ESTADÍSTICA EMPRESARIAL II

GUIA DE TRABAJOS PRACTICOS

Departamento de Matemática y Métodos Cuantitativos

01/01/2016

CONTENIDOS

PRACTICO 1. DISTRIBUCIONES EN EL MUESTREO

Consideraciones generales

Las medidas de posición y dispersión calculadas a partir de una población se denominan parámetros y son valores únicos, constantes, es decir que no cambian a menos que la población lo haga.
Los mismos estadísticos, pero calculados a partir de una muestra (de tamaño n), dependen de ella, por lo tanto varían al cambiar la muestra. Reciben el nombre de estimadores y son variables aleatorias.
La distribución en el muestreo resulta ser entonces la ley de distribución de dichas variables que hemos denominado estimadores. Nos interesará por lo tanto obtener además su valor esperado y su variancia.
En esta primera parte nos referiremos en particular a la media muestral y a la proporción muestral; en ambos casos demostraremos que su distribución tiende a la distribución normal cuando $n \to \infty$.
Se indica a continuación la forma de caracterizar las medidas poblacionales y las correspondientes muestrales:

Medida	Parámetro	Estimador
media aritmética	μ	$\bar{x}$
variancia	σ^2	s^2
desvío estándar	σ	s
proporción	p	$\hat{p}$

Problemas resueltos:

Distribución normal (Revisión del uso de tablas, aplicativo y Excel):

La producción diaria de una fábrica es una variable normalmente distribuida con promedio igual a 54 kilogramos y desvío estándar igual a 7 kilogramos.

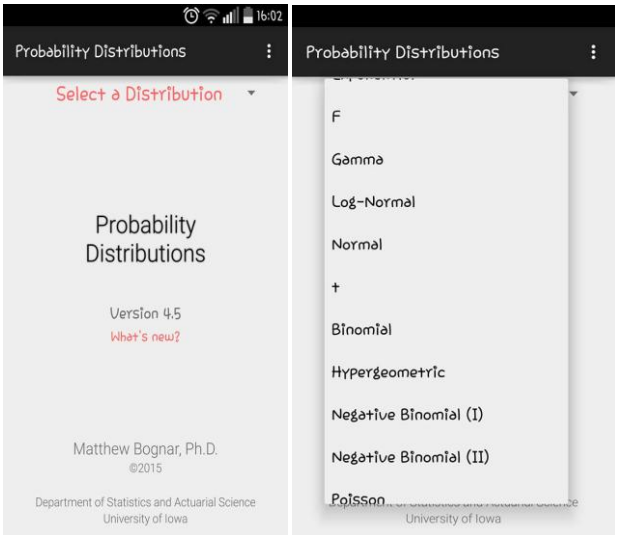
a) Calcular la probabilidad de que la producción de un día resulte inferior a 65 kilogramos.
b) ¿Cuál es la probabilidad de que mañana se produzcan más de 47 kilogramos?
c) ¿En qué porcentaje de los días se produce entre 50 kilogramos y 60 kilogramos?
d) ¿En qué porcentaje de los días la producción supera los 20 kilogramos?
e) ¿Cuál es la producción no superada en el 30% de los días?
f) ¿Cuál es la producción sólo superada en la cuarta parte de los días?
g) El encargado de producción eleva un informe al dueño de la fábrica comprometiéndose a producir diariamente determinada cantidad de kilogramos como mínimo. ¿En cuánto debe fijar dicha cantidad mínima para que la probabilidad de cumplir con el compromiso resulte igual a 0,95?

Solución:

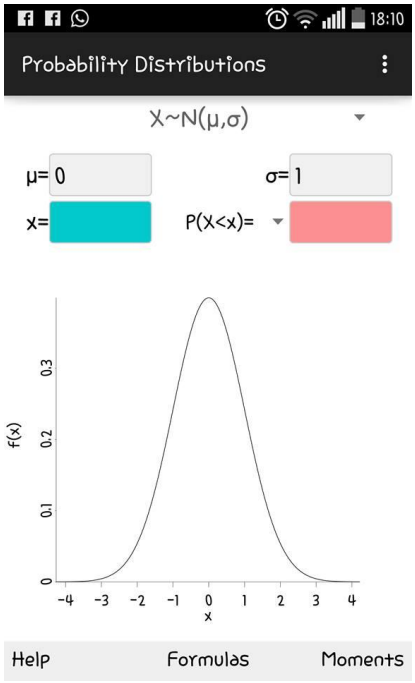
a) $P_N(x<65/\mu=54;\sigma=7)=P_N\left(z<\frac{65-54}{7}\right)=P_N(z<1{,}57)=F(1{,}57)=0{,}94179$

Para el caso de la utilización del aplicativo "Probability Distributions", en principio deberán bajar la aplicación en sus celulares y/o tablets. El ícono que la representa es:

Una vez que la hayan instalado, si la abren, encontrarán la siguiente pantalla, donde seleccionarán la distribución de probabilidad a utilizar, para este caso la normal:



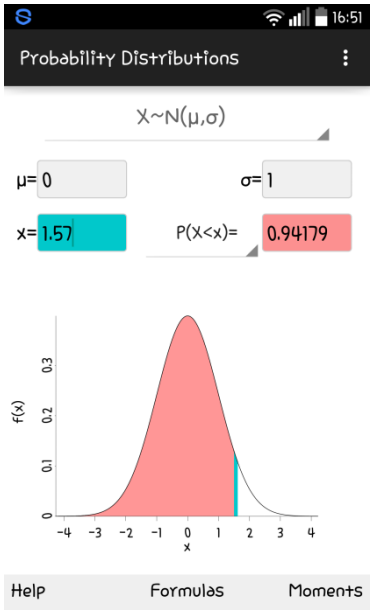
Una vez seleccionada la distribución Normal aparecerán:



Recordemos que por defecto, la aplicación utiliza los parámetros de la distribución normal estándar, es decir, una media de 0 y un desvío de 1. No modificaremos estos valores ya que estamos interesados en trabajar con la normal estándar.

En el rectángulo celeste cargaremos los valores de Z que vayamos calculando para obtener las correspondientes probabilidades. Una vez ingresado el valor de Z, le solicitamos a la app, si deseamos el área a izquierda (P(X<x)) o el área a la derecha (P(X>x)), dependiendo del caso. Podremos ver, que el gráfico nos mostrará la superficie calculada.

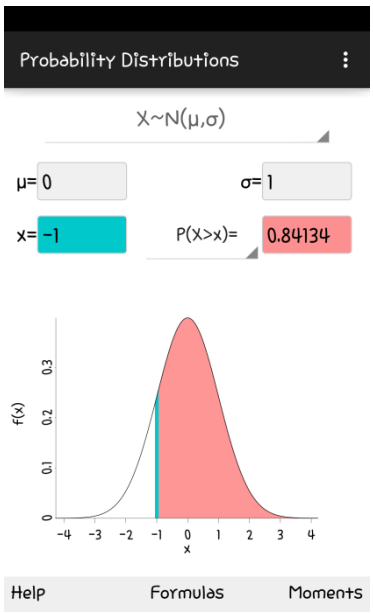
Para la pregunta a) indicamos el valor de Z, previamente calculado, y solicitamos la probabilidad acumulada a la izquierda, llegando de esta manera al resultado.



b) $P_N(x > 47 / \mu = 54 ; \sigma = 7) = P_N\left(z < \frac{47-54}{7}\right) = P_N(z > -1) = 1 - F(-1)$

$= 1 - 0{,}15866 = 0{,}84134$

En esta pregunta, la operación es similar a la utilizada en la pregunta a) con la diferencia de que pediremos la probabilidad acumulada por derecha obteniendo automáticamente el resultado.

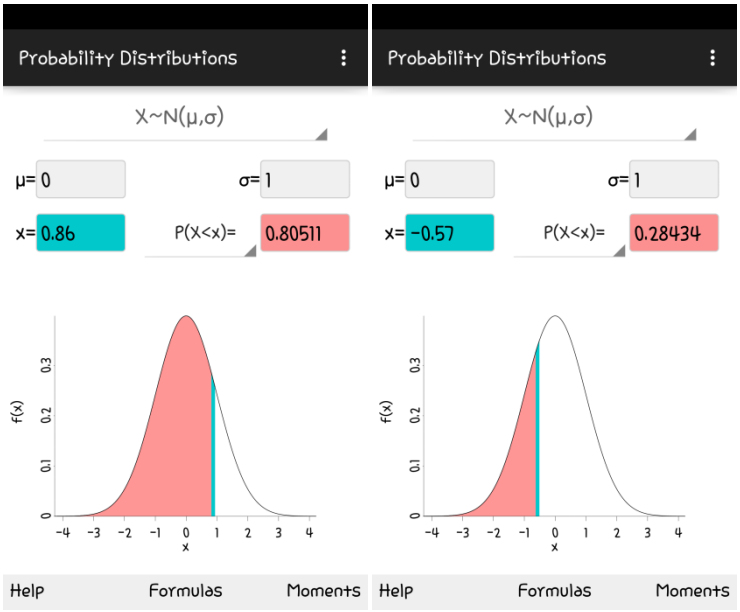


c) $P_N(50 \le x \le 60 / \mu = 54 ; \sigma = 7) = P_N\left(\frac{50-54}{7} \le z \le \frac{60-54}{7}\right) = P_N(-0{,}57 \le z \le 0{,}86)$

$= F(0{,}86) - F(-0{,}57) = 0{,}80511 - 0{,}28434 = 0{,}52077$

Respuesta: 52%

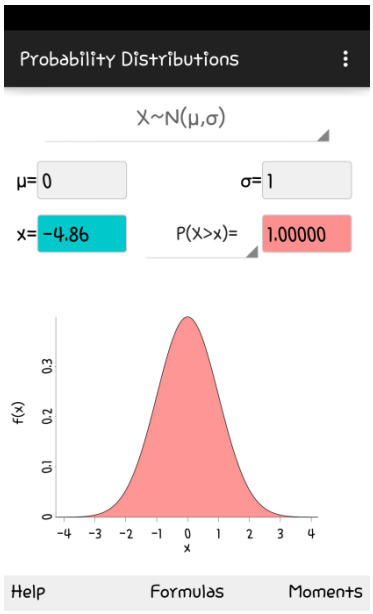
Recordemos que, un área entre dos valores de la variable se calcula restando las probabilidades acumuladas entre ellos, generalmente a la superficie acumulada a la izquierda del valor más grande se le resta la superficie acumulada a la izquierda del valor más pequeño (no siendo esta la única variante, pero si la más utilizada).



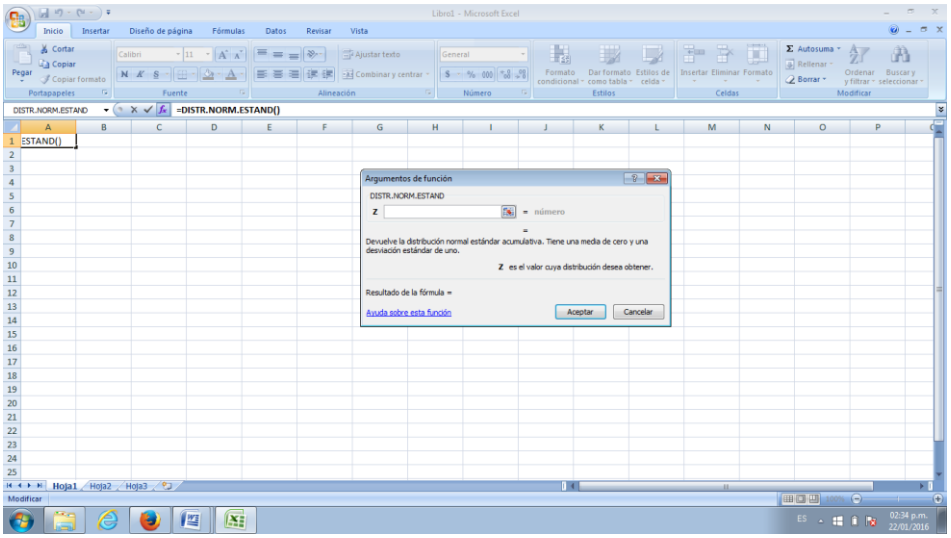
d) $P_N(x > 20 / \mu = 54 ; \sigma = 7) = P_N\left(z < \frac{20-54}{7}\right) = P_N(z > -4{,}86) = 1 - F(-4{,}86) = 1 - 0 = 1$

Respuesta: 100%

La pregunta en cuestión, es igual a la pregunta a), debido a que pide la probabilidad acumulada por izquierda. A diferencia de la tabla, con la aplicación, podemos obtener los valores de probabilidad para cualquier valor de Z. Observemos que la superficie para dicho valor comprende la totalidad de la campana, es decir el 100%. Una de las ventajas del aplicativo será observar si los resultados obtenidos son lógicos si realizan la lectura gráfica que la misma les proporciona.



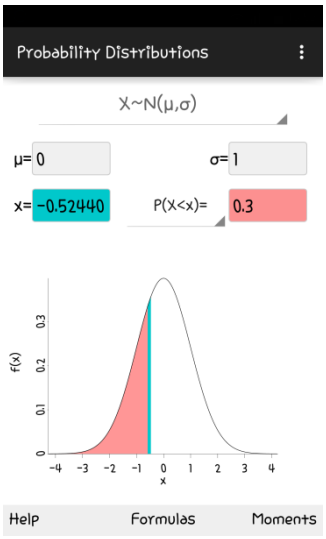
Las primeras cuatro preguntas del ejercicio también pueden ser resueltas utilizando el Excel.
En la opción de *f(x)*, pueden definir, las funciones estadísticas y encontrarán la variante llamada: **=DISTR.NORM.ESTAND**



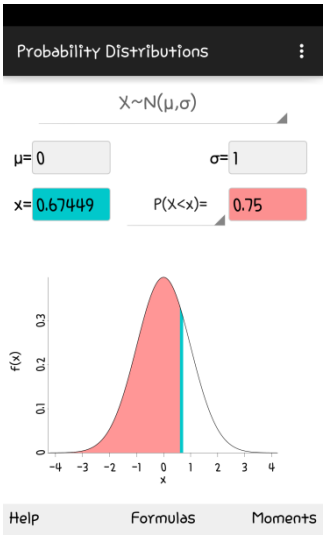
En la cual definiendo el valor de z adecuado, obtendrán el área acumulada a la izquierda del mismo. En la caso de necesitar un área a la derecha se debe proceder como con la tabla.

e) $F(z) = 0{,}30 \qquad z = -0{,}524 \qquad x = z \cdot \sigma + \mu = -0{,}524 \cdot 7 + 54 = 50{,}332 \; kgs$

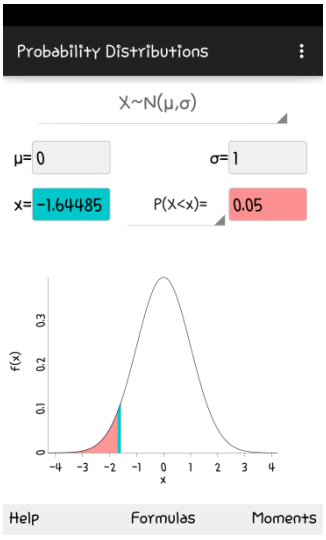
Si en cambio disponemos de los valores de probabilidad y deseamos conocer el valor de Z, en el aplicativo, ingresaremos nuestro dato (la probabilidad) en el rectángulo rosado, indicando que es una probabilidad acumulada a la izquierda (P(X<x)).



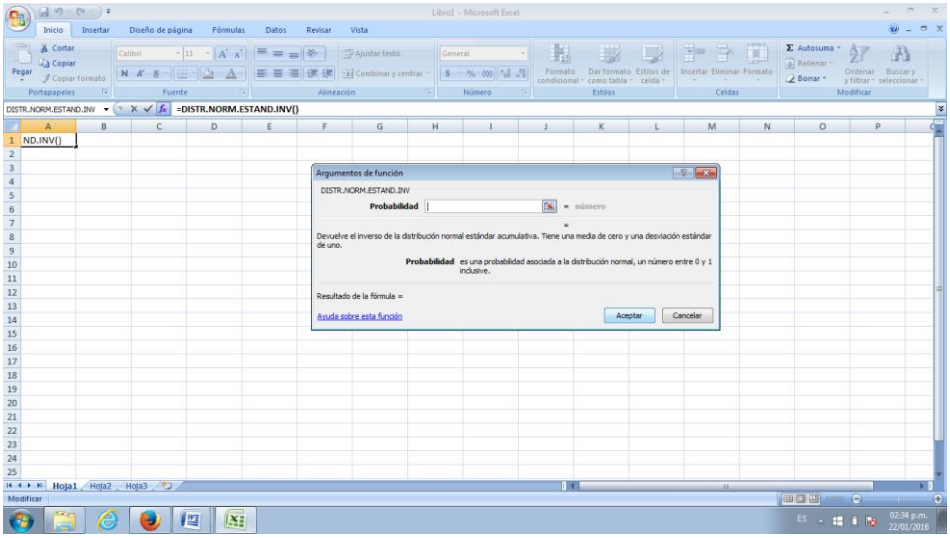
f) $F(z) = 1 - 0{,}25 = 0{,}75 \qquad z = 0{,}674 \qquad x = z \cdot \sigma + \mu = 0{,}674 \cdot 7 + 54 = 58{,}718\ kgs$



g) $F(z) = 1 - 0{,}95 = 0{,}05 \qquad z = -1{,}645 \qquad x = z \cdot \sigma + \mu = -1{,}645 \cdot 7 + 54 = 42{,}485\ kgs.$



En las últimas tres preguntas, si se desea trabajar con la planilla de cálculo tenemos que invertir la función estadística, esto es **=DISTR.NORM.ESTAND.INV**



e indicar la probabilidad acumulada a la izquierda para obtener el valor de Z asociado.

Distribución de la media muestral (con variancia poblacional conocida):

Una profesora de estadística ha determinado que el tiempo necesario para que los estudiantes concluyan un examen final se distribuye normalmente con media igual a 84 minutos y desvío estándar igual a 18 minutos.
Si se toma una muestra de 9 estudiantes, ¿cuál es la probabilidad de que el tiempo promedio de finalización de sus respectivos exámenes no exceda los 90 minutos?
Si se extrajeran muestras de 4 estudiantes para calcular el tiempo promedio de finalización de los exámenes, ¿cuál sería el tiempo promedio sólo superado en el 5% de dichas muestras?

Solución:

Datos: $x: \textit{tiempo en minutos por examen}$ $\mu_x = 84$ $\sigma_x = 18$

a) $n = 9$ *estudiantes*

$\bar{x}: \textit{tiempo promedio de la muestra}$ $\mu_{\bar{x}} = 84$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{9}} = \frac{18}{3} = 6$

$$P_N\left(\bar{x} \le 90\right) = P_N\left(z \le \frac{90-84}{6}\right) = P_N\left(z \le 1\right) = F(1) = 0{,}84134$$

b) $n = 4$ *estudiantes*

$\bar{x}: \textit{tiempo promedio de la muestra}$ $\mu_{\bar{x}} = 84$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{4}} = \frac{18}{2} = 9$

$F(z) = 1 - 0{,}05 = 0{,}95$ $z = 1{,}645$ $\bar{x} = 1{,}645 \cdot 9 + 84 = 98{,}805 \text{ min } utos$

Respuesta: 98,8 minutos

Distribución de la proporción muestral:

En una localidad, donde el 36% de las mujeres casadas trabaja fuera de su hogar, se entrevistarán 256 mujeres casadas seleccionándolas al azar.

a) Calcular la probabilidad de que más del 40% de las entrevistadas trabaje fuera de su hogar.
b) Si se entrevistaran 400 mujeres casadas, ¿cuál sería la probabilidad de encontrar a lo sumo 244 que no trabajen fuera de su hogar?

Solución:

a) n = 256 p = 0,36 (proporción de mujeres casadas que trabajan)
q = 1 – p = 1 – 0,36 = 0,64

$\hat{p}$: *proporción de mujeres que trabajan en la muestra*

$$\mu_{\hat{p}} = 0{,}36 \qquad \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}36 \cdot 0{,}64}{256}} = 0{,}03$$

$$P_N(\hat{p} > 0{,}40) = P_N\left(z > \frac{0{,}40 - 0{,}36}{0{,}03}\right) = P_N(z > 1{,}33) = 1 - F(1{,}33) = 1 - 0{,}90824 = 0{,}09176$$

b) n = 400 p = 0,64 (proporción de mujeres casadas que no trabajan)
q = 1 – p = 1 – 0,64 = 0,36

$\hat{p}$: *proporción de mujeres que no trabajan en la muestra*

$$\mu_{\hat{p}} = 0{,}64 \qquad \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}64 \cdot 0{,}36}{400}} = 0{,}024 \qquad \hat{p}_0 = \frac{244}{400} = 0{,}61$$

$$P_N(\hat{p} \le 0{,}61) = P_N\left(z \le \frac{0{,}61 - 0{,}64}{0{,}024}\right) = P_N(z \le -1{,}25) = F(-1{,}25) = 0{,}10565$$

Ejercitación:

Problema 1: Sabiendo que los montos abonados mensualmente en concepto de comisiones por ventas en la firma "Norte S. A." se distribuyen normalmente con un promedio igual a US$ 4.700 y un desvío estándar igual a US$ 640, contestar:

a) ¿Cuál es la probabilidad de que en un mes se abone como máximo US$ 5.400 en concepto de comisiones por ventas?
b) ¿Cuál es la probabilidad de que en un mes se abone como mínimo US$ 4.900 en concepto de comisiones por ventas?
c) ¿Cuál es la probabilidad de que en un mes se abone entre US$ 4.000 y US$ 5.000 en concepto de comisiones por ventas?
d) ¿Cuál es la probabilidad de que en un mes se abone a lo sumo US$ 8.000 en concepto de comisiones por ventas?

e) ¿En qué porcentaje de los meses el monto abonado en concepto de comisiones por ventas supera los US$ 1.000?
f) ¿En qué porcentaje de los meses el monto abonado en concepto de comisiones por ventas queda comprendido entre US$ 4.200 y US$ 7.500?
g) ¿Cuál es el monto no superado en el 24 % de los meses?
h) ¿Cuál es el monto sólo superado en el 15 % de los meses?
i) ¿Qué monto se debe reservar para pagar las comisiones por ventas del mes próximo, si se desea que la probabilidad de que dicha reserva resulte suficiente sea igual a 0,88?

