2$

(b) De acuerdo a la ecuación estimada, ¿Qué tiene más peso en la nota de Cálculo: la nota de Matemática General o la nota del examen de Admisión? $R/$ La nota de Matemática General, pues $b_1 > b_2$.

5. Para estudiar la relación de publicidad e inversión de capital con utilidades corporativas, los datos siguientes, registrados en unidades de \$100000, se recolectaron para 10 empresas de mediano tamaño en el mismo año. La variable y representa utilidad, x_1 representa inversión de capital y x_2 representa gasto en publicidad.

y	15	16	2	3	12	1	16	18	13	2
x_1	25	1	6	30	29	20	12	15	6	16
x_2	4	5	3	1	2	0	4	5	4	2

(a) Encuentre una ecuación que exprese a y como función lineal de x_1 y x_2. R/$y = -8.177 + 0.292\,13x_1 + 4.4343x_2$

Ayuda:

$$\begin{aligned}
\left(XX^t\right)^{-1} &= \begin{pmatrix} \frac{62\,431}{38\,510} & -\frac{343}{7702} & -\frac{1038}{3851} \\ -\frac{343}{7702} & \frac{13}{7702} & \frac{45}{7702} \\ -\frac{1038}{3851} & \frac{45}{7702} & \frac{226}{3851} \end{pmatrix} \\
XY^t &= \begin{pmatrix} 98 \\ 1433 \\ 383 \end{pmatrix}
\end{aligned}$$

(b) ¿Cuánto se esperaría que sea el gasto en la inversión de capital si la utilidad es \$1650000 y el gasto en publicidad es \$510000? R/7.05

10 Regresión no lineal múltiple

1. Se desea estimar los coeficientes β_0, β_1 y β_2 en la ecuación $z = \dfrac{x}{\beta_0 x + \beta_1 x^2 + \beta_2 y}$, para lo cual se obtiene la siguiente muestra de cuatro observaciones.

$X:$	1	-2	-4	-2
$Y:$	2	1	0	4
$Z:$	1	2	$-\frac{1}{2}$	$\frac{1}{2}$

   (a) Transforme el problema en uno de regresión lineal m ultiple, y escriba la tabla de datos transformados. $R/\quad 1/z = \beta_0 + \beta_1 x + \beta_2 (y/x)$.

   (b) Escriba el sistema de ecuaciones para el problema de regresión lineal múltiple.

$$R/\quad \begin{pmatrix} 4 & -7 & -\frac{1}{2} \\ -7 & 25 & 7 \\ -\frac{1}{2} & 7 & \frac{33}{4} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ 4 \\ -\frac{9}{4} \end{pmatrix}$$

   (c) ¿Cuáles son las estimaciones de β_0, β_1 y β_2 en el problema original? $R/\quad b_0 = 2$, $b_1 = 1$, $b_2 = -1$

2. Considere las siguientes observaciones:

x	0	1	2	3
y	5	3	2	1

   Estime los coeficientes β_0, β_1 y β_2 en la ecuación $y = \dfrac{1}{\beta_0 + \beta_1 x + \beta_2 x^2}$. $R/\quad b_0 = 0.214\,995$, $b_1 = -0.018\,355$, $b_2 = 0.091\,675$

3. Se desea estimar los coeficientes β_0, β_1 y β_2 en la ecuación $z = \dfrac{\beta_0 x + \beta_1 + \beta_2 y}{x}$, para lo cual se obtiene la siguiente muestra de cuatro observaciones.

$X:$	1	0.5	0.25	0.0625
$Y:$	1	2	3	4
$Z:$	5	15	43	228

   (a) Transforme el problema en uno de regresión lineal múltiple, y escriba la tabla de datos transformados. $R/\quad z = \beta_0 + \beta_1 (1/x) + \beta_2 (y/x)$

   (b) Escriba el sistema de ecuaciones para el problema de regresión lineal múltiple.

$$R/\quad \begin{pmatrix} 4 & 23 & 81 \\ 23 & 277 & 1081 \\ 81 & 1081 & 4257 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 291 \\ 3855 \\ 15\,173 \end{pmatrix}$$

(c) ¿Cuáles son las estimaciones de 0, 1 y 2 en el problema original? $R/\quad b_0 \approx 3.30165$, $b_1 \approx -2.39463$, $b_2 \approx 4.1095$

4. Considere las siguientes observaciones:

x	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{2}$	1
y	0.008	0.02	0.04	0.08

Estime los coeficientes β_0, β_1 y β_2 en la ecuación $y = \dfrac{x^2}{\beta_0 + \beta_1 x + \beta_2 x^2}$. $R/\quad b_0 = 4.82955$, $b_1 = -0.738\,636$, $b_2 = 8.068\,18$

5. Dada las siguientes observaciones, estime los coeficientes en la ecuación $z = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 y$ R/ $z = 14.546 + 62.917x - 5.183x^2 - 13.733y$

x	2	4	5	7	11	15	16	21
y	3	5	8	9	12	16	21	25
z	80	117	91	70	-90	-421	-590	-1297