RESPUESTAS: a) 0,86214 b) 0,37828 c) 0,54 d) 1
e) 100 % f) 78,23 % g) US$ 4.248 h) US$ 5.363 i) US$ 5.452

Problema 2: Cierta empresa de radiotaxis ha calculado un gasto de mantenimiento promedio por cada unidad de 1250 $ por mes, con un desvío estándar de 650 $. En una muestra de 50 vehículos de la empresa,

a) Indique el promedio, variancia y distribución de la variable "gasto promedio de mantenimiento mensual por unidad" para la muestra de 50 taxis.
b) Calcule la probabilidad de que en un mes la empresa gaste en promedio en mantenimiento más de 1300 $ por unidad.
c) Calcule la probabilidad de que la empresa gaste a lo sumo 1100 $ promedio mensuales en mantenimiento por unidad.
d) Calcule cuánto se gastará en promedio como mínimo en el 40% de los meses.
e) Complete las siguientes frases:
    1. En el 40% de los meses el gasto promedio de mantenimiento es de a la sumo …….. $.
    2. En promedio en el 80% de los meses se gasta en mantenimiento de la flotilla como mínimo ……. $ por unidad
    3. En el …..% de los meses se gasta en el mantenimiento de la flotilla más de 140$ promedio por unidad.

RESPUESTAS: a) $ 1250, ^{2} 8450, normal b) 0.2932
c) 0.0514 d) $ 1273.3 e1)$ 1226.7; e2) $ 1172.6; e3) 5.14%

Problema 3: Una multinacional llevó a cabo un relevamiento de los sueldos anuales de sus gerentes generales en los 14 países donde opera. Los resultados (en miles de US$) fueron:

125 79 82 62 109 158 102 55 120 105 91 88 104 100

a) Calcule el promedio y el desvío estándar de los sueldos anuales de todos los gerentes (¿son poblacionales o muestrales?)
b) Extraiga una muestra aleatoria de 5 sueldos y calcule su promedio y su desvío estándar (¿son poblacionales o muestrales?). Si tomara otra muestra de 5 sueldos,

también al azar, ¿se mantendrían el promedio y el desvío? ¿Cómo se comporta entonces $\bar{x}$?

c) Indique qué error se está cometiendo si se estima la media poblacional basándose en alguna de las medias calculada en el ítem anterior.

RESPUESTAS: a) μ = US$ 98571 ; σ = US$ 25278 b) $\bar{x}$ es una variable aleatoria
c) error muestral

Problema 4: El peso de paquetes de café envasados automáticamente tiene distribución normal con un promedio de 500 gramos y un desvío típico de 12 gramos.

a) Se selecciona un paquete al azar. Calcular la probabilidad de que dicho paquete pese entre 494 gramos y 506 gramos.
b) Se extrae una muestra al azar de 9 paquetes. Calcular la probabilidad de que el peso promedio de dicha muestra quede comprendido entre 494 gramos y 506 gramos.
c) Se extrae una muestra al azar de 25 paquetes. Calcular la probabilidad de que el peso promedio de dicha muestra quede comprendido entre 494 gramos y 506 gramos.
d) Se extrae una muestra al azar de 100 paquetes. Calcular la probabilidad de que el peso promedio de dicha muestra quede comprendido entre 494 gramos y 506 gramos.
e) Represente gráficamente la distribución de probabilidades de la media muestral en cada uno de los puntos anteriores. ¿Cambia la forma, el centro o la dispersión de la variable?

RESPUESTAS: a) 0,38292 b) 0,86638 c) 0,98758 d) ≈ 1

Problema 5: Calcular la probabilidad de que la longitud media de una muestra de 50 piezas de precisión supere los 5,34 mm, sabiendo que la longitud de dichas piezas se distribuye normalmente con promedio igual a 5 mm y variancia igual a 2,25 mm^2.

RESPUESTA: 0,0548 (aproximadamente 5 %)

Problema 6: Se extrae una muestra al azar de 16 elementos de una población normal con desvío estándar igual a 6. Calcular la probabilidad de que el promedio de dicha muestra difiera del promedio poblacional en 2 unidades como máximo.

RESPUESTA: 0,81648

Problema 7: Sabiendo que una fábrica de repuestos produce con un 13% de defectuosos, contestar:

a) Si se toma una muestra de 200 repuestos, ¿cuál es la probabilidad de que la proporción de repuestos defectuosos en la muestra resulte mayor a 0,10?

b) Recalcular la probabilidad pedida en el punto anterior para una muestra de 500 repuestos.

RESPUESTAS: a) 0,89617 b) 0,97725

Problema 8: En un sindicato donde el 20 % de los afiliados tiene menos de 25 años de edad, se seleccionan al azar 400 afiliados para efectuar una encuesta acerca de la aplicación de las normas de seguridad en sus respectivos lugares de trabajo. Calcular:

a) La probabilidad de que la proporción de afiliados menores de 25 años seleccionados para la encuesta resulte inferior a 0,14.
b) La probabilidad de que el porcentaje de encuestados con 25 años de edad como mínimo resulte inferior al 85 %.

RESPUESTAS: a) 0,00135 b) 0,99379

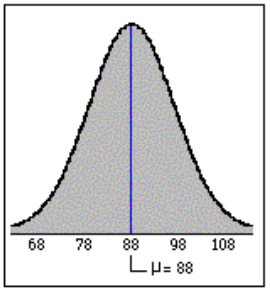
Problema 9: En una universidad donde el 32 % de los alumnos son mujeres, se tomará una muestra de 240 alumnos. Calcular la probabilidad de que el porcentaje de mujeres en dicha muestra difiera en más de 3 puntos del porcentaje de mujeres en la universidad.

RESPUESTA: 0,31732

Revisión conceptual

Responda las siguientes preguntas:

a) ¿Cuál es la diferencia entre parámetro y estimador?
b) ¿Qué es un estimador insesgado? De algún ejemplo.
c) ¿Cuál es la diferencia entre una estimación eficiente y una ineficiente?
d) ¿Qué estadístico usaría para estimar la media poblacional? ¿La media de la muestra o la mediana? ¿Por qué?
e) Un contador selecciona una muestra aleatoria de 100 cuentas bancarias y resulta que promediadas dan un saldo de 725,80$. El contador afirma que ese será el saldo de TODAS las cuentas del banco ya que el promedio muestral es un estimador insesgado de la media de una población. El contador creía saber estadística pero... en esto estaba equivocado, ¿por qué?
f) La figura de la derecha muestra la distribución de probabilidades de la variable X = largo del fruto (en mm) de cierta especie leguminosa. Para una muestra de n = 5 frutos elegidos al azar se define la variable aleatoria $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$



Indique cuál de las siguientes afirmaciones con respecto a la distribución de probabilidades de $\overline{X}$ es verdadera, justificando sus dichos:

1. La distribución de probabilidades no es normal ya que el

tamaño de la muestra no es lo suficientemente grande
2. Tiene una esperanza igual a 0
3. Tiene una varianza 5 veces menor
4. La forma de la distribución es más aplanada que la de la figura

PRACTICO 2. INTERVALOS DE CONFIANZA PARA UNA POBLACIÓN

Consideraciones generales:

La inferencia estadística utiliza la información que proporciona la muestra a través de sus estimadores para concluir sobre los parámetros de la población (de la cual se extrajo la muestra).
La Inferencia Estadística está formada por dos grandes capítulos: la estimación y el ensayo o prueba de hipótesis.
En primer lugar nos referiremos a la media muestral, estimador de la media poblacional, bajo el supuesto de que se conoce la variancia poblacional.
En segundo lugar trabajaremos en inferencia para el promedio poblacional, desconociendo la variancia poblacional e introduciendo la distribución t de Student.
En tercer lugar nos referiremos a la inferencia sobre la proporción poblacional para muestras grandes.
Por último veremos inferencia sobre la variancia poblacional e introduciremos la distribución chi cuadrado.
En todos estos casos trabajaremos con una muestra aleatoria que será obtenida de la población que se desea estudiar.

Problemas resueltos:

Intervalo de confianza para el promedio con variancia poblacional conocida:

Un fabricante de líquido acondicionador para el cabello, que comercializa su producción a través de una gran cantidad de negocios minoristas de su zona, alarmado por la ostensible baja de las ventas debida a la profunda recesión económica, decide cambiar la presentación de su producto, sustituyendo el envase tradicional por otro de menor costo que le permitirá reducir el precio de venta. Para evaluar los efectos del cambio efectuado, visita 49 minoristas y verifica que durante la semana anterior vendieron en total 1.323 litros de acondicionador. Suponiendo que la variancia de las ventas semanales por minorista, que antes del cambio de presentación era igual a 156,25 litros^2, no ha variado:

a) Construir un intervalo de confianza del 95% para estimar el nuevo promedio semanal de ventas por minorista.
b) Si, basándose en la misma muestra, se efectuara una estimación de dicho promedio con un error de muestreo igual a 1,50 litros ¿cuál sería el nivel de confianza de la misma?
c) ¿Cuántos minoristas más habría que incluir en la muestra para lograr que el error de muestreo de la estimación no supere los 1,50 litros y que el NC siga siendo igual a 95%?

Solución:

a) Datos: $n = 49$ $\sum x = 1.323$ $\sigma^2 = 156,25$ $NC = 95\%$

Intervalo de confianza: P{Límite Inferior ≤ μ ≤ Límite Superior } = 1 – α

$$\bar{x} = \frac{\sum x}{n} = \frac{1.323}{49} = 27 \qquad \sigma = \sqrt{156,25} = 12,5 \qquad \alpha = 1 - NC = 1 - 0,95 = 0,05$$

$$\alpha/2 = 0,05/2 = 0,025 \qquad z_{\alpha/2} = z_{0,025} = -1,96 \qquad 1 - \alpha/2 = 1 - 0,025 = 0,975 \qquad z_{1-\alpha/2} = z_{0,975} = 1,96$$

$$Límite\ inferior = LI = \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n} = 27 + (-1,96) \cdot 12,5/\sqrt{49} = 27 - 1.96 \cdot 12,5/7 = 27 - 3,5 = 23,50$$

$$Límite\ superior = LS = \bar{x} + z_{1-\alpha/2} \cdot \sigma/\sqrt{n} = 27 + 1,96 \cdot 12,5/\sqrt{49} = 27 + 1,96 \cdot 12,5/7 = 27 + 3,5 = 30,50$$

En resumen:

$$\bar{x} \pm E \qquad donde: \qquad E = z_{1-\alpha/2} \cdot \sigma/\sqrt{n} = 1,96 \cdot 12,5/\sqrt{49} = 3,5$$

$$Respuesta: \qquad 27 \pm 3,5 \qquad \Rightarrow \qquad [23,50\ ;\ 30,50]$$

b) Datos: n = 49 σ = 12,5 E = 1,50 NC = ?

$$E = z_{1-\alpha/2} \cdot \sigma/\sqrt{n} \qquad 1,50 = z_{1-\alpha/2} \cdot \frac{12,5}{\sqrt{49}} \qquad z_{1-\alpha/2} = \frac{1,50 \cdot 7}{12,5} = 0,84 \qquad F(0,84) = 0,80$$

$$0,8 = 1 - \alpha/2 \qquad \alpha/2 = 1 - 0,8 = 0,2 \qquad \alpha = 2 \cdot 0,2 = 0,4 \qquad NC = 1 - \alpha = 1 - 0,4 = 0,6$$

Respuesta: El nivel de confianza sería igual a 60%

c) Datos: E = 1,50 σ = 12,5 NC = 95% n = ?

$$NC = 0,95 \qquad \Rightarrow \qquad z_{1-\alpha/2} = 1,96 \qquad entonces: \qquad E = 1,50 = \frac{1,96 \cdot 12,5}{\sqrt{n}}$$

$$n = \left[\frac{z_{1-\alpha/2} \cdot \sigma}{E}\right]^2 = \left[\frac{1,96 \cdot 12,5}{1,5}\right]^2 = 266,78 \qquad n = 267 \qquad 267 - 49 = 218$$

Respuesta: Habría que incluir 218 minoristas más.

Intervalo de confianza para el promedio con variancia poblacional desconocida:

Para estimar el coeficiente intelectual (CI) promedio de los alumnos de una universidad se toma una prueba a una muestra de 6 estudiantes obteniéndose los siguientes resultados:
128 – 117 – 125 – 136 – 110 – 134
Suponiendo que los CI siguen una distribución normal:

a) Efectuar la estimación con un riesgo del 5%.
b) ¿Cuál debe ser el tamaño de muestra para que, manteniendo el mismo nivel de confianza, el error de muestreo de la estimación anterior sea igual a 4?

Solución:

a) Datos: n = 6 x: 128 – 117 – 125 – 136 – 110 – 134 $\alpha = 5\%$

$$\bar{x} = \frac{\sum x}{n} = \frac{128+117+125+136+110+134}{6} = \frac{750}{6} = 125$$

$$S^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{(128-125)^2 + \cdots\cdots + (134-125)^2}{6-1} = \frac{9+64+0+121+225+81}{5} = \frac{500}{5} = 100$$

$$S = \sqrt{100} = 10$$

$$NC = 1-0{,}05 = 0{,}95 \qquad \alpha = 0{,}05 \qquad \alpha/2 = 0{,}025 \qquad 1-\alpha/2 = 1-0{,}025 = 0{,}975$$

$$\nu = grados\ de\ libertad = GL = n-1 = 6-1 = 5$$

$$\bar{x} \pm E \qquad E = t_{\nu;1-\alpha/2} \cdot S/\sqrt{n} = t_{5;0,975} \cdot 10/\sqrt{6} = 2{,}571 \cdot 10/2{,}4495 = 10{,}5 \qquad 125 \pm 10{,}5$$

Respuesta: [125 – 10,5 ; 125 + 10,5] = [114,50 ; 135,50]

b) Datos: E = 4 NC = 95% n = ?

$$\nu = n-1 = 6-1 = 5 \qquad t_{5;0,975} = 2{,}571 \qquad n = \left(\frac{t_{\nu;1-\alpha/2} \cdot S}{E}\right)^2 = \left(\frac{2{,}571 \cdot 10}{4}\right)^2 = 41{,}3 \qquad n_1 = 42$$

$$\nu = n-1 = 42-1 = 41 \qquad t_{41;0,975} = 2{,}020 \qquad n = \left(\frac{t_{\nu;1-\alpha/2} \cdot S}{E}\right)^2 = \left(\frac{2{,}020 \cdot 10}{4}\right)^2 = 25{,}5 \qquad n_1 = 26$$

$$\nu = n-1 = 26-1 = 25 \qquad t_{25;0,975} = 2{,}060 \qquad n = \left(\frac{t_{\nu;1-\alpha/2} \cdot S}{E}\right)^2 = \left(\frac{2{,}060 \cdot 10}{4}\right)^2 = 26{,}5 \qquad n_1 = 27$$

$$\nu = n-1 = 27-1 = 26 \qquad t_{41;0,975} = 2{,}056 \qquad n = \left(\frac{t_{\nu;1-\alpha/2} \cdot S}{E}\right)^2 = \left(\frac{2{,}056 \cdot 10}{4}\right)^2 = 26{,}4 \qquad n_1 = 27$$

Respuesta: El tamaño de la muestra debería ser igual a 27 alumnos.

Intervalo de confianza para la proporción:

En un importante supermercado, que cuenta con varias sucursales en distintos puntos del país, se está estudiando la incidencia de las tarjetas de débito como medio de pago. A tal efecto, se ha analizado una muestra de 125 compras efectuadas durante el último fin de semana en el local ubicado en Constitución, observándose que 79 fueron abonadas con tarjetas de débito y el resto con otros medios de pago.

a) Estimar la proporción de compras que se abonan con tarjetas de débito en la sucursal Constitución. (Utilizar $\alpha = 2\%$).

b) Determinar el tamaño de muestra necesario para reducir en un 40% el error de muestreo de la estimación anterior.

c) ¿Cuántas compras deberían analizarse en la sucursal Mar del Plata para estimar la proporción en estudio con un error de muestreo que no supere el 3% y una confianza del 92%? Cabe aclarar que se carece de datos previos acerca de la incidencia del pago con tarjetas de débito en esta sucursal.

Solución:

a) Datos: n = 125 r = 79 α = 2%

$$\hat{p} = \frac{r}{n} = \frac{79}{125} = 0{,}632 \qquad \hat{q} = 1 - \hat{p} = 1 - 0{,}632 = 0{,}368 \qquad 1 - \alpha/2 = 1 - 0{,}01 = 0{,}99$$

$$E = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = z_{0{,}99} \cdot \sqrt{\frac{0{,}632 \cdot 0{,}368}{125}} = 2{,}326 \cdot 0{,}043 = 0{,}10 \qquad \hat{p} \pm E = 0{,}632 \pm 0{,}10$$

Respuesta: [0,532 ; 0,732]

b) Datos: $\hat{p} = 0{,}632$ α = 2% E(anterior) = 0,10 n = ?

Reducir el error anterior en un 40%: nuevo error = E = 0,10.0,6 = 0,06

$$n = \frac{z_{1-\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2} = \frac{z_{0{,}99}^2 \cdot 0{,}632 \cdot 0{,}368}{0{,}06^2} = \frac{2{,}326^2 \cdot 0{,}232576}{0{,}0036} = 349{,}5279 \qquad n = 350$$

Respuesta: El tamaño de muestra necesario es igual a 350 compras.

c) Datos: NC = 92% E = 0,03 n = ?

Al carecerse de datos acerca de la proporción, se considerará: $\hat{p} = \hat{q} = 0{,}50$

$$NC = 0{,}92 \qquad \alpha = 1 - NC = 1 - 0{,}92 = 0{,}08 \qquad \alpha/2 = 0{,}04 \qquad 1 - \alpha/2 = 1 - 0{,}04 = 0{,}96$$

$$n = \frac{z_{1-\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2} = \frac{z_{0{,}96}^2 \cdot 0{,}50 \cdot 0{,}50}{0{,}03^2} = \frac{1{,}751^2 \cdot 0{,}25}{0{,}0009} = 851{,}6669 \qquad n = 852$$

Respuesta: Deberían analizarse 852 compras.

Intervalo de confianza para la variancia

Se desea estimar el desvío estándar de la longitud de un lote de piezas fabricadas. Es razonable suponer que la longitud de la pieza se distribuye normalmente. Una muestra de 12 piezas del lote produjo un desvío estándar de 32 mm. Basándose en estos datos, construir un intervalo de confianza del 95% para el desvío estándar.

Solución:

Datos: n = 12 S = 32 NC = 95%

$v = n - 1 = 12 - 1 = 11$ $\alpha = 1 - NC = 1 - 0{,}95 = 0{,}05$ $\alpha/2 = 0{,}025$ $1 - \alpha/2 = 0{,}975$

$\chi^2_{v;\alpha/2} = \chi^2_{11;0{,}025} = 3{,}82$ $\chi^2_{v;1-\alpha/2} = \chi^2_{11;0{,}975} = 21{,}92$

$$Límite\ Inferior\ = LI = \frac{S^2 \cdot (n-1)}{\chi^2_{v;1-\alpha/2}} = \frac{32^2 \cdot 11}{21{,}92} = \frac{11.264}{21{,}92} = 513{,}8686$$

$$Límite\ Superior = LS = \frac{S^2 \cdot (n-1)}{\chi^2_{v;\alpha/2}} = \frac{32^2 \cdot 11}{3{,}82} = \frac{11.264}{3{,}82} = 2.948{,}6911$$

Intervalo de confianza para estimar la variancia : $[513{,}8686\,;\,2.948{,}6911]$

$\sqrt{513{,}8686} = 22{,}67$ $\sqrt{2.948{,}6911} = 54{,}30$

Intervalo de confianza para estimar el desvío estandar : $[22{,}67\,;\,54{,}30]$

Respuesta: Se estima con una confianza del 95% que el desvío estándar de la longitud de todo el lote está comprendido entre 22,67 mm y 54,30 mm.

Ejercitación:

Problema 1: En una ciudad del interior del país en la que habitan 100.000 familias se tomó una muestra al azar de 285 familias con la finalidad de analizar el ingreso mensual familiar y se obtuvo una media de $2131. Suponga que el desvío estándar de los ingresos asciende a $1772.

a) Indique cuál es la unidad de observación, la variable aleatoria en estudio, la población de referencia y la muestra. ¿$2131 es el valor de un estimador o de un parámetro? ¿Y $1772? Justifique su respuesta.
b) Estimar el ingreso promedio mensual familiar con una confianza del 90 %. ¿Los resultados se aplican a las familias encuestadas, a todas las familias de la ciudad o a todas las familias del interior?

c) Repetir la estimación pero utilizando una confianza del 99%.
d) Calcule la amplitud de ambos intervalos. ¿Es razonable que el segundo intervalo tenga una amplitud mayor que el primero?
e) ¿Cuántas familias más se debería incluir en la muestra para reducir el error de muestreo de la estimación del punto b) en un 20 %?
f) Indique los supuestos para la validez de las estimaciones efectuadas.
g) Indique cuál es el estimador utilizado y cuál es su esperanza, desvío estándar y distribución de probabilidades.

RESPUESTAS:

a) la unidad de observación es cada familia, la variable aleatoria en estudio es el ingreso mensual, la población de referencia está constituida por las familias de la ciudad del interior del país y la muestra está formada por 285 familias. $2131 es el valor de un estimador ya que se calculó sobre la muestra y $ 1772 es el valor de un parámetro, ya que al ser un dato histórico, se asume que se calculó sobre una gran cantidad de datos.
b) Se estima que el ingreso promedio mensual familiar está comprendido entre $1958 y $2303.
c) [$ 1860 ; $ 2401]
d) $ 345y $ 541
e) Se deberían incluir 161 familias más en la muestra.
f) muestreo aleatorio, distribución normal de la media, desvío poblacional conocido
g) El estimador es $\bar{x}$, su esperanza es μ, su desvío estándar (también llamado error estándar) es $\sigma/\sqrt{n}$ y distribución de probabilidades es normal ya que el tamaño de muestra es grande.

Problema 2: El dueño de un comercio minorista desea estimar el tiempo promedio que demanda la atención de cada cliente, y sabe por estudios anteriores que dicho tiempo se distribuye normalmente con desvío estándar igual a 4,215 minutos. A tal efecto, registró la cantidad de minutos que le insumió la atención de seis clientes elegidos al azar y obtuvo los siguientes datos:

15 – 12 – 8 – 23 – 15 – 11

a) Efectuar la estimación requerida con un nivel de riesgo igual al 2 %.
b) ¿Cuántos clientes más se debería observar para reducir el error de muestreo anterior en 1 minuto?
c) Basándose en la muestra original el comerciante estimó que el tiempo promedio de atención por cliente oscila entre 12,142 minutos y 15,858 minutos. ¿Cuál es el nivel de confianza de esta estimación?

RESPUESTAS: a) [10; 18 min] b) 5 clientes más c) NC = 72 %

Problema 3: Como parte de su control de calidad, la Química Erovne mide la temperatura, en °C, durante el ciclo de fabricación de un producto. Se sabe por registros históricos que la temperatura en dicho paso se distribuye normalmente con una variancia de 9°C^2.

a) Si se desea estimar la temperatura media de un ciclo de fabricación con una confianza del 95% y un máximo error muestral de 2.5°C, ¿cuántas mediciones deberán efectuarse?
b) Efectuadas las mediciones indicadas en a) se obtuvo una temperatura promedio de 92 °C. Efectúe la estimación solicitada.

RESPUESTAS: a) 6 mediciones b) [89.6°C ; 94.4°C]

Problema 4: Las anchoas en filetes se envasan a mano en cierta empresa, a fin de garantizar una presentación óptima del producto. Interesa obtener una estimación de la velocidad de llenado de las latas por los operarios, para lo cual se registró la cantidad de latas completadas por cada operario por hora. Los resultados fueron:

Velocidad (latas/hora)	Cantidad de operarios
10-20	3
20-30	11
30-40	9
40-50	2

Dado que esta estimación se efectúa periódicamente en la empresa, se conoce el desvío estándar de la velocidad de llenado, que es de 8 latas/hora.

a) Estime puntualmente la velocidad promedio de llenado manual de latas.
b) Estime la velocidad promedio de llenado manual de latas con una confianza del 90%.
c) Idem anterior, pero con una confianza del 99%. ¿Cómo afecta el aumento del nivel de confianza al error de estimación del intervalo de confianza?

RESPUESTAS: a) 29 latas/h b) [26.37 ; 31.63 latas/h] c) [24.88 ; 33.12 latas/h]

Problema 5: A fin de mejorar la programación de turnos con pacientes, cierto centro médico desea estimar el tiempo que pasan los médicos de cabecera con cada paciente en el consultorio. Con tal fin se toma una muestra aleatoria de 20 citas, con las siguientes duraciones de consulta (en min):

15 5 18 32 28 10 14 19 25 7 12 8 16 12 9 13 20 5 17 20

a) Estime puntualmente el tiempo promedio de cada consulta y su desvío estándar.
b) Estime el tiempo promedio de cada consulta con una confianza del 90%.
c) ¿Cuántas observaciones más se deberían obtener para reducir el error muestral del punto b) a la mitad?
d) Compare sus resultados con la salida de Excel generada en Herramientas > Análisis de datos > Estadística descriptiva. Para activar la opción Análisis de datos del menú Herramientas, seleccionar: Herramientas > Complementos > Herramientas para análisis

e) Indique cuál es el estimador utilizado y cuál es su esperanza, desvío estándar y distribución de probabilidades.

RESPUESTAS: a) 15.25 y 7.38 min b) [12.4 min; 18.1 min] c) 55 citas más

Estadística descriptiva	
Media	15,25
Error típico	1,65
Mediana	14,5
Moda	5
Desviación estándar	7,38
Varianza de la muestra	54,41
Curtosis	0,06
Coeficiente de asimetría	0,64
Rango	27
Mínimo	5
Máximo	32
Suma	305
Cuenta	20
Nivel de confianza(90,0%)	2,85

Problema 6: El contador de una firma comercial elige al azar 10 de las facturas emitidas en el día de ayer y encuentra los siguientes montos en pesos:

142 – 38 – 76 – 24 – 187 – 95 – 129 – 82 – 63 – 74

a) Estimar el monto promedio de las facturas emitidas ayer, con $\alpha = 0.10$, bajo el supuesto (poco realista) que el monto de las facturas se comporta normalmente.
b) Determinar el tamaño de muestra necesario para efectuar la estimación con un error de muestreo igual a $15 manteniendo el mismo nivel de confianza.
c) Estime el monto promedio máximo de las facturas emitidas ayer, con $\alpha = 0.10$. ¿Por qué no coincide con el límite superior del intervalo construido en el punto a)?
d) Explique el procedimiento que debió seguirse para extraer la muestra.

RESPUESTAS: a) [62,37 ; 119,63] b) 32 facturas c) $ 112.60 d) muestreo aleatorio

Problema 7: El consumo de bebidas alcohólicas por adolescentes constituye un problema creciente. Una comisión integrada por profesionales de la salud en cierta localidad del conurbano está interesada en conocer el nivel de consumo de alcohol en dicho grupo de riesgo. Para ello llevaron a cabo una encuesta anónima en 40 adolescentes elegidos al azar, a los cuales se interrogó sobre la cantidad y tipo de bebida que aproximadamente consumían por semana. Los resultados, convertidos en litros netos de alcohol etílico, fueron:

Consumo (litros)	Cant. Encuestados
0-0,2	12
0,2-0,4	11
0,4-0,6	8
0,6-0,8	5
0,8-1,2	4

a) Estime puntualmente el consumo semanal promedio de alcohol de los adolescentes de dicha localidad y su desvío estándar.
b) Estime mediante un intervalo de confianza del 95% el consumo semanal promedio de alcohol.
c) ¿A cuántos adolescentes más se debería encuestar si se desea efectuar la estimación con un error muestral de ± 0.06 litros?
d) El consumo semanal de alcohol no parece distribuirse normalmente. Sin embargo, la utilización de intervalos de confianza basados en la distribución normal está justificada en este estudio. ¿Por qué?
e) En realidad, los 40 adolescentes que integraron la muestra fueron seleccionados en forma aleatoria entre los alumnos de un colegio secundario de la localidad. ¿Qué cambia de sus conclusiones anteriores?

RESPUESTAS: a) 0.4 y 0.285 litros b) [0.31 ; 0.49 litros] c) 50 más
d) Por el teorema central del límite
e) la población sobre la que se efectúa la inferencia

Problema 8: La Dirección General de Estadísticas y Censos de la Ciudad de Buenos Aires informa periódicamente los resultados de la Encuesta Permanente de Hogares (EPH). Se trata de un operativo por muestreo que involucra un número importante de viviendas particulares distribuidas en el territorio de la Ciudad y está principalmente orientada a medir los niveles de ocupación y de ingresos de la población económicamente activa. En el informe correspondiente a 2010 se indica que el ingreso medio per cápita familiar mensual de la población en hogares en la Ciudad de Buenos Aires es de $ 2943. Una nota al pie indica: *excluye la población que no declara ingresos y a la población sin ingresos*. Además se informa un error de estimación de 441 $, con un nivel de confianza del 95%.

a) ¿El valor $ 2943 es un estimador o un parámetro? ¿Por qué?
b) ¿Cuál considera que es la población de este estudio? ¿Sospecha de la existencia de sesgo?
c) Calcule el intervalo de confianza para la media poblacional e interprete el resultado.
d) Explique a alguien que no sepa estadística lo que significa "nivel de confianza del 95%".

Problema 9: Una universidad privada con 11000 alumnos, desarrolla políticas activas para desalentar el consumo de tabaco. Al inicio del ciclo lectivo se implementó una encuesta a fin de determinar la incidencia actual del tabaquismo entre su alumnado. Para ello fueron seleccionados 500 alumnos al azar, de los cuales 140 se declararon fumadores.

a) Indique cuál es la unidad de observación, la variable aleatoria en estudio, la población de referencia y la muestra.
b) Estime con una confianza del 90% la proporción de fumadores entre los alumnos de la universidad. Concluya.
c) ¿Los resultados del punto anterior se aplican a los 500 alumnos seleccionados, a todos los alumnos de dicha universidad o a todos los alumnos universitarios?
d) Si se desea que el error de muestreo de la estimación no supere el 2 % y que el nivel de confianza de la misma se mantenga igual al 90%, ¿cuántos alumnos más deberían consultarse?
e) Con la misma muestra otro investigador estimó que el porcentaje de fumadores de la universidad estaba comprendido entre 26 % y 30 %. ¿Cuál fue el NC de esta estimación?
f) En otra universidad se desea realizar una investigación similar y obtener una estimación del porcentaje de alumnos fumadores con las mismas características planteadas d), pero no se cuenta con ningún dato previo acerca del valor de dicho porcentaje. ¿Cuántos estudiantes se debería consultar?
g) Indique cuál es el estimador utilizado y cuál es su esperanza, desvío estándar y distribución de probabilidades.

RESPUESTAS:

a) la unidad de observación es cada estudiante, la variable aleatoria en estudio es condición del alumno (fumador/no fumador), la población de referencia son los 11000 alumnos de la universidad y la muestra son los 500 alumnos encuestados.
b) [0,247 ; 0,313]
c) a todos los alumnos de esa universidad
d) 864 alumnos más
e) NC = 68 %
f) n = 1692 alumnos
g) El estimador es $\hat{p}$, su esperanza es p, su desvío estándar (también llamado error estándar) es $\sqrt{\frac{p.(1-p)}{n}}$ y distribución de probabilidades es normal ya que el tamaño de muestra es grande, $pn > 5$ y $(1-p)n > 5$

Problema 10: El comercio electrónico o e-commerce consiste en la distribución, compra, venta, marketing y suministro de información de productos o servicios a través de Internet. En nuestro país se encuentra en fuerte expansión, relacionado con la fuerte penetración de Internet. Una consultora llevó a cabo una encuesta entre 352 usuarios de Internet mayores de 30 años residentes en la ciudad de Buenos Aires y comprobó que sólo 39 de ellos efectuaron alguna compra por ese medio en el último año.

a) Estime puntualmente la proporción de todos los usuarios de Internet mayores de 30 años residentes en la ciudad de Buenos Aires que efectúan compras por ese medio.
b) Explíquele a alguien que no sabe estadística por qué no podemos sencillamente decir que el porcentaje de usuarios de Internet mayores de 30 años residentes en la ciudad de Buenos Aires que efectúan compras online es del 11,08%.

c) Estime la proporción de todos los usuarios de Internet mayores de 30 años residentes en la ciudad de Buenos Aires que efectúan compras online, con una confianza del 95%.
d) ¿Cuántos usuarios deberían ser encuestados si se quiere tener un error muestral máximo del 1%?

RESPUESTAS: a) 0.1108 c) [0.078 ; 0.144] d) 3785 usuarios

Problema 11: Para una investigación de mercado, se desea estimar el porcentaje de actuales compradores de cierto yogurt que comprarían una presentación del mismo producto pero de mayor contenido.

a) Si se desea estimar el porcentaje de futuros compradores en ± 10% con un riesgo del 5%, ¿cuántos consumidores deberán ser encuestados?
b) Efectuada la encuesta, 56 clientes respondieron que comprarían la nueva presentación. ¿Cuál sería el intervalo de confianza resultante?
c) ¿Por qué se utilizó $p = 0.5$ al no contarse con una estimación previa? Asigne distintos valores a p y calcule el tamaño de muestra resultante. ¿Qué observa?
d) Indique los supuestos necesarios para la validez de los cálculos efectuados.

RESPUESTAS:
a) 97 consumidores
b) [0.479 ; 0.676]
c) el tamaño de muestra alcanza un máximo cuando $p = 1\text{-}p = 0.5$
d) La muestra debe ser aleatoria. Se supone que la variable r = cantidad de compradores que adquirirán la nueva presentación sigue una distribución binomial. Para que la aproximación a la distribución normal que se utiliza en la resolución del problema sea válida se requiere un tamaño de muestra lo suficientemente grande y que $pn > 5$ y $(1\text{-}p)n > 5$.

Problema 12: Una consultora lleva a cabo un sondeo a fin de estimar la intención de voto de cierto partido político a dos semanas de las elecciones. El último sondeo arrojó un 29% de intención de voto, pero los analistas quieren confirmar si este porcentaje se mantiene.

a) ¿Cuántos votantes deberán ser encuestados si se desea estimar el porcentaje de votantes de dicho partido en ± 2% con un riesgo del 1%?
b) Suponga que la consultora tiene presupuesto para 800 encuestas. ¿Cuál será el error muestral resultante si se mantiene el nivel de confianza? ¿Y cuál será el nivel de confianza si lo que se desea es mantener el error muestral?
c) ¿Cuál de los dos intervalos anteriores es más preciso? ¿Y cuál es más exacto?

RESPUESTAS: a) 3416 votantes b) EM = 4.13% ; NC = 0.7887

c) Es más preciso el de menor EM y más exacto el de mayor NC.

Problema 13: Como parte de la política de satisfacción total del cliente, una empresa automotriz desea conocer el nivel de aceptación del modelo Theo que salió al mercado hace 6 meses y ya vendió 7300 unidades. Para ello, contacta a 523 compradores de dicho modelo (elegidos en forma aleatoria entre todos los compradores) y los somete a un cuestionario, que entre otras cosas, indaga lo siguiente:

- ¿Está satisfecho con su compra? Muy satisfecho Medianamente satisfecho Insatisfecho
- ¿Ha tenido algún desperfecto con su vehículo? SI NO

Luego del procesamiento de las respuestas se obtuvo el siguiente resumen:

¿Está satisfecho con su compra?	**Muy satisfecho**	**Medianamente Satisfecho**	**Insatisfecho**
Cant. respuestas	291	159	73

¿Ha tenido algún desperfecto con su vehículo?	**SI**	**NO**
Cant. respuestas	138	385

a) Estime con un nivel de confianza del 90% el porcentaje de todos los compradores que están satisfechos con el producto.
b) Estime con la misma confianza el porcentaje de todos los compradores que sufrieron algún desperfecto con el vehículo.
c) En otra investigación se envía un cuestionario con respuesta postal paga a los 7300 compradores del modelo y se reciben 500 respuestas. ¿Qué opina de la calidad de esta muestra? ¿Es comparable a la obtenida anteriormente? Determine el tipo de muestreo utilizado en cada caso.

RESPUESTAS: a) [83.55 ; 88.53] b) [23.22 ; 29.56]
c) Se trata de un muestreo no probabilístico, por conveniencia. En cambio, el muestreo anterior era probabilístico y se trató de un muestreo simple al azar.

Problema 14: Para controlar la precisión de una máquina que corta piezas pequeñas de acero se toma una muestra de 35 piezas y al analizar la longitud de las mismas se encuentra una variancia de 64 milímetros^2. Estimar con una confianza del 99 % el desvío estándar de las longitudes de toda la producción, suponiendo que la longitud de las piezas se distribuye normalmente.

RESPUESTA: Se estima que el desvío estándar de toda la producción está comprendido entre 6,08 mm y 11,48 mm.

Problema 15: Se desea invertir en una empresa y se sabe que las utilidades mensuales provenientes de dicha concesión se distribuyen en forma aproximadamente normal. Sin embargo, interesa la variabilidad de las utilidades, que es una medida del riesgo que se asume en el negocio. Se decide efectuar una estimación del desvío estándar máximo de las utilidades y para ello se extrae una muestra de 12 meses tomados al azar, obteniéndose un desvío estándar de U$S 14000. Efectúe la estimación solicitada con una confianza del 95%.

RESPUESTA: U$S 21709

Problema 16: Un inspector tomó una muestra de 5 latas de gaseosas, verificó el contenido en litros de las mismas y encontró: 0,48 – 0,51 – 0,45 – 0,43 – 0,52 litros

a) Estimar puntualmente el correspondiente desvío estándar poblacional.
b) Estimar el desvío estándar poblacional con NC = 90 %.
c) ¿Qué supuestos deben asumirse?

RESPUESTA: a) 0.038 litros b) [0,025 ; 0,090 litros].

c) la variable contenido de las latas de gaseosa sigue una distribución normal y la muestra es aleatoria.

Revisión conceptual

Responda las siguientes preguntas, justificando las respuestas:

a) Si se aumenta el tamaño de la muestra ¿el error muestral aumenta o disminuye?
b) Si aumenta la variabilidad de la población ¿el error muestral aumenta o disminuye? ¿Qué se podría hacer al respecto?
c) Si aumenta el riesgo de la estimación ¿el error muestral aumenta o disminuye?

Se estimó el contenido de conservantes en cierto lote de pan lactal, obteniéndose un intervalo de confianza del 95% de 0.25-0.35 (mg/100g)
Indique si las siguientes afirmaciones son verdaderas o falsas, justificando su respuesta:

a) El 95% de las muestras posee entre 0.25 y 0.35 mg/100g
b) El promedio del lote está entre 0.25 y 0.35 mg/100g
c) El promedio de la marca está entre 0.25 y 0.35 mg/100g
d) Si se aumenta la confianza el intervalo de confianza se achica (es más preciso)
e) Si se quiere achicar el intervalo de confianza se debe aumentar el tamaño de la muestra

PRACTICO 3. PRUEBAS DE HIPÓTESIS PARA UNA POBLACIÓN

Problemas resueltos:

Prueba de hipótesis para el promedio con variancia poblacional conocida:

Los aspirantes a ingresar como conductores a una importante línea de autobuses deben someterse a una serie de controles entre los que se cuenta una evaluación de los reflejos, que consiste en presentar sorpresivamente 4 obstáculos al aspirante mientras maneja y medir el tiempo que tarda en reaccionar ante los mismos. La empresa considera que el tiempo de reacción promedio ante un obstáculo imprevisto no debe superar los 0,48 segundos. Se sabe que el tiempo de reacción se distribuye normalmente con un desvío estándar de 0,04 segundos, y la decisión de dar por aprobada o considerar desaprobada la evaluación se toma con un nivel de significación del 10%.

a) Indicar las hipótesis adecuadas, la condición de rechazo y la regla de decisión.
b) Un aspirante registró los siguientes tiempos de reacción en segundos: 0,48 – 0,52 – 0,59 – 0,46. ¿Se dará por aprobada esta evaluación? ¿Por qué?
c) Calcular la probabilidad de aprobar a un aspirante cuyo verdadero tiempo de reacción promedio es igual a 0,525 segundos.
d) ¿Cuántos obstáculos más debería incluirse en la evaluación para que la probabilidad calculada en el punto anterior valga 0,05?

Solución:

a) Datos: n = 4 σ = 0,04 segundos μ_0 = 0,48 segundos α = 10%

Para aprobar, el promedio no debe superar los 0,48 seg. ($\mu \leq 0,48$)

$H_0) \ \mu \leq 0,48$ (*se aprueba la evaluación*)

$H_1) \ \mu > 0,48$ (*no se aprueba la evaluación*)

$$\bar{x}_c = \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n} = 0,48 + z_{0,90} \cdot 0,04/\sqrt{4} = 0,48 + 1,282 \cdot \frac{0,04}{2} = 0,48 + 0,02564 = 0,50564$$

Condición de rechazo : *Si* $\bar{x} > \bar{x}_c$ *se rechaza la hipótesis nula.*

(CR : *Si* $\bar{x} > 0,50564$ *se rechaza* H_0)

Re*gla de decisión* : RD : *Si se rechaza* H_0, *no se aprueba la evaluación*.

b) x: 0,48 – 0,52 – 0,59 – 0,46

$$\bar{x} = \frac{\sum x}{n} = \frac{0{,}48 + 0{,}52 + 0{,}59 + 0{,}46}{4} = \frac{2{,}05}{4} = 0{,}5125$$

Conclusión : *Como* $\bar{x} > \bar{x}_c$ $(0{,}5125 > 0{,}50564)$ *se rechaza* H_0, *por lo tanto*, *el aspirante no será aprobado.*

c) $\mu_1 = 0{,}525$ $\beta = ?$

P(aprobar si $\mu = 0{,}525$) = P(No rechazar H_0 siendo falsa) = β

$$\beta = P_N\left(\bar{x} \le 0{,}50564 \,/\, \mu_1 = 0{,}525 \,;\, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}04}{\sqrt{4}} = 0{,}02\right)$$

$$\beta = P_N\left(z \le \frac{0{,}50564 - 0{,}5250}{0{,}02}\right) = P_N\left(z \le -0{,}97\right) = F(-0{,}97) = 0{,}16602$$

Respuesta: La probabilidad de aprobar con μ=0,525 segundos es igual a 0,166

d) $\mu_1 = 0{,}525$ $\beta = 0{,}05$ $1-\beta = 0{,}95$ $\mu_0 = 0{,}48$ $\alpha = 0{,}10$ $1-\alpha = 0{,}90$

$$n = \left[\frac{\left(z_{1-\alpha} + z_{1-\beta}\right)\cdot\sigma}{\mu_0 - \mu_1}\right]^2 = \left[\frac{\left(z_{0{,}90} + z_{0{,}95}\right)\cdot 0{,}04}{0{,}48 - 0{,}525}\right]^2 = \left[\frac{(1{,}282 + 1{,}645)\cdot 0.04}{-0{,}045}\right]^2 = 6{,}77 \qquad n = 7$$

Respuesta: Se deberían incluir 3 obstáculos más.

Prueba de hipótesis para el promedio con variancia poblacional desconocida:

El diámetro interior de los cilindros producidos por una máquina se distribuye normalmente y su promedio debe resultar igual a 1,32 cm. Para controlar la calidad de la producción y detenerla en caso de detectar que los cilindros no cumplen la especificación requerida, se revisa una muestra de 16 cilindros que arroja un diámetro promedio de 1,315 cm. y un desvío estándar igual a 0,02 cm. Ante estos resultados, y fijando en 0,10 la probabilidad de detener equivocadamente la producción, ¿continuaría usted el proceso productivo o lo detendría?

Solución:

Datos: $\mu_0 = 1{,}32$ $n = 16$ $\bar{x} = 1{,}315$ $S = 0{,}02$ $\alpha = 0{,}10$

Se detiene la producción si no se cumple la especificación de que el diámetro promedio resulte igual a 1,32 cm. (Es decir, se detiene si: $\mu \neq 1{,}32$).

$H_0)\ \mu = 1{,}32$ (*no se detiene el proceso productivo*)

$H_1)\ \mu \neq 1{,}32$ (*se detiene el proceso productivo*)

$$\nu = n-1 = 16-1 = 15 \qquad \phi = 1-\alpha/2 = 1-0{,}10/2 = 0{,}95 \qquad t_{\nu;\phi} = t_{15;0{,}95} = 1{,}753$$

$$\bar{x}_{c1} = \mu_0 - t_{\nu;\phi} \cdot S/\sqrt{n} = 1.32 - 1{,}753 \cdot \frac{0{,}02}{\sqrt{16}} = 1{,}32 - 0{,}008765 = 1{,}311235$$

$$\bar{x}_{c2} = \mu_0 + t_{\nu;\phi} \cdot S/\sqrt{n} = 1{,}32 + 1{,}753 \cdot \frac{0{,}02}{\sqrt{16}} = 1{,}32 + 0{,}008765 = 1{,}328765$$

CR: *Si* $\bar{x} < 1{,}311235$ *o* $\bar{x} > 1{,}328765$ *se rechaza* H_0.

RD: *Si se rechaza* H_0, *entonces se* det*iene el proceso productivo.*

Conclusión:
Como $\bar{x}$ (*que resultó igual a* 1,315) *no es menor a* $\bar{x}_{c1}$(1,311235) *ni es mayor a* $\bar{x}_{c2}$(1,328765) *no se rechaza la hipótesis nula. En consecuencia, no se detendrá el proceso productivo.*

Respuesta: Continuaría el proceso productivo.

Prueba de hipótesis para la proporción:

Como parte de una política orientada a mejorar las condiciones de trabajo, una fábrica de grandes dimensiones instaló el año pasado un comedor para sus operarios y otorgó la explotación del mismo en forma transitoria, por el lapso de un año, a una conocida empresa del rubro. Una vez transcurrido dicho período se encuestará a un grupo de operarios elegidos al azar para conocer la opinión del personal acerca de la calidad del servicio recibido, y se decidirá renovar la concesión de la explotación del comedor a la misma empresa en caso de detectar que más del 70% de los operarios se muestra satisfecho con dicho servicio. En caso contrario, se tomará la decisión de cambiar el concesionario.

a) Determinar la cantidad de operarios que serán consultados y la regla de decisión que se adoptará, sabiendo que se desea que valga 0,05 la probabilidad de decidir equivocadamente la renovación de la concesión; y que se ha fijado en 0,15 la probabilidad de decidir el cambio de concesionario en caso de que el verdadero porcentaje de operarios satisfechos con el servicio de comedor ascendiera al 83%.
b) Adoptando esta regla de decisión, ¿cuál es la probabilidad de renovar la concesión si las tres cuartas partes del personal está satisfecho con el servicio de comedor brindado por el actual concesionario?
c) Cumplido el año de plazo, se toma la muestra prevista y se comprueba que el 82% de los operarios consultados expresó su satisfacción con el actual concesionario del comedor. ¿Qué decisión se adoptará?

d) Estimar, con una confianza del 90%, el porcentaje de operarios satisfechos con el actual servicio de comedor en toda la fábrica.

Solución:

a) Datos: La concesión se renovará si más del 70% ($p_0 = 0,7$) de los operarios están satisfechos con el servicio actual (se renueva si: $p > 0,70$).

$H_0)\ p \le 0,70$ (*no se renueva la concesión*)

$H_1)\ p > 0,70$ (*se renueva la concesión*)

P (decidir la renovación de la concesión equivocadamente) = 0,05
P (Rechazar H_0 / H_0 es cierta) = 0,05 = α $\quad 1 - \alpha = 0,95$

P (decidir el cambio de concesionario si en realidad $p_1 = 0,83$) = 0,15
P (No se rechaza H_0 / H_0 es falsa) = 0,15 = β $\quad 1 - \beta = 0,85$

$$n = \left[\frac{z_{1-\alpha} \cdot \sqrt{p_0 \cdot q_0} + z_{1-\beta} \cdot \sqrt{p_1 \cdot q_1}}{p_1 - p_0} \right]^2 = \left[\frac{z_{0,95} \cdot \sqrt{0,70 \cdot 0,30} + z_{0,85} \cdot \sqrt{0,83 \cdot 0,17}}{0,83 - 0,70} \right]^2$$

$$n = \left[\frac{1,645 \cdot \sqrt{0,21} + 1,036 \cdot \sqrt{0,1411}}{0,13} \right]^2 = \left[\frac{0,7538 + 0,38916}{0,13} \right]^2 = 77,3 \qquad n = 78$$

$$\hat{p}_c = p_0 + z_\phi \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} = 0,70 + z_{0,95} \cdot \sqrt{\frac{0,70 \cdot 0,30}{78}} = 0,70 + 1,645 \cdot 0,0519 = 0,70 + 0,085 = 0,785$$

CR : Si $\hat{p} > 0,785$ *se rechaza la hipótesis nula.*
RD : Si se rechaza H_0*, se renueva la concesión.*

Respuesta: Se consultarán 78 operarios y se renovará la concesión si el porcentaje de operarios satisfechos en esta muestra resulta mayor a 78,5 % (es decir, si como mínimo 62 de los consultados se muestran satisfechos).

b) P (renovar la concesión si las ¾ partes de los operarios están satisfechos)
P (Rechazar H_0 / $p_1 = 0,75$) = P (Rechazar H_0 / H_0 es falsa) = $1 - \beta$

$$P_N\left(\hat{p} > 0,785 \,/\, \mu = p_1 = 0,75 \,;\, \sigma = \sqrt{\frac{p_1 \cdot q_1}{n}} = \sqrt{\frac{0,75 \cdot 0,25}{78}} = 0,049 \right) = P_N\left(z > \frac{0,785 - 0,75}{0,049} \right)$$

$$P_N(z > 0,71) = 1 - F(0,71) = 1 - 0,76115 = 0,23885$$

Respuesta: La probabilidad pedida es igual a 0,23885

c) Datos: n = 78 $\hat{p} = 0{,}82$

Respuesta : Como $\hat{p}$ *resultó mayor que* $\hat{p}_c$ $(0{,}82 > 0{,}785)$, *se rechaza* H_0.
Por lo tanto, se adoptará la decisión de renovar la concesión del comedor.

d) Datos: n = 78 $\hat{p} = 0{,}82$ NC = 90%

$\hat{q} = 1 - \hat{p} = 1 - 0{,}82 = 0{,}18$ $\alpha = 1 - NC = 1 - 0{,}90 = 0{,}10$ $1 - \alpha/2 = 1 - 0{,}05 = 0{,}95$

$$E = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = z_{0{,}95} \cdot \sqrt{\frac{0{,}82 \cdot 0{,}18}{78}} = 1{,}645 \cdot 0{,}0435 = 0{,}07 \qquad \hat{p} \pm E = 0{,}82 \pm 0{,}07$$

Respuesta: Se estima que el porcentaje de operarios satisfechos con el servicio, en toda la fábrica, está comprendido entre 75% y 89%.

Prueba de hipótesis para la variancia:

El gerente de una importante empresa de servicios desea mejorar la atención a los clientes en lo que respecta al tiempo que les insume completar los distintos trámites que deben efectuar, pues si bien el promedio de dicho tiempo es relativamente bajo, su variabilidad es muy alta, tornando imprevisible la cantidad de minutos que demandarán. Un asesor, consultado al respecto, aconseja la implementación de un sistema de turnos rotativos y capacitación del personal que, en su opinión, redundará en una reducción significativa de la variabilidad del tiempo de espera y atención que se distribuye normalmente, y cuyo desvío estándar en la actualidad es igual a 73 minutos.
Después de aplicar durante dos meses el sistema propuesto por el asesor, se selecciona una muestra de 12 trámites que arroja un tiempo promedio de 40 minutos y un desvío estándar de 32 minutos. ¿Se puede considerar que el nuevo sistema logró su objetivo? (Utilizar α = 5%).

Solución:

Datos: $\sigma_0 = 73$ n = 12 S = 32 α= 5%

Para lograr su objetivo el nuevo sistema debe redundar en una reducción significativa de la variabilidad del tiempo de espera y atención. (Es decir, que se considerará que logró su objetivo si $\sigma^2 < 73^2$).

$H_0) \ \sigma^2 \geq 73^2$ (*el nuevo sistema no logró su objetivo*)

$H_1) \ \sigma^2 < 73^2$ (*el nuevo sistema logró su objetivo*)

$v = n - 1 = 12 - 1 = 11 \qquad \phi = \alpha = 0{,}05 \qquad \chi^2_{crít.} = \chi^2_{v;\phi} = \chi^2_{11;0{,}05} = 4{,}57$

CR: *Si* $\chi^2_{calc} < \chi^2_{crít.}$ *se rechaza* H_0. (*Si* $\chi^2_{calc.} < 4{,}57$ *se rechaza la hipótesis nula*).

RD: *Si se rechaza* H_0, *se considera que el nuevo sistema logró su objetivo*.

$$\chi^2_{calc.} = \frac{S^2 \cdot (n-1)}{\sigma^2} = \frac{32^2 \cdot (12-1)}{73^2} = \frac{1.024 \cdot 11}{5.329} = 2{,}1137$$

Conclusión: *Como* $\chi^2_{calc.} < \chi^2_{crít.}$ (2,1137 < 4,57), *se rechaza* H_0. *Por lo tanto se considera que el nuevo sistema logró su objetivo*.

Ejercitación:

Problema 1: El dueño de una casa de comidas rápidas con entrega domiciliaria, decide controlar el rendimiento de sus empleados y se comunica telefónicamente con 14 clientes para verificar el tiempo de entrega de los pedidos y encuentra un promedio de 32 minutos. El encargado le había asegurado que en promedio de entrega de pedidos era de media hora como máximo. Sabiendo que el tiempo de entrega se distribuye normalmente con desvío estándar igual a 8 minutos, y utilizando α = 10 %,

a) ¿Considera usted que la afirmación del encargado es incorrecta?
b) ¿Cómo explicaría el significado del valor de α?

RESPUESTAS:

a) No hay pruebas de que la afirmación del encargado sea incorrecta (32 es < a 32.74 min)
b) La máxima probabilidad de considerar incorrecta la afirmación del encargado, cuando en realidad es correcta, es igual a 0,10.

Problema 2: Una importante firma dedicada a la comercialización de artículos médicos compra mensualmente grandes partidas de jeringas descartables a un proveedor local que las entrega en cajas. Cada partida contiene 3000 cajas[1] se ha pactado que el contenido promedio de la misma debe ser de 120 jeringas/caja. Para evitar la compra de cajas con

[1] El conocimiento del tamaño de la población implicaría en los procedimientos utilizados una corrección sobre el error de muestreo dado por $\sqrt{\frac{(N-n)}{(N-1)}}$, denominado, factor de corrección por finitud. En este caso, su valor es prácticamente 1, con lo cual su consideración no modificaría las conclusiones a las cuales se arriban sin su empleo. Todos los ejercicios de la guía que presentan el tamaño de la población como dato, tienen este mismo resultado, motivo por el cual no lo emplearemos. Por otra parte, conceptualmente se excede al campo de cobertura de ésta materia

contenido inferior al pactado, se efectúa un control de recepción revisando una muestra de 42 cajas elegidas al azar. Suponga que el desvío estándar es igual a 6 jeringas y que la probabilidad de rechazar equivocadamente una partida se fija en 5%.

a) Establecer la condición de rechazo y la regla de decisión.
b) Se recibe una partida de la que se extraen al azar 42 cajas, siendo el contenido promedio de las mismas de 117 jeringas. ¿Qué decisión se tomará acerca de la compra de esta partida?
c) Indicar cuál es el riesgo del comprador y cuál el del vendedor.
d) Calcular la probabilidad de rechazar el pedido cuando la partida tenga un promedio de 118 jeringas por caja.
e) ¿Qué cantidad de cajas deberán ser revisadas si se desea que la probabilidad calculada en d) valga 0.9?
f) Discuta cómo se modificaría la región de rechazo y la potencia de la prueba si: 1) aumenta la variabilidad en el contenido de jeringas por caja; 2) aumenta la cantidad de cajas revisadas en el control; 3) aumenta el riesgo que está dispuesto a cometer el proveedor; 4) se recibe un lote muy malo.

RESPUESTAS: a) CR: Si $\bar{x} < 118{,}5$ se rechaza H_0. RD: Si se rechaza H_0 no se compra la partida.

b) Se rechaza la partida.

c) Riesgo del comprador: comprar una partida a pesar de que su contenido promedio es menor a 120 jeringas; la probabilidad de que esto ocurra es ß. Riesgo del proveedor: que le rechacen una partida a pesar de que su contenido promedio es de 120 jeringas o más; la máxima probabilidad de que esto ocurra es α.

d) 0.697

e) 78 cajas

Problema 3: En un laboratorio se controla periódicamente la calidad de los productos elaborados examinando muestras al azar de la producción y deteniendo el proceso de elaboración en caso de detectar alguna anomalía. A tal efecto se ha analizado una muestra de 20 comprimidos cuyo contenido promedio de calcio resultó igual a 247 mg. Por otro lado se sabe que el contenido promedio de calcio por comprimido debe ser igual a 250 mg y que el desvío estándar es de 2 mg.

a) ¿Debería detenerse el proceso de elaboración de estos comprimidos? (utilizar un nivel de significación del 1 %).
b) ¿En qué consisten el error de tipo I y de tipo II en este problema?
c) ¿Cuál es la probabilidad de detener el proceso cuando el contenido promedio de los comprimidos se incrementa en un 1%?
d) ¿Qué pasaría con los valores críticos si se aumentase el tamaño de la muestra?

RESPUESTAS: a) Sí, debería detenerse el proceso de elaboración porque el contenido promedio de calcio es inferior a 248.84 mg.

b) Error de tipo I: detener el proceso cuando este funciona correctamente; Error de tipo II: no detener el proceso cuando este funciona mal.

c) 0.9987

Problema 4: Una pequeña empresa disminuye sus gastos administrativos encargando la confección de su documentación a un centro de tipeo que trabaja con un promedio de 1,8 errores por página con un desvío igual a 0,6. Otro centro similar le ofrece sus servicios a un costo un poco más alto, pero le asegura que la calidad de su trabajo es muy superior, es decir que trabajan con un promedio de errores por hoja significativamente menor. Con el fin de tomar una decisión al respecto, la empresa encarga al segundo centro la confección de 50 páginas y al revisarlas se verifica un total de 85 errores. Considere que el desvío estándar no varía y fije en 0,05 la probabilidad de decidir equivocadamente el cambio de centro.

a) Establecer la condición de rechazo y la regla de decisión.
b) ¿Qué decisión se tomará acerca del cambio de centro? ¿Qué tipo de error se puede estar cometiendo? ¿Con qué probabilidad?

RESPUESTAS:

a) CR: Si $\bar{x} < 1{,}66$ errores por página se rechaza H_0. RD: Si se rechaza H_0 se cambia de centro.
b) Se continuará con el centro actual (1.7 no es menor que 1.66). Se puede estar cometiendo un error de tipo II, con probabilidad desconocida.

Problema 5: En una fábrica se producen pilas cuya vida útil promedio es de 78 horas con una variancia igual a 49 horas^2. Un ingeniero propone al dueño de la fábrica la adopción de un nuevo método de producción cuya implementación resultaría bastante costosa, pero si se comprobara que la duración de las pilas realmente se incrementa, el dueño estaría dispuesto a adoptarlo. La comprobación consistió en tomar una muestra de 28 pilas fabricadas con el nuevo método, y al hacerlo se observó una duración promedio de 82 horas.

a) ¿Con un nivel de significación del 2 % aconsejaría cambiar el método de producción?
b) ¿Qué tipo de error se puede estar cometiendo? ¿Con qué probabilidad máxima?
c) Indique lo supuestos requeridos para la validez de la prueba.
d) Estime la duración promedio de las pilas fabricadas con el nuevo método con una confianza del 95%.

RESPUESTA: a) Sí, porque el promedio de la muestra es mayor a 80,72 hs.
b) Se puede estar cometiendo un error de tipo I, es decir cambiar de método de producción cuando en realidad la duración de las pilas no aumentó.
c) La variable debe seguir una distribución normal y la muestra debe ser tomada al azar.
d) [79.4 ; 84.6 hs]

Problema 6: El diámetro promedio de ciertas piezas producidas automáticamente debe ser igual a 3 mm para que el proceso de producción se considere bajo control. El ingeniero industrial a cargo de la producción desea establecer un control rutinario consistente en la extracción de una muestra cada hora y si se detectara que el proceso de producción no está bajo control, detenerlo y revisar. El ingeniero establece que la probabilidad de detener innecesariamente la producción debe ser igual a 0,05 y que la probabilidad de detectar que

el proceso está fuera de control cuando el diámetro promedio sea de 3,5 mm debe ser igual a 0,90. Por registros históricos se conoce que el desvío estándar de los diámetros analizados es igual a 0,84 mm.

a) Determinar el tamaño de la muestra.
b) Si se implementa el control y una de las muestras arroja un diámetro promedio de 2,89 mm. ¿detendría usted el proceso de producción? ¿Por qué?

RESPUESTAS: a) n = 30 piezas

b) No, porque el promedio de la muestra está comprendido entre 2,7 mm y 3,3 mm.

Problema 7: La fórmula del latex utilizado para guantes de uso en cirugía es exclusividad de cada fabricante. Uno de los fabricantes estudia la posibilidad de cambiar la fórmula actual por otra más costosa, siempre que pueda asegurar que el promedio de duración sea superior al de la fórmula actual, que es de 110 días. Se fija en un 5% la probabilidad de cambiar equivocadamente la fórmula actual por la nueva fórmula y en un 10 % probabilidad de no cambiar la fórmula cuando el promedio de duración con la nueva fórmula es de 126 días. Por otro lado, el desvío estándar del tiempo de duración es de 25 días y se piensa que no se modificará con la nueva fórmula.

a) Indique las hipótesis apropiadas a esta situación, el tamaño de muestra necesario, la condición de rechazo y la regla de decisión.
b) Calcule la probabilidad de efectuar el cambio si el promedio de duración con la nueva fórmula es de 122 días.
c) Si en la muestra se obtuvo una duración promedio de 130 días, ¿qué decisión se debería tomar?
d) Estime con una confianza del 90% el promedio máximo de duración de los nuevos guantes.

RESPUESTAS:

a) n = 21 guantes; CR: Si $\bar{x} > 119$ días se rechaza Ho. RD: Si se rechaza Ho se cambia a la fórmula más costosa.

b) 0.71048
c) Se debería cambiar a la fórmula más costosa.
d) 137 días

Problema 8: Una fábrica dedicada a la producción en serie de cierto tipo de pieza tiene un tiempo de manufactura que se distribuye normalmente con un promedio de 6,2 minutos y un desvío estándar de 0,7 minutos. Se considera la posibilidad de incorporar una nueva máquina recientemente lanzada al mercado ya que se piensa que se pueden disminuir los tiempos de producción. Teniendo en cuenta el costo de la nueva máquina, solo se la adquirirá si se tiene una razonable seguridad de que con ella se logrará una disminución del tiempo medio actual. Si el promedio es igual al actual se fija en 5% la probabilidad de comprar la nueva máquina. En cambio, si el promedio es inferior en un 10% se desea que la probabilidad de compra valga 99%.

a) Indique que cantidad de ensayos deberían efectuarse con la nueva máquina y qué resultado debería observarse para decir comprarla.
b) Calcule la probabilidad de comprar la nueva máquina cuando el tiempo medio de producción es un 5% inferior al actual.
c) Si en la muestra se obtuvo un promedio de 5 minutos, ¿qué decidiría Ud?

RESPUESTAS: a) Deberían efectuarse 21 ensayos y si el tiempo promedio de producción de los mismos es inferior a 5.95 min, se aconsejará su compra.
b) 0.65
c) Aconsejaría la compra de la nueva máquina.

Problema 9: La evaluación sensorial constituye una de las herramientas fundamentales del proceso de aseguramiento de la calidad de los alimentos. Se lleva a cabo un trabajo que tiene como finalidad medir la aceptabilidad de una nueva formulación de pasta de aceituna, utilizando consumidores habituales de este producto. Para ello, se efectúa una prueba de sabor con la nueva formulación, en la que una muestra de consumidores valora el agrado general en una escala estructurada de 0 a 100 (desagrado-agrado). Se decide que sólo si se encuentran pruebas de que los consumidores de pasta de aceituna valorarán la nueva formulación con un puntaje superior a 80 puntos en promedio, ésta será comercializada. Participaron de la experiencia 25 consumidores. Los resultados obtenidos luego del procesamiento de las respuestas en Excel fueron:

Estadística descriptiva	
Media	85,25
Mediana	84,11
Desviación estándar	17,38
Varianza de la muestra	302,06
Error típico	60,41
Rango	45
Mínimo	55
Máximo	100
Cuenta	25

a) ¿Qué decisión debería tomarse con estos resultados con respecto a la comercialización de la nueva formulación? Se fijó el nivel de significación en 5%.
b) El gerente de ventas afirma que el producto debe ser comercializado, ya que el puntaje promedio obtenido fue superior a 80. ¿Qué le respondería Ud?

RESPUESTAS: a) no se aconseja la comercialización de la nueva formulación (85,25 no es mayor a 85,95).
b) El requerimiento para lanzar el producto hace referencia a un promedio poblacional, mientras que el puntaje promedio al que hace referencia el gerente es muestral. Como sabemos, éste último es variable y por lo tanto, no es comparable al promedio poblacional.

Problema 10: Una compañía cerealera de transportes desea investigar si el costo medio de mantenimiento de los camiones es inferior a $ 2500 mensuales cuando se utiliza un aceite especial más caro que el actual. Se experimentó el aceite especial con 30 camiones durante un mes, obteniéndose un costo medio de $ 2420 con un desvío estándar de $ 645. Se establece en un 5% la probabilidad de cometer error de tipo I.

a) ¿Recomendaría Ud. el cambio?
b) Interprete α y β en términos del problema. ¿Cuál de los dos podría estar cometiendo según su respuesta al punto anterior?
c) Indique si se cumplen los supuestos de la prueba estadística efectuada.

RESPUESTAS: a) No, porque el costo medio con el nuevo aceite no es inferior a 2300$

b) α = máxima probabilidad de afirmar que el nuevo aceite disminuyó el costo medio de mantenimiento de los camiones cuando en realidad no lo hizo. β = probabilidad de afirmar que el nuevo aceite no disminuyó el costo medio de mantenimiento de los camiones cuando en realidad sí lo hizo.
c) Los 30 camiones deben haber sido elegidos al azar entre el plantel de camiones de la compañía. Además, la variable costo mensual debe ajustar a una distribución normal, lo cual es poco probable. Sin embargo, como el tamaño de la muestra es relativamente grande, por el teorema central del límite podemos asumir una distribución aproximadamente normal para la media muestral.

Problema 11: La ley 26.687 sancionada en 2011 regula la publicidad, promoción y consumo de productos elaborados con tabaco. Entre otros considerandos, establece un máximo de once miligramos (11 mg) de alquitrán por cigarrillo, en promedio, debido a sus comprobados efectos nocivos sobre la salud. El Ministerio de Salud controla una nueva marca de cigarrillos, para lo cual mide el contenido de alquitrán, con los siguientes resultados (en mg por cigarrillo):

11,3	13,3	13,1	12,3	12,5	12,0	11,8	11,0	12,4	12,2	12,7	11,1
10,8	12,6	10,2	9,7	10,8	10,9	11,2	12,3	14,3	11,8	10,5	11,1

a) Utilizando un nivel de significación del 1%, ¿existe evidencia de que la nueva marca supera los niveles máximos permitidos de alquitrán en sus cigarrillos?
b) Estime con una confianza del 90 % el contenido medio de alquitrán de la nueva marca de cigarrillos.
c) ¿Cuántos cigarrillos más deberán analizarse si se desea disminuir el error muestral en un 25%?

RESPUESTAS: a) Sí existen evidencias porque 11,73 es superior a 11,55 mg.
b) [11,37; 12,13 mg]
c) 18 cigarrillos más

Problema 12: El dengue se transmite por un mosquito presente en aguas estancadas. Se lleva a cabo un estudio sanitario en varias localidades del GBA y se declarará emergencia

sanitaria si en más del 40% de los depósitos de agua en lugares públicos están presentes larvas de dicho mosquito. Se toman 300 muestras de agua al azar de dichas localidades, observándose que 165 están libres del mosquito. Se establece en un 10% el riesgo de declarar erróneamente la emergencia sanitaria.

a) ¿Considera que existen suficientes evidencias como para declarar la emergencia sanitaria?
b) Estime el porcentaje de cuerpos de agua contaminados en dichas localidades con una confianza del 90%.

RESPUESTAS: a) Sí, porque el porcentaje de muestras contaminadas es superior al 43.6%

b) [40.3 ; 49.7%]

Problema 13: Una empresa que se dedica a la fabricación de insecticidas en aerosol, recibe reclamos en el 10% de sus unidades debido a fallas en el sistema de spray. Se está considerando cambiar el sistema de aerosol actual por otro más costoso pero más efectivo. A fin de tomar una decisión, se fabrican 500 unidades con el nuevo sistema, encontrándose fallas en 38 de ellas.

a) Con un riesgo del 5% de tomar una decisión incorrecta, ¿aconsejaría cambiar al sistema más costoso?
b) Calcule la probabilidad de aconsejar el cambio al sistema más costoso si con él se lograse un 7% de unidades defectuosas. ¿Cómo se denomina esta probabilidad? ¿Es una decisión correcta?
c) ¿Cuántas unidades más deberían haber sido fabricadas y probadas si se desea que la probabilidad calculada en el punto anterior valga 0.9?
d) Interprete α y β en términos del problema.

RESPUESTAS: a) Sí, porque $\hat{p}$ es inferior a 0.078

b) 1-β = 0.757 c) 248 unidades más

d) α es la máxima probabilidad de decidir cambiar al sistema más costoso cuando éste en realidad no es más efectivo. β es la probabilidad de decidir no cambiar al sistema más costoso cuando éste en realidad es más efectivo que el sistema actual.

Problema 14: Una población infantil se dice que es susceptible de recibir una campaña de educación e higiene si su porcentaje de niños con dientes cariados es superior al 15%. En una población de 12637 niños, un analista evalúa si debe efectuarse la campaña, siendo que de 387 niños elegidos al azar, 65 tenían algún diente cariado.

a) Utilice α= 0.1 y determine qué debería concluir el analista. ¿Qué error puede estar cometiendo con la decisión tomada?
b) El analista se pregunta si, en caso de que el porcentaje de niños con caries en la población fuese grande, de 20% por ejemplo, qué probabilidad tendría de detectarlo con esta prueba.

RESPUESTAS:

a) No se aconseja iniciar la campaña ya que no existen evidencias de que el porcentaje de niños con caries sea superior al 15% (0.168 no es mayor que 0.173). Dado que no se rechazó Ho, el analista puede estar cometiendo un error de tipo II.
b) 1-ß = 0,9047

Problema 15: En un control efectuado en una fábrica de mermeladas, se extrae una muestra al azar de 120 envases de un lote y se observa que 14 presentan un incorrecto pegado de las etiquetas. Las especificaciones de calidad admiten como máximo un 9% de envases incorrectamente etiquetados por lote y se fija en un 10% la probabilidad concluir erróneamente que se supera dicho valor.

a) Con la información de la muestra ¿qué podría concluir? ¿Qué error podría estar cometiendo?
b) ¿Cuál es la probabilidad de no detectar que en el lote el porcentaje de envases incorrectamente etiquetados es del 13%? ¿Cuáles son las consecuencias de cometer este error?
c) ¿Cuántos envases deberían revisarse si se desea que la probabilidad calculada en el punto anterior valga la mitad?

RESPUESTAS: a) Se concluye que no existen evidencias para asegurar que el porcentaje de envases incorrectamente sellados está por encima del máximo admitido (0.117 < 0.124). Se podría estar cometiendo un error tipo II.
b) 0.416; producir productos fuera de la especificación c) 257 envases

Problema 16: Una nueva serie de TV en horario central debe demostrar que capta más del 20 % de la audiencia después de su período inicial de 4 semanas para decir que tuvo éxito. Después de un mes del lanzamiento de una nueva serie, de una muestra aleatoria de 400 hogares, contactados telefónicamente, se encontró que 104 estaban viéndola.

a) Utilizando un nivel de significación del 10 % y con la información de la muestra, ¿puede Ud. concluir que la serie tuvo éxito?
b) ¿Cuál es la probabilidad, con una muestra de 400 hogares, de concluir que la serie no tuvo éxito si en realidad el 24 % de los hogares viese la serie?
c) Si se desea que la probabilidad de decir que la serie tuvo éxito cuando el 24 % de las familias ven la serie valga 95%, ¿cuantas familias más deberían incluirse en la muestra?

RESPUESTAS: a) Sí, porque $\hat{p}$ es superior a 0.226 b) 0.251 c) 523 familias más

Problema 17: En cierto establecimiento avícola, los huevos son lavados previo a su envasado en cajas de cartón. Durante este procedimiento las pérdidas estimadas por rotura son del 13 %. Con el fin de mejorar el rendimiento, se decide implementar un nuevo proceso de lavado, levemente más costoso, fijándose en un 5 % la probabilidad de tomar una decisión equivocada. Por otro lado se desea una probabilidad del 70% de implementar el nuevo proceso de lavado si con él se reduce el porcentaje de rotura al 10%.

a) Calcule el tamaño de muestra adecuado y la cantidad de huevos rotos que se deben encontrar como máximo en dicha muestra para implementar el nuevo proceso.
b) ¿Cuál es la probabilidad de no implementar el nuevo sistema si con el nuevo método se logra reducir el porcentaje de rotura en 2 puntos?
c) Si en la muestra se observó que 50 huevos resultaron rotos, ¿recomendaría la implementación del nuevo sistema?
d) Estime el porcentaje de huevos sanos con el nuevo sistema con una confianza del 90%.

RESPUESTAS: a) Se deberán revisar 561 huevos y a lo sumo 61 de los mismos deberán estar rotos para implementar el nuevo método de lavado.
b) 0.60 c) Sí (0.089 < 0.107) d) [89.1 ; 93.1%]

Problema 18: Una empresa manufacturera que cuenta con un gran plantel de operarios distribuidos en tres turnos de trabajo ha contratado los servicios de un experto en relaciones laborales quien, como parte de su asesoramiento, recomienda que los operarios escuchen música mientras trabajan, generando así condiciones laborales más agradables. Esta reforma redundará, según su opinión, en un aumento de la productividad. El dueño de la empresa decide implementar esta reforma si le demuestran que más de la mitad de los operarios está de acuerdo. El asesor propone entonces, tomar una muestra de operarios y consultar su opinión acerca de escuchar música mientras trabajan.

a) Determinar el tamaño de la muestra necesario para que la probabilidad de decidir equivocadamente la implementación de la reforma propuesta por el asesor valga 0,05 y que la probabilidad de no implementarla, cuando en realidad el 62 % de los operarios está de acuerdo, resulte igual a 0,10.
b) Calcular la probabilidad de detectar, mediante este test, la conveniencia de implementar dicha reforma si el 55% de los operarios estuviera de acuerdo con la misma.
c) Una vez tomada la muestra, se encontró que 87 operarios se manifestaron de acuerdo con la reforma propuesta. ¿Aconsejaría usted implementarla?
d) Estimar con NC = 90 % el porcentaje de operarios de la empresa que está de acuerdo con escuchar música mientras trabajan.

RESPUESTAS: a) n = 145 operarios b) 0,33 c) Sí (porque 0,60 > 0,5683) d) [53,3% ; 66,7%]

Problema 19: Una empresa de productos alimenticios ha lanzado una campaña publicitaria sobre una mayonesa cuya penetración en el mercado era del 15 %. Se desea realizar un relevamiento a fin de determinar si la campaña ha sido efectiva, en cuyo caso se realizará una campaña similar para otro de los productos de la empresa. Si no se ha conseguido el resultado deseado, se establece en un 5% la probabilidad de tomar una decisión errónea y en cambio un aumento en la penetración de 3 puntos se considera un resultado razonablemente bueno, en cuyo caso la probabilidad de realizar la nueva campaña se establece en un 90%.

a) Indicar el tamaño de muestra necesario.
b) ¿Qué cantidad de personas deberán como mínimo contestar afirmativamente para que se comience la nueva campaña?
c) Calcular la probabilidad de no realizar otra campaña cuando con esta se ha obtenido un aumento de 2 puntos.

RESPUESTAS: a) n = 1296 b) 216 personas c) 0.362

Problema 20: En una fábrica de bebidas refrescantes que comercializa su producción en botellas que contienen en promedio 65 calorías con un desvío estándar de 4 calorías, se está analizando la posibilidad de reformar el proceso de elaboración para disminuir los costos, siempre que esta reforma no aumente la variabilidad del contenido de calorías por botella. Se toma una muestra de 30 botellas elaboradas con el nuevo proceso y se observa que la variancia del contenido de calorías resulta igual a 18,24. Con un nivel de significación del 5 %, ¿qué decisión recomendaría? ¿Por qué?

RESPUESTA: Recomendaría reformar el proceso de elaboración, ya que al no rechazar Ho, no puede afirmarse que la variabilidad haya aumentado.

Problema 21: Los profesores de primer año de cierta universidad han advertido que existe gran disparidad en el nivel de conocimientos matemáticos de los alumnos recién ingresados, disparidad que quedó evidenciada en el hecho de que el desvío estándar de las notas que dichos alumnos obtuvieron en una prueba calificada de cero a cien, resultó igual a 32. Con el objeto de subsanar este inconveniente, las dos primeras semanas de clase se dedicaron al dictado de un curso de nivelación intensivo. Al finalizar el mismo se seleccionaron al azar 18 alumnos y se les tomó una prueba similar a la anterior, observándose que el desvío estándar de las notas obtenidas fue igual a 26. ¿Considera usted que el curso de nivelación dio resultado? (Usar $\alpha = 0{,}10$)

RESPUESTA: No, porque al no rechazar Ho, no puede afirmarse que la variabilidad de las notas haya disminuido.

Problema 22: Se supone que las latas de cierta conserva de tomates contienen 170 grs. Sin embargo, existe cierta variación entre las latas ya que las máquinas envasadoras no son absolutamente precisas. La distribución de contenido de conserva de una máquina envasadora es aproximadamente normal con un desvío de 10 grs, que se considera excesivo. Se desea reemplazar la máquina actual por otra solo si se tiene razonable evidencia de un mejor desempeño en relación a la homogeneidad de la dosificación. Se planea efectuar una prueba piloto con 30 latas, fijándose en un 5% la probabilidad de concluir erróneamente que la nueva máquina tiene mejor desempeño en cuanto a su variabilidad.

a) Establezca el juego de hipótesis adecuado, la regla de decisión y la condición de rechazo.
b) Si en la muestra se obtuvo un desvío de 8.5 grs, ¿qué conclusión debería sacarse?

c) En base a su respuesta anterior, ¿qué error podría estar cometiendo?

RESPUESTAS: a) CR: si χ^2_{calc} es menor a 17.71, se rechaza Ho

b) Como χ^2_{calc} es 20.95, no se rechaza Ho y se concluye que no existen evidencias para afirmar que la variabilidad del contenido de las latas disminuyó.

c) Error tipo II: suponer que la variabilidad con la nueva máquina no es menor cuando en realidad sí lo es.

Problema 23: Para un trabajo de microbiología se requieren placas para preparaciones microscópicas que tengan un espesor uniforme. La firma Placa ofrece unas placas que según ellos tienen un desvío de a lo sumo 0,1 μm. A fin de corroborarlo, se toma una muestra de 20 placas obteniéndose una variancia de 0.0169 μm^2. Asumiendo un riesgo del 5%, ¿dudaría en la afirmación de la empresa?

RESPUESTA: Como χ^2_{calc} = 32.11 es mayor a $\chi^2_{crít}$ = 30.144, se rechaza Ho y se concluye que existen evidencias para afirmar que la empresa está mintiendo.

Revisión conceptual

Indique si las siguientes afirmaciones son verdaderas o falsas, justificando su respuesta

a) El nivel de significación de una prueba de hipótesis mide la probabilidad de que Ho sea falsa
b) Si no se rechaza Ho significa que Ho es verdadera
c) Cuando se rechaza Ho es porque una muestra aleatoria no es coherente con la hipótesis nula
d) Las hipótesis se plantean sobre los estimadores
e) El error de tipo II consiste en aceptar que un tratamiento ineficaz produce efectos útiles.
f) ß es la potencia del ensayo

EJERCICIOS INTEGRADORES

Problema 1: La consultora Z llevó a cabo un estudio sobre reinserción laboral en directivos de empresas, para lo cual se basó en 60 casos de gerentes recientemente reincorporados a la actividad laboral. Con respecto al medio por el cual consiguieron su nuevo trabajo, 18 lo hicieron mediante avisos publicados, 6 mediante presentación espontánea a empresas o consultoras y el resto mediante contactos.

a) Con una confianza del 95%, estime qué porcentaje de los gerentes lograron su nuevo trabajo mediante contactos.
b) ¿Qué cantidad adicional de gerentes debería encuestarse si se desea reducir el error muestral anterior a la mitad?
c) Con respecto al tiempo que tardaron los 60 gerentes en obtener empleo los datos fueron:

Tiempo (meses)	0-2	2-4	4-6	6-8	8-10	10-12
cant.gerentes	25	6	15	8	3	3

En las actuales condiciones del mercado laboral, ¿cuánto estima que tardan, en promedio, en reinsertarse los gerentes? Trabaje con una confianza del 95%.

RESPUESTA: a) 47,6%; 72,4% b) 180 gerentes más c) 3,12 ; 4,68

Problema 2: En cierto establecimiento avícola se ha detectado últimamente que los pollitos recién nacidos son de bajo peso (34 g en promedio), lo que redunda en mayores tasas de mortalidad. Se sospecha que la causa de este problema reside en una temperatura demasiado elevada en las incubadoras, por lo que se disminuye la temperatura en las mismas. Luego de tres semanas de aplicada la modificación se eligieron al azar 70 pollitos recién nacidos y se les registró el peso, con los siguientes resultados:

Peso (g)	30-32	32-34	34-36	36-38	38-40
Cant. pollitos	8	16	25	18	3

a) ¿Hay evidencia de una mejora en el peso de pollitos? Asuma un riesgo del 5%.
b) Estime, en las nuevas condiciones de incubación y con una confianza del 95%, el peso promedio de los pollitos recién nacidos y su variabilidad.
c) ¿Cuántos pollitos más deberían pesarse si se quiere reducir el error muestral de la estimación efectuada en el punto anterior en un 20%?
d) Estime, en las nuevas condiciones de incubación y con una confianza del 95%, el porcentaje de pollitos recién nacidos con un peso de por lo menos 36g.

RESPUESTA: a) Sí (34.77 > 34.42) b) [34.27 ; 35.27 g]
c) 40 pollitos más d) [19.3 ; 40.7%]

Problema 3: El departamento de investigación de mercados de una empresa de proveedora de Internet desea conocer los hábitos de sus abonados. Con tal fin, seleccionó una muestra aleatoria de 120 usuarios particulares (no empresas) y obtuvo los siguientes datos:

- tiempo medio de conexión a Internet / día: 2.8 hs con un desvío de 1.8 hs
- solo 29 de los usuarios fueron mujeres

a) Estime con un 95% de confianza el tiempo medio de conexión diario a Internet de los usuarios particulares de dicha empresa y su desvío estándar.
b) Estime con un 10% de riesgo el porcentaje de mujeres usuarias de Internet de dicha empresa.
c) En base a la información precedente, la empresa proveedora de Internet desea aumentar la participación de las mujeres en la red, por lo que lanza una intensa campaña publicitaria en ese sentido. A los 3 meses de iniciada la campaña una muestra aleatoria de 150 usuarios particulares determinó que 51 de ellos eran mujeres. Con un riesgo del 5%, ¿considera que el porcentaje de mujeres en la red es superior al 30%?
d) Determine el tamaño de muestra necesario para que la probabilidad de detectar que el porcentaje de mujeres usuarias de Internet aumentó al 35% sea de 0.8.

RESPUESTAS: a) [2.48 ; 3.13 hs] y [1.59 ; 2.05 hs] b) [17.7 ; 30.6%]
c) No (0.34 no es mayor a 0.362)

Problema 4: La ganancia porcentual promedio para un proyecto de inversión no es la única preocupación para un administrador de inversiones, quien debe preocuparse por una gran variabilidad en las ganancias ya que una pérdida podría sacarlo del negocio. Un promotor de inversiones proyecta alcanzar una ganancia promedio de más del 12% con un desvío del 5%. Una muestra de 25 proyectos produjo una media del 12.5% con un desvío estándar del 3.6%.

a) El promotor desea estar seguro de que el desvío es menor al 5% con un nivel de significación del 10%. ¿Qué podría Ud. decir al respecto?
b) Estime el desvío estándar máximo que el promotor puede esperar con un riesgo del 10%.
c) Con un riesgo del 10%, ¿considera que el promotor logró la ganancia promedio esperada?

RESPUESTAS: a) Existen evidencias de que el desvío es inferior al 5% (12.442 < 15.659)
b) 4.46% c) No (12.5 no es mayor a 12.95)

Problema 5: El gobierno está pensando en crear un mejor sistema de financiamiento a las exportaciones en determinados rubros. Para ello debe estimar el monto exportado por empresa para las ubicadas en tales rubros. Una muestra aleatoria de empresas proporcionó los siguientes datos:

Monto exportado (en millones de $)	N° de empresas
0-10	3
10-20	6
20-30	11
30-40	9

a) Estimar el monto de exportaciones por empresa para el nivel de confianza de 0.95.
b) Si se sabe que anteriormente el monto exportado era de 27 millones de pesos ¿puede afirmarse que se ha modificado? Realice la prueba con un nivel de significación del 2%.
c) Si el desvío estándar del monto exportado era de 10 millones de pesos, construya la regla de decisión que le permita probar si también el desvío se ha modificado. Utilice $\alpha = 2\%$.
d) ¿Con qué supuestos trabajó?

RESPUESTAS: a) [20253015 ; 27678985]
b) No se rechaza Ho
c) No se rechaza Ho

Problema 6: Se desea realizar una campaña publicitaria con el fin de presentar en el mercado un nuevo producto de la empresa. El lanzamiento sería viable si se puede evidenciar un aumento significativo en los gastos mensuales promedio de las personas para productos de este tipo y si el porcentaje de compradores potenciales para el producto supera al 45%. Actualmente los gastos mensuales de las personas en productos similares (que tienen una distribución normal, ¿por qué?) poseen un promedio de 350$ con un desvío estándar 40$. Para decidir por la conveniencia o no del lanzamiento de la campaña se tomaron datos de una encuesta realizada sobre 200 personas, en una empresa del mismo ramo con la oferta de un producto similar, y se observó que 97 personas estarían dispuestas a comprar dicho producto y que el gasto promedio mensual en productos similares era de 383$ con un desvío estándar de 45$. Si se asume una probabilidad del 10% de decidir lanzar la campaña cuando no es aconsejable:

a) ¿Considera que el desvío estándar se mantendrá, asumiendo un riesgo del 10%?
b) ¿Qué decisión tomaría? Justifique estadísticamente dicha decisión.
c) Realice una explicación clara y sencilla de cuál fue el procedimiento empleado en el análisis del punto anterior, qué temas de los expuestos en clase se utilizaron y cómo los combinó.
d) ¿Por qué posee una distribución normal la variable en estudio? ¿Es necesario que sea normal para realizar los procedimientos que usted utilizó?

RESPUESTAS: a) El desvío no se mantiene (251,85 > 232,91). b) Decidiría no lanzar ya que si bien se prueba que el promedio es mayor a lo esperado (383 > 354.09) no puede probarse que la proporción supere al 45% (0.485 < 0.495)

Problema 7: Se desea diseñar una encuesta en la ciudad de Formosa a fin de determinar el porcentaje de hogares con necesidades básicas insatisfechas.

a) Indique el tipo de muestreo que considera más apropiado para llevar a cabo el estudio. Justifique su elección.
b) Si se desea efectuar la estimación sólo en la zona residencial de la ciudad de Formosa, ¿cambiaría la respuesta dada en el punto anterior? Fundamente su respuesta.
c) Determine el tamaño de la muestra para efectuar la estimación del punto b, sabiendo que se desea un error de estimación de ±5%, con una confianza del 95%. Indique las unidades del valor obtenido.

RESPUESTAS: a) Muestreo estratificado
b) Muestreo aleatorio simple o muestreo sistemático
c) 385 hogares

Problema 8: Una empresa dedicada a la fabricación de piezas se encuentra interesada en controlar su maquinaria con el fin de saber si su producción cumple los requisitos exigidos por sus compradores. Las variables influyentes en la producción son la dimensión de la pieza, que se sabe que poseen una distribución normal y la cantidad de piezas producidas con rugosidades

Para efectuar los controles pertinentes se tomaron muestras al azar de 100 piezas durante la producción del día y se encontró una dimensión promedio de 15.8 cm, una desviación estándar de 2.3 cm y 85 piezas que no presentaron rugosidad.

a) Estimar con una confianza del 90% la dimensión promedio de las piezas producidas.
b) ¿Cuál debería ser el tamaño de la muestra si se pretende que el error de muestreo anterior disminuya en un 30%?
c) Estimar el porcentaje máximo de unidades con rugosidad con un riesgo del 4%
d) ¿Cuántas muestras deberían tomarse si se pretende que el error de la estimación anterior disminuya a la mitad?
e) ¿cuál será el nivel de confianza para la estimación del ítem c) si se pretende un error de muestreo de 2% y se toma una muestra de 100 piezas?
f) Estimar el desvío estándar en la dimensión de las piezas asumiendo una confianza del 90%

RESPUESTAS: a) 15,4181 – 16,1819; b) 203; c) 0,2128; d) 401; e) 71,226%;
f) 2,0616 -2,6072

Problema 9: En los laboratorios de investigación de una compañía metalúrgica se ha desarrollado una nueva aleación para los rodillos con lo que se espera mejorar la duración promedio de los mismos que, con la aleación actual es una variable aleatoria con media 520 y desvío 38 (millones de revoluciones). Después de algunas discusiones, la gerencia decide efectuar un experimento a efectos de verificar en forma definitiva las cualidades del nuevo producto. Dado que la aleación actual está sólidamente impuesta en el mercado, el gerente considera que le parece razonable una probabilidad del 20% de lanzar el nuevo producto en caso de que su duración media sea un 4% superior a la actual. Dado el costo de la nueva aleación, se solicita al jefe de ingeniería que si el nuevo producto tuviera la misma duración

media que el actual, la probabilidad de lanzarlo sea muy pequeña, del 1 por mil, esto es 0,001.

a) Indicar hipótesis nula apropiada a esta situación, su condición de rechazo, el tamaño de muestra a tomar y la regla de decisión
b) ¿Cuál es la probabilidad de no lanzar el producto si tuviera una duración media 8% superior a la actual?
c) Luego de tomar la muestra calculada en el punto a) se obtuvo un promedio de 555 con un desvío de 65 (millones de revoluciones) y se encontró que la tercera parte de los rodillos presentaban asperezas que perjudicaban al proceso ¿Existen pruebas para afirmar que el desvío no se mantiene con un riesgo del 5%?
d) Teniendo en cuenta su conclusión anterior ¿Considera que existen pruebas para concluir que la duración promedio ha mejorado con la nueva aleación?
e) Estimar el porcentaje de todos los rodillos que no poseen asperezas con una confianza del 80%.
f) Estime el promedio de duración mínimo y el desvío estándar máximo en la duración con la nueva aleación, asumiendo un riesgo del 5%.
g) Si se desea reducir el error de la estimación promedio en un 40%, ¿cuántas muestras deberán tomarse?

RESPUESTAS: a) n= 17; b)0.077; c) Si, hay evidencias de que el desvío se modificó.
d) Con la información muestral No Rechazo la H0, no hay pruebas de que haya mejorado la duración promedio; e) 52%;81%;
f)527.48 ; 92.14; g) 44

PRACTICO 4. INFERENCIA PARA DOS POBLACIONES

Consideraciones generales:

En numerosas situaciones, un profesional puede encontrarse ante el problema de tener que tomar una decisión en base a la comparación de dos métodos de trabajo o entre dos productos, para elegir el que resulte mejor; o bien sólo determinar si existe una diferencia significativa entre los mismos. En todas estas situaciones deberá tomar decisiones teniendo como información sólo los resultados de las muestras obtenidas. Bajo estas circunstancias deberá utilizar las técnicas estadísticas para comparar las poblaciones de las cuales se extrajeron las muestras, utilizando conceptos ya introducidos como ensayos de hipótesis y estimación por intervalos de confianza.
Los casos que se verán son los más comunes, relacionados con dos muestras -independientes o no -, siempre bajo el supuesto de que las muestras deben provenir de poblaciones normales.

Se desarrollarán los siguientes casos:

1. Comparación de dos variancias poblacionales, utilizando la distribución F de Snedecor.
2. Comparación de dos medias poblacionales, utilizando la distribución normal o t
    a. Muestras independientes:
        i. Variancias conocidas
        ii. variancias desconocidas e iguales
        iii. variancias desconocidas y distintas
    b. Muestras dependientes o pareadas

Problemas resueltos:

Comparación de dos medias poblacionales con variancias desconocidas y supuestamente iguales

Una sucursal bancaria recibe numerosas quejas de sus clientes debido al excesivo tiempo de espera en los cajeros automáticos (definido como el tiempo que transcurre desde que el cliente se incorpora a la fila hasta que inicia la operación). Se registra entonces durante una semana el tiempo de espera de 50 clientes elegidos al azar, obteniéndose un promedio de 12 min con un desvío estándar de 5 min. Estos resultados son considerados excesivos, por lo que se implementa un proceso de mejora que incluye personal de orientación al cliente. Al cabo de dos meses de implementado el proceso se toma una nueva muestra de 60 clientes, obteniéndose esta vez un tiempo de demora promedio de 10 min con un desvío de 4

min. Analice la información y decida si la implementación del proceso fue efectiva, con un riesgo del 5%.

Solución:

El proceso de mejora será considerado efectivo si el tiempo promedio de demora de los clientes una vez implementado dicho proceso (μ_2) es inferior al tiempo de demora antes de la implementación del mismo (μ_1). Es decir:

$$H_0)\ \mu_1 \le \mu_2 \rightarrow \mu_1 - \mu_2 \le 0 \quad \textit{(el proceso no fue efectivo)}$$
$$H_1)\ \mu_1 > \mu_2 \rightarrow \mu_1 - \mu_2 > 0 \quad \textit{(el proceso si fue efectivo)}$$

Datos: $n_1 = 50$ $\bar{x}_1 = 12\ min$ $s_1 = 5$ min
$n_2 = 60$ $\bar{x}_2 = 10\ min$ $s_2 = 4$ min $\alpha = 0{,}05$

Como las varianzas poblacionales son desconocidas, primero debe probarse si son iguales o no.

Comparación de dos varianzas

$H_0)\ \sigma^2_1 = \sigma^2_2$ *(las varianzas de las dos poblaciones son iguales)*
$H_1)\ \sigma^2_1 \neq \sigma^2_2$ *(las varianzas de las dos poblaciones son distintas)*

CR: Si $F_{calc} < F_{crít1}$ o si $F_{calc} > F_{crít2}$ se rechaza la hipótesis nula
RD: Si se rechaza Ho se concluye que las varianzas son distintas

siendo $F_{crít1} = F_{\nu 1;\nu 2;\alpha/2} = F_{49;59;0.025} \cong 0{,}40576$ $F_{crít2} = F_{\nu 1;\nu 2;1-\alpha/2} = F_{49;59;0.975} \cong 2{,}464$

$$F_{calc} = \frac{S_1^2}{S_2^2} = \frac{5^2}{4^2} = 1{,}56$$

Conclusión: Como F_{calc} (que resultó igual a 1,56) no es menor que $F_{crít1}$ (0,40576) ni es mayor a $F_{crít2}$ (2,464) no se rechaza Ho. En consecuencia se infiere que las varianzas de las dos poblaciones (tiempo de demora antes y después de la implementación del proceso de mejora)no difieren significativamente.

Volvamos a la comparación de las dos medias poblacionales con varianzas poblacionales desconocidas, y en base al resultado de la prueba de hipótesis anterior, supuestamente iguales. Las hipótesis eran:

$$H_0)\ \mu_1 \le \mu_2 \rightarrow \mu_1 - \mu_2 \le 0 \quad \textit{(el proceso no fue efectivo)}$$
$$H_1)\ \mu_1 > \mu_2 \rightarrow \mu_1 - \mu_2 > 0 \quad \textit{(el proceso si fue efectivo)}$$

CR: Si $t_{calc} > t_{crít}$ se rechaza la hipótesis nula
RD: Si se rechaza Ho se concluye que el proceso fue efectivo

siendo $t_{crít} = t_{v1+v2-2;1-\alpha} = t_{108;0.95} = 1{,}6591$

Para obtener el t_{calc} debe calcularse previamente la varianza amalgamada s^2_a:

$$S_a^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{5^2.49 + 4^2.59}{50 + 60 - 2} = 20{,}08 \qquad \text{siendo } S_a = \sqrt{S_a^2} = \sqrt{20{,}08} = 4{,}48$$

$$t_{calc} = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{S_a\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(12 - 10) - 0}{4{,}48\sqrt{\frac{1}{50} + \frac{1}{60}}} = 2{,}33$$

Conclusión: Como t_{calc} (que resultó igual a 2,33) es mayor que $t_{crít}$ (1,6591) se rechaza Ho. En consecuencia se infiere que el tiempo de demora medio de todos los clientes después de la implementación del proceso de mejora es menor al tiempo de demora antes de la implementación del mismo y por lo tanto el proceso puede considerarse efectivo, siendo aconsejable su implementación en otras sucursales.

Ejercitación:

Problema 1: El ingeniero a cargo de una planta de envasado de detergente desea saber si existen diferencias en el volumen de llenado de los envases de detergente en dos tipos de máquinas. Los desvíos estándar del volumen de llenado son 0.03 y 0.05 litros respectivamente. Una muestra aleatoria de 15 envases llenados por la máquina 1 indicó un contenido promedio de 1.03 litros y una muestra similar de 20 envases de la máquina 2 indicó un contenido promedio de 0.98 litros. Con una significación del 5%, ¿existen diferencias entre los dos tipos de máquinas en el volumen promedio de llenado?

RESPUESTA: Sí ($z_{calc} = 3.68 > 1.96$)

Problema 2: Una compañía de transporte utiliza habitualmente neumáticos marca P. Recientemente han salido al mercado neumáticos de otra marca, que denominaremos M, que ofrecen una mayor duración aunque a un mayor costo. La compañía debe renovar constantemente sus neumáticos, pero antes de decidir la compra lleva a cabo una prueba donde se toman 12 de cada marca y se evalúa su rendimiento. Los neumáticos se utilizan hasta su desgaste. Se obtienen los siguientes resultados:

marca P : promedio de 36000 km
marca M : promedio de 38000 km

Por datos suministrados por las empresas se conoce el desvío estándar de la duración de los neumáticos, que es de 5000 km para la marca P y de 5800 km para la marca M. Con un nivel de significación del 10%, ¿aconsejaría a la compañía comprar la marca M?

RESPUESTA: No existen evidencias para aconsejar una marca sobre la otra por su rendimiento ($z_{calc} = -0.905$ no es menor que -1.282), por lo tanto se aconseja seguir con la marca P, de menor costo.

Problema 3: Algunas empresas incurren en gastos considerables para entrenar nuevos empleados. Existe un costo directo debido al programa de entrenamiento y un costo indirecto debido a que los empleados en entrenamiento deben alcanzar un cierto grado de eficiencia para contribuir significativamente al proceso de manufactura. Es por ello que se buscan programas que lleven a los empleados a un grado de eficiencia máxima en el menor tiempo posible. Cierta operación requiere de un entrenamiento de un mes para que el empleado alcance el grado deseado. Se ha sugerido un nuevo método que se desea comparar con el actual. Para ello se seleccionaron 20 empleados que se dividieron en dos grupos iguales. Un grupo fue entrenado con el método en uso (A) y el otro con el nuevo (B). Al cabo de un mes se obtuvieron los siguientes tiempos de operación (en min):

Método en uso	32	37	35	28	41	44	35	31	34	38
Método nuevo	35	31	29	25	34	40	27	32	31	33

a) ¿Presentan estos datos evidencia suficiente de que el nuevo método disminuye los tiempos de operación? Utilice un nivel de significación del 5 %.

RESPUESTAS: a) Sí (1.89 > 1.73)

Problema 4: Una empresa de investigación de mercado proporciona a un fabricante de electrodomésticos estimaciones sobre las ventas de sus productos al por menor a partir de muestreos en negocios minoristas. El gerente de marketing tiende a fijarse en la estimación y a ignorar el error de ésta. Este mes, una muestra aleatoria simple de 75 negocios da una media de ventas de 52 unidades para cierto electrodoméstico, con un desvío estándar de 13 unidades. Durante el mismo mes del año anterior, otra muestra aleatoria simple de 53 negocios da una media de ventas de 49 unidades con un desvío estándar de 11 unidades. Un aumento de 49 a 52 unidades representa un incremento del 6%. El gerente de marketing está contento porque las ventas han aumentado un 6%.

a) Efectúe la prueba de hipótesis tendiente a determinar si las ventas promedio de este mes difieren significativamente de las del mismo mes del año anterior, con un nivel de significación del 5%. Asuma igualdad de varianzas.
b) Compare los resultados de los dos ítems anteriores. Explique con un lenguaje que el gerente pueda entender, por qué no estamos seguros de que las ventas hayan subido un 6% y que incluso podrían haber bajado.
(extraído de Moore, 2000)

RESPUESTAS: a) No hay evidencias (t_{calc} = 1,37 no es mayor $t_{crít2}$ = 1,98)

Problema 5: Se desea comparar los niveles de contaminación sonora de dos áreas de una ciudad. Para ello, en cada uno de las áreas se eligen puntos al azar y se determina la intensidad sonora en la vía pública. Los resultados (en decibeles) fueron:

Área Norte	Área Sur
70.1	74.1
70.4	75.4
75.8	76.2
67.5	79.9
68.4	70.5
73.6	70.1
76.9	74.9
75.7	75.3
71.4	70.3
70.3	70.7
71.1	
74.4	
70.2	
74.3	

Con $\alpha = 0.05$ ¿podría afirmar si existen diferencias en los niveles promedio de contaminación sonora de las dos áreas? Asuma que la intensidad sonora sigue una distribución normal (¿Por qué es necesario efectuar tal suposición?) Si fuese el intendente de la ciudad, ¿cuál de las dos áreas consideraría como prioritaria por mayor contaminación acústica? Fundamente su respuesta.

En Excel: Herramientas > Análisis de datos > Prueba F para varianzas de dos muestras

Prueba F para varianzas de dos muestras

	Área Norte	**Área Sur**
Media	72,15	73,74
Varianza	8,62	10,61
Observaciones	14	10
Grados de libertad	13	9
F	0,813	
P(F<=f) una cola	0,356	
Valor crítico para F (una cola)	0,368	

En Excel: Herramientas > Análisis de datos > Prueba t para dos muestras suponiendo varianzas iguales

Prueba t para dos muestras suponiendo varianzas iguales

	Área Norte	**Área Sur**
Media	72,15	73,74
Varianza	8,62	10,61
Observaciones	14	10
Varianza agrupada	9,44	
Diferencia hipotética de las medias	0	
Grados de libertad	22	
Estadístico t	-1,250	
P(T<=t) una cola	0,112	
Valor crítico de t (una cola)	1,717	
P(T<=t) dos colas	0,224	
Valor crítico de t (dos colas)	2,074	

Problema 6: Como consecuencia de los resultados que se muestran en el problema 7 del Práctico 2, los responsables del área de salud de dicha localidad decidieron lanzar una campaña de prevención de adicciones. Luego de tres meses de campaña, una muestra de 50 adolescentes elegidos al azar arrojó un promedio de 0.38 lts de alcohol consumidos semanalmente con un desvío de 0.24 lts.

a) ¿Considera, con un riesgo del 5%, que la campaña ha sido efectiva?
b) ¿Qué error puede estar cometiendo como resultado de la prueba efectuada en el punto anterior? Indique en qué consiste, en términos del problema.

RESPUESTAS: a) No (t_{calc} = 0.36 no es mayor que 1.6624)
b) Como no se rechazó Ho se puede estar cometiendo un error de tipo II.

Problema 7: La densidad de la cerveza es una cualidad importante para mantener la calidad de la misma. Puede verse afectada por el tiempo de fermentación, variaciones en los ingredientes o diferencias en el equipo de fermentación. Un fabricante con dos líneas de producción a hecho ligeros cambios en la línea 2, buscando reducir la variabilidad en la densidad de la cerveza producida. Se tomaron 23 muestras de cerveza de ambas líneas de producción y se determinó la densidad de la misma, obteniéndose los siguientes resultados:

promedio de la línea 1 = 3,2 desvío estándar 1 = 1,04
promedio de la línea 2 = 3,0 desvío estándar 2 = 0,69

a) Determine para ambas líneas el coeficiente de variabilidad. ¿Qué podría Ud. concluir?
b) Con los datos obtenidos, ¿existe evidencia para indicar una variabilidad menor en la línea 2 asumiendo un riesgo de equivocarse del 5%?
c) Las modificaciones efectuadas, ¿alteraron la densidad promedio de la cerveza? Asuma un riesgo del 5%.

RESPUESTAS: a) $CV_1 = 32.5\%$, $CV_2 = 23\%$ b) Sí ($F_{calc} = 2.27$)
c) No ($t_{calc} = 0.77$ no es mayor que $t_{38,\ 0.975} = 2.024$)

Problema 8: En un estudio efectuado a fin de caracterizar la producción de aceite de oliva en la provincia de Catamarca, se estudiaron dos de las variedades más difundidas de aceitunas. Muestras de distintos ejemplares elegidos al azar fueron secadas en estufa y se les determinó el contenido porcentual en aceite, con los siguientes resultados:

Variedad	% aceite									
Arauco	10.5	12.2	9.3	12.5	11.1	10.5	11.4	9.6	12.6	13.4
Carolea	16.4	14.8	17.8	8.3	11.9	15.5	13.4	16.0	15.8	18.2

Las diferencias observadas ¿se deben al azar? Es decir, ¿existen diferencias significativas en el contenido promedio porcentual de aceite de ambas variedades? De ser así, ¿cuál de ellas recomendaría por su mayor rendimiento promedio? Asuma un riesgo del 5%. Compare sus resultados con la salida generada en Excel.

Prueba F para varianzas de dos muestras		
	Arauco	*Carolea*
Media	11,31	14,81
Varianza	1,841	8,741
Observaciones	10	10
Grados de libertad	9	9
F	0,210616	
P(F<=f) una cola	0,01484	
Valor crítico para F (una cola)	0,314575	

Prueba t para dos muestras suponiendo varianzas desiguales

	Arauco	*Carolea*
Media	11,31	14,81
Varianza	1,841	8,741
Observaciones	10	10
Diferencia hipotética de las medias	0	
Grados de libertad	13	
Estadístico t	-3,4024	
P(T<=t) una cola	0,0024	
Valor crítico de t (una cola)	1,7709	
P(T<=t) dos colas	0,0047	
Valor crítico de t (dos colas)	2,1604	

Problema 9: BJ y Asociados están probando dos comerciales para una compañía productora de jugos. Ambos comerciales se mostraron a 12 individuos, a quienes se pidió que los calificaran en una escala de 1 a 100. Los resultados fueron:

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Aviso 1	95	59	73	65	32	45	60	83	27	50	63	95
Aviso 2	87	65	80	73	45	39	57	81	33	40	66	93

Utilice un nivel de significación del 10 % para la prueba que determina si los panelistas apreciaron más alguno de los dos avisos. ¿A qué conclusión debería llegar la empresa BJ y Asociados? Compare sus resultados con la salida generada con Excel.

Prueba t para medias de dos muestras emparejadas

	AVISO1	*AVISO2*
Media	62,25	63,25
Varianza	483,6591	415,1136
Observaciones	12	12
Coeficiente de correlación de Pearson	0,9453	
Diferencia hipotética de las medias	0	
Grados de libertad	11	
Estadístico t	-0,48207	
P(T<=t) una cola	0,31960	
Valor crítico de t (una cola)	1,36343	
P(T<=t) dos colas	0,63920	
Valor crítico de t (dos colas)	1,79589	

Problema 10: La gerencia de ventas de una cadena de mueblerías diseñó un plan de incentivos para los vendedores. A fin de evaluar este plan innovador, seleccionaron 6 agentes de ventas aleatoriamente y registraron las ventas promedio diarias antes y después del plan (en pesos).

Vendedor	Antes	Después
RL	320	340
MP	290	295
BA	421	475
FF	510	510
EG	210	228
PF	402	500

a) ¿Hubo un incremento significativo en las ventas promedio semanal de los vendedores debido al plan innovador de incentivos? Utilice un nivel de significación del 0,05.

d) Explique en qué consiste el error tipo I y el error tipo II en términos del problema. ¿Cuál de los dos podría estar cometiendo según su conclusión en el punto anterior? ¿Cuáles serían las consecuencias de cometer ese error?

RESPUESTA: a) Sí (-2.14 < -2.02); b) Error tipo I: concluir que las ventas semanales de los vendedores aumentaron, es decir que el plan de incentivos fue efectivo, cuando en realidad no fue así. Error tipo II: concluir que las ventas semanales de los vendedores no aumentaron, es decir que el plan de incentivos no fue efectivo, cuando en realidad sí lo fue. En este caso se podría estar cometiendo un error tipo I, con una probabilidad máxima de 0,05.

Problema 11: El dueño de una inmobiliaria desea comprobar la eficacia de un nuevo empleado en la tasación de propiedades. Selecciona 8 propiedades cualquiera y envía por separado al nuevo empleado y a un empleado con probada experiencia en el oficio a tasarlas. Los resultados (en miles de US$) fueron:

		Propiedad							
		1	2	3	4	5	6	7	8
Vendedor	nuevo	25	42	150	100	78	29	62	95
	experimentado	24	45	131	98	65	30	58	86

¿Cuáles son las conclusiones de la prueba? Asuma α =0.10.

RESPUESTA: Se concluye que existen evidencias para afirmar que las tasaciones del vendedor nuevo son mayores a las del experimentado (2.06 > 1.8946).

Problema 12: Una importante compañía alimenticia que fabrica hamburguesas adquiere la materia prima a dos proveedores. El contenido promedio de grasa de ambos proveedores es el mismo, pero se sospecha que la variabilidad en el contenido graso puede diferir. El desvío estándar del contenido graso en una muestra de 18 lotes adquiridos al proveedor 1 fue de 8.9%, mientras que una muestra aleatoria de 15 lotes de la compañía 2 arrojó un desvío estándar de 5.1%. ¿Existe evidencia suficiente para concluir que la variabilidad de las dos poblaciones es diferente? Utilice α =0.05.

RESPUESTA: Sí (F_{calc} = 3.04 > 2.90)

Revisión conceptual

Indique si las siguientes afirmaciones son verdaderas o falsas, justificando su respuesta

Se realiza un estudio para saber si dos tratamientos de quimioterapia presentan diferencias en cuanto al tiempo de supervivencia de los pacientes. No se encontró diferencia estadísticamente significativa. ¿Cuál de las siguientes razones podrían ser causantes del resultado?

a) Los tratamientos ofrecen tiempos de supervivencia muy diferentes.
b) El nivel de significación es demasiado alto.
c) Las muestras son demasiado grandes.
d) Las muestras son demasiado pequeñas.
e) Nada de lo anterior.

PRACTICO 5. PRUEBAS CHI-CUADRADO

Consideraciones generales:

En los temas anteriores se determinó que la aplicación de determinados conceptos de la inferencia estadística quedaba supeditada al cumplimiento de ciertos supuestos. Uno de ellos se refiere a la ley de distribución de la población de la que extrae la muestra. La forma de determinar si la población responde a determinada ley de distribución teórica es la prueba de bondad de ajuste.
En algunos otros estudios es necesario trabajar con variables cualitativas; en estos casos se utilizan las pruebas de independencia entre atributos.

Problemas resueltos:

Prueba de Bondad de ajuste

El Gerente de Personal de cierta empresa está preocupado por el ausentismo. Desea saber si el mismo está distribuido uniformemente durante los 5 días laborales. Para ello extrajo una muestra al azar de 120 ausencias, obteniendo los siguientes resultados:

Días	lunes	martes	miércoles	jueves	viernes
Cant. ausentes	35	18	21	17	29

Con esta información, ¿qué concluiría el Gerente de Personal? Utilice un nivel de significación del 5%.

Solución:

Se trata de probar si existe un buen ajuste a una proporción dada; en este caso el gerente desea probar si es igualmente probable faltar cualquier día de la semana. Es decir:

Ho) el ausentismo está distribuido uniformemente en los 5 días laborables; o bien:
P(faltar un lunes) = P(faltar un martes) = …. = P(faltar cualquier día)
H_1) el ausentismo no está distribuido uniformemente; o bien:
P(faltar algún día de la semana) ≠ P(faltar)

Para decidir se deben comparar las frecuencias observadas en el muestreo con las frecuencias que se esperaría observar si la hipótesis nula fuera cierta. Si así fuera, las 120 ausencias estarían distribuídas homogéneamente, con una probabilidad de 0.2 de faltar un día dado de la semana, es decir que se esperarían 24 ausencias en cada día laborable:

	lunes	martes	miércoles	jueves	viernes
Frecuencias observadas (OBS)	35	18	21	17	29
Frecuencias esperadas (ESP)	24	24	24	24	24

El estadístico de prueba compara las frecuencias observadas con las esperadas:

$$\chi^2_{calc} = \sum_i \frac{(OBS_i - ESP_i)^2}{ESP_i}$$

CR: Si $\chi^2_{calc} > \chi^2_{crít}$, se rechaza Ho
RD: Si se rechaza Ho se concluye que existen evidencias de que el ausentismo no está distribuido uniformemente.
Siendo $\chi^2_{crit} = \chi^2_{k-1-m;1-\alpha}$ donde
k = cantidad de categorías, en este caso 5 (los días laborables)
m = cantidad de parámetros estimados, en este caso ninguno

Por lo tanto $\chi^2_{crit} = \chi^2_{4;0.95} = 9.488$
y $\chi^2_{calc} = \frac{(35-24)^2}{24} + \frac{(18-24)^2}{24} + ... = 10$
Conclusión: Como $\chi^2_{calc} > \chi^2_{crít}$, se rechaza Ho y el gerente debería concluir que existen evidencias para afirmar que el ausentismo no está distribuido uniformemente.

Tablas de Contingencia: Prueba de Independencia

Una marca de cervezas decidió efectuar un estudio de mercado a fin de identificar las preferencias de sus consumidores. De un total de 200 personas encuestadas, 110 manifestaron preferir la cerveza rubia y 60 la cerveza sin alcohol, mientras que el resto prefirió la negra. De aquellos que prefirieron la cerveza rubia, el 40% son mujeres, mientras que de los que prefieren la cerveza sin alcohol, 19 son hombres. Finalmente, 23 encuestados son hombres que prefieren la cerveza negra. Con un riesgo del 5%, ¿presentaron los 3 tipos de cerveza igual preferencia en ambos sexos?

Solución:

Las hipótesis a testear son:

Ho) la preferencia por cierto tipo de cerveza es independiente del sexo del consumidor
H1) la preferencia por cierto tipo de cerveza depende del sexo del consumidor

Al igual que en la situación anterior se deben comparar las frecuencias observadas en el muestreo con las frecuencias que se esperaría observar si la hipótesis nula fuera cierta, es decir si los dos criterios de clasificación de los datos (tipo de cerveza preferida y sexo) fueran independientes. La tabla de contingencia con las frecuencias observadas (OBS) en el muestreo es:

	Cerveza preferida			
	Rubia	Sin alcohol	Negra	Totales
Mujeres	44	41	7	92
Hombres	66	19	23	108
Totales	110	60	30	200

Para construir la tabla de contingencia con las frecuencias esperadas (ESP) suponiendo independencia se procede de la siguiente manera. La proporción de consumidores que prefieren la cerveza rubia, sin importar el sexo de los mismos, es de 110/200, es decir 0.55 (55%). Si los dos sexos tienen idénticas preferencias se esperaría que el 55% de las mujeres y el 55% de los hombres prefieran la cerveza rubia, es decir 50,6 mujeres (el 55% de 92) y 59,4 hombres (el 55% de 108). Efectuando el mismo razonamiento con los otros tipos de cerveza es posible construir la siguiente tabla de frecuencias esperadas:

	Cerveza preferida			
	Rubia	Sin alcohol	Negra	Totales
Mujeres	50.6	27.6	13.8	92
Hombres	59.4	32.4	16.2	108
Totales	110	60	30	200

Obsérvese que los totales no se modificaron. El estadístico de prueba compara las frecuencias observadas con las esperadas:

$$\chi^2_{calc} = \sum_i \frac{(OBS_i - ESP_i)^2}{ESP_i}$$

CR: Si $\chi^2_{calc} > \chi^2_{crít}$, se rechaza Ho
RD: Si se rechaza Ho se concluye que existen evidencias de que la preferencia por cierto tipo de cerveza depende del sexo del consumidor

Siendo $\chi^2_{crit} = \chi^2_{(F-1)(C-1);1-\alpha}$ donde

F = cantidad de filas, en este caso 2
C = cantidad de columnas, en este caso 3

Por lo tanto $\chi^2_{crit} = \chi^2_{2;0.95} = 5.991$

$$\text{y } \chi^2_{calc} = \frac{(44-50.6)^2}{50.6} + \frac{(66-59.4)^2}{59.4} + \ldots = 19.85$$

Conclusión: Como $\chi^2_{calc} > \chi^2_{crít}$, se rechaza Ho y se concluye que la preferencia por la cerveza depende del sexo del consumidor.

Ejercitación:

Problema 1: Las estadísticas indican que en una universidad privada el 60% de los inscriptos se anotan en la facultad de Cs. Económicas, el 15% en la de Abogacía y el resto en la de Ingeniería. Antes del inicio del ciclo lectivo y a fin de planificar la distribución de docentes y cursos, las autoridades están interesadas en determinar si se mantienen las proporciones históricas de inscripción en las distintas facultades. Hasta la fecha registran un total de 1200 inscripciones, de las cuales 760 corresponden a la facultad de Cs. Económicas, 200 a la de Abogacía y el resto a Ingeniería. Con un riesgo del 1% ¿considera que en el presente ciclo lectivo se detectan cambios con respecto a las proporciones históricas?

RESPUESTA: Existen evidencias de que existen cambios en la proporciones históricas de inscriptos en las distintas facultades (16.44 > 9.21).

Problema 2: El listado generado por computadora del gerente de una tienda contiene a todos los empleados de venta e indica que el 70% son empleados de tiempo completo, el 20% son empleados de medio tiempo y el 10% están suspendidos temporariamente o se encuentran con licencia. Una muestra aleatoria de 50 empleados del listado indica que 40 son empleados de tiempo completo, 6 son de medio tiempo y 4 están suspendidos o con licencia. Determine si esta muestra es representativa de la población con un riesgo del 10%, es decir, ¿Considera que hay evidencias para probar que existen diferencias entre las proporciones de empleados en la muestra y las del listado general?

RESPUESTA: Puede considerarse que la muestra es representativa, ya que no existen evidencias de que la proporción de cada tipo de empleado difiera de las proporciones poblacionales (2.51 no es mayor a 4.605)

Problema 3 Se desea probar con un nivel de significación del 5 % si un dado está cargado. Para ello se realizan 90 tiradas y se obtiene la siguiente información:

cara del dado	1	2	3	4	5	6
cantidad de veces que salió	10	16	20	15	17	12

RESPUESTA: De las tiradas no surgen evidencias para decir que el dado está cargado (4.27 < 11.07).

Problema 4: La agencia publicitaria Atlántico investiga la relación entre el tipo preferido de mensaje para una campaña contra el SIDA y el nivel socioeconómico para una muestra de jóvenes de ambos sexos. La cantidad de individuos de cada nivel socioeconómico que prefirieron cada uno de los mensajes fue:

	Mensaje preferido		
	A	B	C
Bajo	25	40	70
Medio	29	30	31
Alto	45	20	10

a) Pruebe si el nivel socioeconómico se relaciona con la preferencia de los mensajes. Utilice un nivel de significación del 1%.
b) ¿Recomendaría utilizar distintos mensajes según el nivel socioeconómico de los jóvenes a los que va dirigido? Fundamente su respuesta.
c) Indique los supuestos de la prueba.

RESPUESTA: a) La preferencia por el mensaje depende del nivel socioeconómico (45,34 > 13.277)

b) Sí, ya que se encontraron evidencias de que la preferencia por el mensaje depende del nivel socioeconómico. Para el nivel socioeconómico bajo recomendaría el mensaje C, para el alto, el A, mientras que para el medio, es indistinto.
c) Se supone que las respuestas son independientes entre sí. Además todas las frecuencias esperadas deben ser de por lo menos 5, por lo que en determinadas ocasiones puede ser necesario combinar dos o más categorías.

Problema 5: Una fábrica de equipos de refrigeración selecciona la producción de un día cualquiera y efectúa un control total de los equipos producidos. Se observa que en el turno mañana se fabricaron 119 equipos de los cuales 16 estaban fallados; en el turno tarde se fabricaron 252 equipos, siendo 24 defectuosos; finalmente en el turno noche, de 93 equipos fabricados, 18 estaban fallados.

a) ¿Existen diferencias al 10% en el desempeño de los turnos?
b) ¿Cuál de los 3 turnos considera que tiene peor desempeño? Justifique.

RESPUESTAS: a) Sí (6.13 > 4.605)
b) el turno noche, con un 19,35% de equipos defectuosos.

Problema 6: La confianza ciudadana es un signo de democracia sana y un requisito indispensable para lograr mayor gobernabilidad en una sociedad. Si en una sociedad hay confianza ciudadana en sus representantes, los miembros de esa sociedad querrán cumplir con sus obligaciones y ser partícipes activos en la esfera de la vida pública. El Observatorio de la Deuda Social Argentina efectúa encuestas en ciudadanos de 18 años o más de todo el país con respecto a la confianza en las instituciones públicas. A continuación se muestran los resultados obtenidos a partir de 2000 encuestas con respecto al grado de confianza en el Congreso según nivel educativo.

Confianza en el Congreso Nacional	*Secundario incompleto o menos*	*Secundario completo o más*
Alta confianza	186	125
Poca confianza	682	359
Ninguna confianza	482	166

a) Plantee las hipótesis que considere apropiadas, en términos del problema. Concluya, con un riesgo del 1%.
b) De existir diferencias, ¿podría decir en qué nivel educativo existe mayor confianza en el Congreso? Justifique su respuesta.

RESPUESTA: a) El grado de confianza en el congreso está asociado al nivel de estudios (24.25 > 9.21); b) Los individuos con mayor nivel educativo muestran mayor confianza. Por ejemplo, el 19% de ellos manifiesta alta confianza contra un 13,7% del otro grupo

PRACTICO 6. ANALISIS DE REGRESION Y DE CORRELACION LINEAL SIMPLE

Consideraciones generales:

En este caso se analizan situaciones que involucran al menos dos variables y el objetivo es estudiar la relación entre ellas. Se estudiarán dos análisis: el de regresión y el de correlación.
El análisis de regresión estudia la dependencia de una variable, la variable dependiente o de respuesta, en una o más variables, las variables independientes o explicativas. Se utiliza con propósitos de estimación o predicción. La variable dependiente debe ser aleatoria, en cambio la independiente puede serlo o no. Se estudiará el ajuste a un modelo de regresión lineal simple, que utiliza una única variable independiente, y múltiple, que utiliza más de una variable independiente para predecir la variable dependiente.
El análisis de correlación, en cambio, se utiliza para medir la fuerza de asociación entre dos variables aleatorias.

Regresión
Las técnicas de regresión permiten hacer predicciones sobre los valores de cierta variable Y (dependiente o explicada), a partir de los de otra X (independiente o explicativa), entre las que intuimos que existe una relación. Para ilustrarlo supongamos que sobre un grupo de personas observamos los valores que toman las variables
X = altura medida en centímetros
Y = altura medida en metros
No es necesario hacer grandes esfuerzos para intuir que la relación que hay entre ambas es:

$$Y = \frac{X}{100}$$

Obtener esta relación es menos evidente cuando lo que medimos sobre el mismo grupo de personas es
X = altura medida en centímetros
Y = peso en kilogramos
La razón es que no es cierto que conocida la altura Xi de un individuo, podamos determinar de modo exacto su peso Yi (por ej. dos personas que miden 1,70 m pueden tener pesos de 60 y 65 kilos). Sin embargo, alguna relación entre ellas debe existir, pues parece mucho más probable que un individuo de 2 m pese más que otro que mida 1,20 m. Es más, nos puede parecer más o menos aproximada una relación entre ambas variables como la siguiente:

$$Y = X - 110 \pm \text{error}$$

A la deducción, a partir de una serie de datos, de este tipo de relaciones entre variables, es lo que denominamos regresión. Mediante las técnicas de regresión expresamos una variable $\hat{Y}$ como función de otra variable X

$$\hat{Y} = f(X)$$

Esto es lo que denominamos relación funcional. El criterio para construir el modelo, tal como citamos anteriormente, es que la diferencia entre el valor real de Y y el valor teórico o estimado de Y a partir de dicha relación ($\hat{Y}$) sea pequeña. Dicha diferencia se conoce como error o residuo:

$$e_i = Y - \hat{Y}$$

El objetivo será buscar la función (también denominada modelo de regresión) $\hat{Y} = f(X)$ que lo minimice.

Regresión lineal

La forma de la función f en principio podría ser cualquiera, lineal o no lineal, pero por el momento nos vamos a limitar al caso de la regresión lineal. Con este tipo de regresiones nos conformamos con encontrar relaciones funcionales de tipo lineal, es decir que el modelo que utilizaremos, conocido como modelo de regresión lineal, es:

$$Y = \alpha + \beta x + \varepsilon$$

donde α y β son los parámetros del modelo, siendo α la ordenada al origen y β el coeficiente de regresión de Y sobre X o pendiente de la recta. La letra ε corresponde al término del error, y es la variable aleatoria que explica la variabilidad en Y que no se puede explicar con la relación lineal entre X e Y.

Obsérvese que la relación anterior explica cosas como que si X varía en 1 unidad, Y varía la cantidad β. Es decir que β mide la variación de Y por incremento unitario de X, mientras que α indica el valor de Y cuando X=0. Por lo tanto:

- Si $\beta>0$, cuando X aumenta, Y también lo hace.
- Si $\beta<0$, cuando X aumenta, Y disminuye.
- Si $\beta=0$, cuando X aumenta o disminuye, Y no se ve afectada (Y no depende de X)

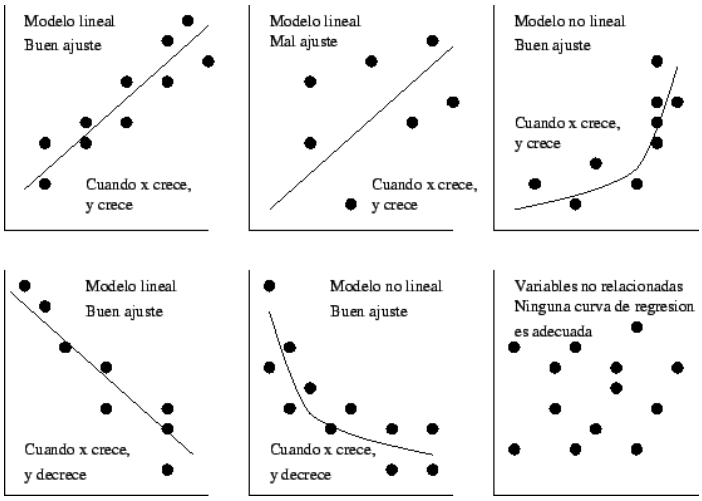
En consecuencia, en el caso de las variables peso y altura lo lógico sería encontrar que $\beta>0$.

Desafortunadamente, los parámetros α y β son usualmente desconocidos, por lo que es necesario estimarlos a partir de una muestra. Sus estimadores, a y b, permiten expresar la ecuación estimada de regresión:

$$\hat{y} = a + bx$$

con el menor error posible entre $\hat{Y}$ e Y, es decir que el modelo supone que la media o valor esperado de ε es cero.

Figura 1: Diferentes nubes de puntos y modelos de regresión para ellas.



El problema que se plantea es entonces el de cómo estimar las constantes α y ß a partir de un conjunto de n observaciones de forma que se minimice el error. El error que se comete al aproximar Y mediante $\hat{Y}$ se mide calculando la suma de las diferencias entre los valores reales y los estimados (residuos) elevadas al cuadrado (para que sean positivas y no se compensen los errores):

$$\sum (y_i - \hat{y}_i)^2 = \sum e_i^{\ 2}$$

y se hallan los estimadores a y b que hagan mínima dicha sumatoria. Este método se conoce como el método de los cuadrados mínimos.
Mediante una serie de procedimientos matemáticos se llega a la expresión:

$$b = Sxy / Sxx \qquad a = \bar{y} - b\bar{x}$$

$$\text{siendo } S_{xy} = \sum xy - n\bar{x}\bar{y} \quad \text{y} \quad S_{xx} = \sum x^2 - n\bar{x}^2$$

Supuestos del modelo de regresión lineal

- La variable independiente X se supone medida sin error. Es decir se supone fija; sus distintos valores están fijados de antemano.
- Los valores esperados de la variable aleatoria Y para cada valor de la variable X están alineados, es decir: $\mu_{(Y/X)} = \alpha + \beta X$.
- Supuesto de normalidad: Para cada valor de la variable X la subpoblación de la variable Y sigue una distribución normal; las subpoblaciones son independientes.
- Supuesto de homocedacia: Las varianzas de las subpoblaciones son iguales.
- Estos supuestos pueden reunirse en uno solo diciendo que en el modelo $Y = \alpha + \beta x + \varepsilon$ los ε_i son variables aleatorias independientes con distribución normal, media $\mu = 0$ y varianza σ^2_ε

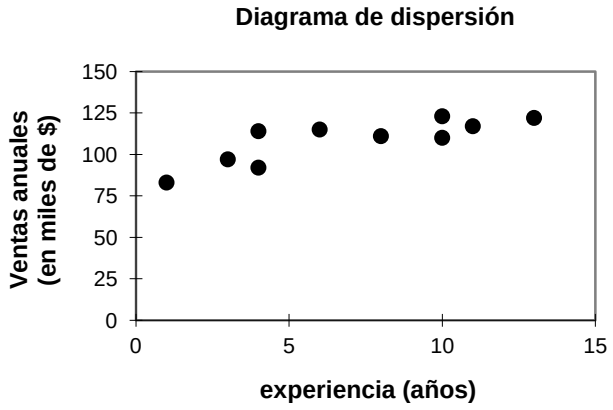
Ejemplo

Un gerente de ventas sostiene que la experiencia es fundamental a la hora de concretar una operación. A fin de darle mayor crédito a su aseveración, selecciona un grupo de diez

vendedores de su empresa con distintos años de experiencia en el rubro y estudia los reportes de ventas anuales (en miles de $) de los mismos:

Vendedor	1	2	3	4	5	6	7	8	9	10
Años de experiencia	13	1	11	4	6	3	10	8	4	10
Ventas anuales	122	83	117	92	115	97	110	111	114	123

Aquí lo que interesa es determinar si las ventas están relacionadas con la experiencia de los vendedores y eventualmente hallar un modelo que permita predecir las ventas anuales (Y, variable dependiente o explicada) en función de los años de experiencia de los vendedores (X, variable independiente o explicativa).
Con los datos de la muestra el primer paso consiste en graficar el diagrama de dispersión para ver si la variable respuesta Y depende o no, en cierto grado, de la variable independiente X y si la relación entre ellas puede ser razonablemente expresada por una ecuación lineal.



El segundo paso consiste en hallar la ecuación de regresión aplicando el método de los cuadrados mínimos. Para ello es necesario efectuar una serie de cálculos:

Vendedor	*x = años de experiencia*	*y = ventas anuales*	*x.y*	x^2	y^2
1	13	122	1586	169	14884
2	1	83	83	1	6889
3	11	117	1287	121	13689
4	4	92	368	16	8464
5	6	115	690	36	13225
6	3	97	291	9	9409
7	10	110	1100	100	12100
8	8	111	888	64	12321
9	4	114	456	16	12996
10	10	123	1230	100	15129
Total	70	1084	7979	632	119106

n = 10 $\quad \bar{x} = \Sigma x / n = 70/10 = 7$ $\quad \bar{y} = \Sigma y / n = 1084/10 = 108.4$

$S_{xx} = \sum x^2 - n\bar{x}^2 = 632 - 10 \text{ x } 7^2 = 142$

$S_{yy} = \sum y^2 - n\bar{y}^2 = 1191062 - 10 \text{ x } 108.4^2 = 1600.4$

$S_{xy} = \sum xy - n\bar{x}\bar{y} = 7979 - 10 \text{ x } 7^2 \text{ x } 108.4^2 = 391$

Lo que se busca es la ecuación de la recta, $\hat{y} = a + bx$, que mejor aproxima a la nube de puntos. Los coeficientes de esta recta son:

$b = S_{xy} / S_{xx} = 391 / 142 = 2{,}754$

$a = \bar{y} - b\bar{x} = 108{,}4 - 2{,}754 \text{ x } 7 = 89{,}122$

Así, la ecuación de regresión lineal resulta ser: $\hat{Y} = 89{,}122 + 2.754\ X$. Es decir que por cada año adicional de experiencia, un vendedor incrementa sus ventas anuales en 2,754 miles de $, es decir 2754 $. Por otro lado, se estima que un vendedor sin experiencia reportará ventas anuales por 89122 $.

Por tanto, para un vendedor con 9 años de experiencia (X=9), el modelo lineal predice unas ventas anuales de:

$\hat{y} = a + bx = 89{,}122 + 2{,}754 \text{ x } 9 = 113{,}91$ miles de $

En este punto hay que preguntarse si realmente esta predicción puede considerarse fiable. Para dar una respuesta, es necesario estudiar la bondad de ajuste al modelo lineal.

Evaluación de la recta de regresión: Bondad de ajuste al modelo

Una vez que se ha obtenido la ecuación de regresión estimada, ésta debe ser evaluada para detectar si describe adecuadamente la relación funcional entre las variables y si puede ser utilizada de manera efectiva con fines de estimación y predicción.

Para ello se requiere conocer la variabilidad o dispersión de los puntos alrededor de la recta, es decir la varianza poblacional del modelo $\sigma^2_{y/x}$. Teniendo en mente que dicha varianza mide las diferencias entre el valor real de Y y el teórico, obtenido mediante la ecuación de regresión, resulta obvio que cuanto mayor sea esta varianza, peor resultará el ajuste al modelo lineal propuesto (ver Fig.2).

El estimador puntual de $\sigma^2_{y/x}$ es: $s^2_{y/x} = \Sigma(y_i - \hat{y}_i)^2 / (n-2)$

Recordemos que la diferencia entre el valor observado de Y y el valor estimado a partir de la recta, para cada observación, constituye el residuo e_i, y representa el error que se comete al usar $\hat{Y}$ para estimar *Y*. Por dicha razón esta varianza se conoce también como varianza residual (s^2_e). Los grados de libertad son n-2 dado que se estimaron dos parámetros poblacionales, α y β.

Existen numerosas formas de calcularla:

$$S^2_{y/x} = s^2_e = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2} = \frac{\sum y_i^2 - a\sum y_i - b\sum xy}{n-2} = \frac{S_{yy} - (b^2 S_{xx})}{n-2}$$

Una forma de evaluar el buen ajuste al modelo lineal es comparar la dispersión de los puntos alrededor de la recta, que como ya se explicó constituye la varianza residual s^2_e, con la dispersión de los puntos alrededor de la media o varianza total s^2_y.

La varianza total está dada por: $$s^2{}_y = \frac{\sum (y_i - y)^2}{n-1}$$

Si el modelo lineal es bueno, es de esperar que la variación residual sea mucho menor que la variación total. Por ello se define el coeficiente de determinación de la regresión de *Y* sobre *X*, R^2, como

$$R^2 = 1 - \frac{s^2{}_e}{s^2{}_y}$$

y mide la proporción de la variación total que está explicado por la regresión de Y en X. En otras palabras mide la proximidad del ajuste de la ecuación de regresión de la muestra a los valores observados de Y.

El coeficiente de determinación toma valores entre 0 y 1, es decir $0 \leq R^2 \leq 1$. Se ve que cuando los residuos son pequeños, la s^2_e es pequeña. Esto lleva a un cociente s^2_e / s^2_y pequeño y por lo tanto a un R^2 grande.

Si $s^2_e = 0$ ➔ $R^2 = 1$ Si $s^2_e = s^2_y$ ➔ $R^2 = 0$

En la Fig 2 se pueden apreciar las distintas situaciones de la recta de regresión según el valor de R^2 (buen ajuste al modelo lineal implica R^2 cercano a 1; mal ajuste al modelo lineal implica R^2 cercano a 0.

La fórmula que se utiliza para el cálculo de R^2 es $$R^2 = \frac{S^2_{xy}}{S_{xx} S_{yy}}$$ siendo

$$S_{yy} = \sum y^2 - n\bar{y}^2$$

Un valor de R^2 cercano a 1 indica que la regresión ha tenido en consideración una gran proporción de la variabilidad total en los valores observados de Y, y por lo tanto la ecuación de regresión parece aceptable. Cuando R^2 es cercano a 0, lo cual indica una falla de la regresión para explicar una proporción de la variación total en los valores observados de Y, tiende a arrojar dudas sobre la utilidad de tal ecuación. En general, se considera que el modelo de regresión lineal es válido si el R^2 es de por lo menos 0.5. Sin embargo no se pasa al juicio final sin antes realizar alguna prueba estadística. Esta prueba se realiza por medio de una prueba de hipótesis para el coeficiente de regresión β. Las hipótesis que se plantean son:

Ho: $\beta = 0$ ➔ es decir que Y no depende linealmente de X; el modelo lineal no es válido

H_1: $\beta \neq 0$ ➔ es decir que Y sí depende linealmente de X; el modelo lineal es válido

b es el estimador insesgado de β, es decir que E(b) = β, mientras que el desvío estándar estimado de b (S_b) es:

$$S_b = \sqrt{\frac{S^2_e}{S_{xx}}}$$

y la distribución que se utiliza (dado que se desconoce el desvío poblacional) es la de Student, con n-2 grados de libertad.

Volviendo a la prueba de hipótesis, el estadístico de prueba es: $$t_{calc} = \frac{b - \beta}{S_b}$$

siendo la condición de rechazo: $t_{calc} < t_{crít\ 1}$ o $t_{calc} > t_{crít\ 2}$,

donde $t_{crít\ 1} = t_{n-2;\ \alpha/2}$ y $t_{crít\ 2} = t_{n-2;\ 1-\alpha/2}$

Otra forma de expresar la condición de rechazo es: $|t_{calc}| > t_{crít}$, donde $t_{crít} = t_{n-2;\ 1-\alpha/2}$
Si se rechaza Ho, se puede afirmar que existen evidencias suficientes de una dependencia lineal de Y sobre X, con un nivel de significación de α.

Volviendo al ejemplo de las ventas anuales en función de la antigüedad de los empleados, se desea determinar si el modelo lineal es válido. En primer lugar se calcula el coeficiente de determinación R^2:

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} = \frac{391^2}{142 x 1600.4} = 0.67$$

Es decir que el 67% de la variabilidad de las ventas anuales está explicada linealmente por la antigüedad de los vendedores. Por lo tanto puede decirse, como primera aproximación, que el modelo lineal es válido, ya que queda solo un 33% de variabilidad en las ventas que se debe a otras causas, distintas de la antigüedad.

La forma estadísticamente correcta de evaluar el modelo lineal es a través de la prueba de hipótesis para β:

Ho: β = 0 ➔ las ventas no dependen de la antigüedad del vendedor
H1: β ≠ 0 ➔ las ventas sí dependen de la antigüedad del vendedor

CR: $t_{calc} < t_{crít\ 1}$ o $t_{calc} > t_{crít\ 2}$, donde $t_{crít\ 1} = t_{n-2;\ \alpha/2}$ y $t_{crít\ 2} = t_{n-2;\ 1-\alpha/2}$

Otra forma: $|t_{calc}| > t_{crít}$, donde $t_{crít} = t_{n-2;\ 1-\alpha/2}$

En el ejemplo, para calcular t_{calc} es necesario calcular s^2_e y entonces s_b.

$$s_e^2 = \frac{S_{yy} - (b^2 S_{xx})}{n-2} = \frac{1600.4 - (2{,}754^2 x 142)}{8} = 65{,}42$$

$$S_b = \sqrt{\frac{S_e^2}{S_{xx}}} = \sqrt{\frac{65{,}42}{142}} = 0{,}679$$

$$t_{calc} = \frac{b - \beta}{S_b} = \frac{2.574 - 0}{0.679} = 4.06$$

siendo $t_{crít} = t_{8;\ 0.975} = 2.306$ (asumiendo $\alpha = 0.05$)
Por lo tanto se rechaza Ho y se puede afirmar, con un riesgo del 5%, que las ventas anuales dependen linealmente de los años de experiencia del vendedor; el modelo lineal es válido.

Usos de la ecuación de regresión
Si se demostró que el modelo de regresión lineal es válido, eso significa que la ecuación de regresión puede utilizarse para estimaciones y predicciones:

- Estimación puntual: la ecuación de la recta se aplica para calcular el valor de y para un dado valor de x.

- Estimación por intervalos de confianza para la media de y: se establecen los límites entre los cuales se estima que se encontrará con un cierto nivel de confianza el valor medio de y o E(y) para un determinado valor de x.
- Estimación por intervalos de predicción: se establecen los límites entre los cuales se estima que se encontrará con un cierto nivel de confianza un valor individual de y para un determinado valor de x.
- Estimación del coeficiente de regresión: se establecen los límites entre los cuales se estima que se encontrará con un cierto nivel de confianza el coeficiente de regresión (β).

Intervalo de confianza para el valor medio de Y:

$$\hat{y} \pm t_{GLerror;1-\alpha/2} S_e \sqrt{\frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{XX}}}$$

Intervalo de predicción para un valor individual de Y:

$$\hat{y} \pm t_{GLerror;1-\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{XX}}}$$

Para un determinado valor de x el intervalo de predicción será más grande que el de confianza, debido a la mayor incertidumbre.

Intervalo de confianza para el coeficiente de regresión:

$$b \pm t_{v;1-\alpha/2} S_b$$

Ejemplo:

a) Use la ecuación estimada de regresión para predecir puntualmente las ventas anuales de un vendedor con 9 años de experiencia.

Como se calculó anteriormente, la estimación puntual de las ventas es:
$\hat{y} = a + bx$ = 89,122 + 2,754 x 9 = 113,908 miles de $

b) Estime las ventas promedio anuales de todos los vendedores con 9 años de experiencia, con un nivel de confianza del 95%.

Dado que se desea estimar el valor promedio de Y, corresponde un intervalo de confianza del 95% para el promedio de Y:

$$\hat{y} \pm EM \qquad \text{donde } EM = t_{8,0.975} S_e \sqrt{\frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{XX}}} = 2.306\sqrt{65.42}\sqrt{\frac{1}{10} + \frac{(9-7)^2}{142}} = 6.687$$

reemplazando: 113,908 ± 6,687 = [107,221 ; 120,595]; con lo que resulta que las ventas anuales promedio para todos los vendedores con 9 años de experiencia se encuentran entre 107,221 y 120,595 miles de $ con una confianza del 95%.

c) Pedro Urdemales cumplirá el año próximo 9 años en las ventas. Pronostique las ventas que se esperan de él, con un nivel de confianza del 95%.

En este caso se trata de una estimación para un valor individual de Y, por lo que corresponde un intervalo de predicción.

$\hat{y} \pm EM$ donde

$$EM = t_{8,0.975} S_e \sqrt{1 + \frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{XX}}} = 2{,}306\sqrt{65.42}\sqrt{1 + \frac{1}{10} + \frac{(9-7)^2}{142}} = 19{,}809$$

reemplazando: 113,908 ± 18,809 = [94,100 ; 133,717]; con lo que resulta que las ventas anuales estimadas para Pedro Urdemales, un vendedor con 9 años de experiencia, se encuentran entre 94,100 y 133,717 miles de $ con una confianza del 95%.

d) Estime el incremento en las ventas anuales por cada año adicional de experiencia, con un nivel de confianza del 95%.

En este caso corresponde un intervalo de confianza para el coeficiente de regresión:

$b \pm t_{\nu;1-\alpha/2} S_b$

reemplazando: 2,754 ± 2,306 x 0,679 = [1,188 ; 4,320]; con lo que resulta que las ventas anuales se incrementarán entre 1,188 y 4,320 miles de $ por cada año adicional de experiencia en ventas, con una confianza del 95%.

En Excel: Herramientas > Análisis de datos > Regresión

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,8202
Coeficiente de determinación R^2	0,6727
R^2 ajustado	0,6318
Error típico	8,0915
Observaciones	10

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*
Intercepción	89,125	5,398	16,511	1,83E-07	76,677	101,573
Años de experiencia	2,754	0,679	4,055	0,004	1,188	4,319

Análisis de correlación

Este modelo se aplica cuando ambas variables X e Y son aleatorias y el objetivo es determinar el grado de relación lineal entre dichas variables en estudio; y se dirá si las

variables están o no linealmente correlacionadas. El parámetro que mide tal relación es el coeficiente de correlación (ρ), que se define como:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma^2{}_x * \sigma^2{}_y}}$$

donde σ_{xy} es la covarianza de X e Y (una medida de la variabilidad conjunta de ambas variables); σ^2_X es la varianza de X y σ^2_y es la varianza de Y.
El coeficiente de correlación toma valores entre -1 y 1, es decir $-1 \leq \rho \leq 1$

- ρ = ± 1 si y solo si las variables X e Y están linealmente relacionadas en forma perfecta, es decir vale y = a ± bx, con lo cual en realidad se está midiendo el grado de relación lineal.
- ρ = 0 indica que X e Y no están correlacionadas; luego la variables independientes nunca están asociadas. (Recordar que ρ = 0 es porque cov(x, y) =0).

El estimador puntual de ρ es r, el coeficiente de correlación muestral, y se calcula como:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

En el ejemplo, el coeficiente de correlación vale:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \sqrt{R^2} = 0.82$$

Es decir que existe un buen grado de asociación lineal entre ambas variables, siendo dicha asociación *directa* (al aumentar una variable, la otra también lo hace).

Ejercitación:

Problema 1: Para analizar la incidencia del cansancio sobre la eficiencia en el trabajo se tomó una muestra de 10 empleados, se los hizo trabajar distinta cantidad de horas, luego se les entregó un texto de varias carillas para tipear y finalmente se verificó la cantidad de errores cometidos por cada uno de ellos, obteniéndose los siguientes datos:

Empleado	1	2	3	4	5	6	7	8	9	10
Horas trabajadas	2	2	3	3	4	4	5	5	6	6
Errores cometidos	4	6	7	7	8	10	9	13	11	15

Datos útiles: $Sxx = 20$; $Syy = 100$, $Sxy = 40$; $Se^2 = 2.5$

a) Hallar la recta de mínimos cuadrados.
b) Calcular la variancia de la estimación.
c) Con un nivel de significación del 10%, ¿presentan los datos suficiente evidencia sobre la existencia de una relación lineal entre estas dos variables?
d) Hallar un intervalo de confianza del 90% para el coeficiente de regresión.

e) Estimar con un 90% de confianza la cantidad promedio de errores que cometerá un empleado que ha trabajado tres horas y media.
f) Hallar un intervalo del 90% para la cantidad de errores que cometerá un empleado que ha trabajado 3,5 horas.
g) Calcular el coeficiente de determinación e interpretar su resultado.
h) Indique los supuestos necesarios para la validez de los cálculos efectuados.
i) Compare los resultados obtenidos con la salida generada en Excel.

RESPUESTAS: a) $\hat{y} = 1 + 2x$ b) $S_e^2 = 2{,}50$

c) Sí. (porque 5,66 > 1,86) d) [1,34 ; 2,66]

e) Entre 7 y 9 errores. f) Entre 5 y 11 errores.

g) $R^2 = 0{,}80$ (el 80% de la variación total en la cantidad de errores queda explicada por la cantidad de horas trabajadas).

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,894
Coeficiente de determinación R^2	0,8
R^2 ajustado	0,775
Error típico	1,581
Observaciones	10

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*
Intercepción	1	1,500	0,667	0,524	-2,459	4,459
Horas trabajadas	2	0,354	5,657	0,000	1,185	2,815

Problema 2: Se está analizando la demanda de cierta gaseosa en la localidad A, y a tal efecto se ha tomado una muestra de 50 negocios minoristas que tienen dicho producto a la venta, registrando el precio al que ofrecen el litro (x) y la cantidad de litros demandada durante la semana anterior (y). Los resultados obtenidos se muestran a continuación:

$\sum x = 110$ $\sum x^2 = 1882{,}73$ $\sum y = 3530$ $\sum y^2 = 271268$ $\sum x \cdot y = 2569$

a) Determinar mediante un coeficiente adecuado el grado de relación lineal entre el precio y la cantidad demandada.
b) Estimar con un 95% de confianza la cantidad de litros de esta gaseosa que venderá semanalmente un negocio que la ofrezca a $2 el litro.

RESPUESTAS: a) r = -0,864 b) [49,32 ; 93,14]

Problema 3: Una aplicación importante del análisis de regresión en contabilidad es para estimar costos. Al reunir datos sobre volumen y costo y aplicar el método de mínimos cuadrados para determinar la ecuación de regresión donde se relacionan estas variables, un contador puede estimar el costo asociado con determinada operación de manufactura. Se obtuvo la siguiente muestra de volúmenes de producción y costo total para una operación de manufactura.

Volumen de producción (unidades)	Costo total ($)
400	5250
450	5184
550	5401
600	5892
700	6398
750	6840

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,956
Coeficiente de determinación R^2	0,913
R^2 ajustado	0,892
Error típico	222,380
Observaciones	6

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*
Intercepción	3114,880	427,370	7,288	0,002	1928,310	4301,450
Volumen de producción	4,718	0,726	6,495	0,003	2,701	6,734

Datos útiles: $Sxx = 93750$ $Syy = 2284287{,}5$ $Sxy = 442275$ $Se = 222{,}380$

a) Use estos datos para estimar una ecuación de regresión con la que se pueda predecir el costo total para determinado volumen de producción.
b) Interprete el significado de las componentes de la recta, en función del problema.
c) Calcule el coeficiente de determinación. ¿Qué porcentaje de la variación en el costo total puede explicar el volumen de producción?
d) El programa de producción de la empresa indica que el mes próximo se van a producir 500 unidades. ¿Cuál será el costo total estimado para esta operación?
e) Estime con una confianza del 90% el parámetro del punto anterior.
f) ¿Es correcto estimar el costo total cuando se producen 1000 unidades? Discuta.

RESPUESTAS: a) $\hat{y} = 3114{,}88 + 4{,}718x$

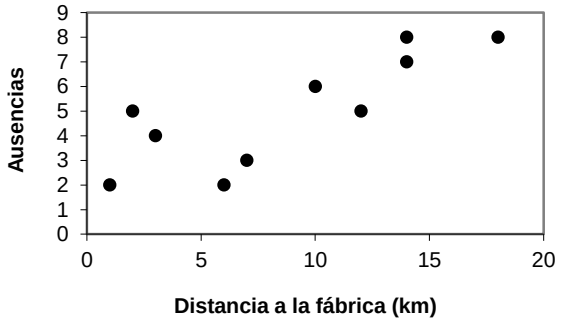
b) Si no se producen unidades, el costo total de manufactura será de 3114,88 $ (costo fijo), para este caso, al no ser observada la producción en 0 unidades estaríamos extrapolando en la interpretación. Mientras que por cada unidad adicional fabricada, el costo total se incrementará en 4,72$ (costo variable). c) 0.9134; 91.34% d) 5474 $
e) [4949 ; 5999 $] f) Es incorrecto extrapolar.

Problema 4: Dos directivos de una fábrica discuten acerca de la importancia de contratar operarios que vivan en las cercanías. El licenciado Gómez alude a los menores costos de transporte, menor tiempo de viaje y menores dificultades para concurrir en caso de problemas con los transportes. El licenciado Lavalle asegura que el domicilio del empleado no afecta los intereses de la compañía. Para dirimir la cuestión investigan la relación entre

la cantidad de días que faltan los empleados por año y la distancia (en km) de su hogar a su trabajo. Se eligió una muestra de 10 empleados y se reunieron los siguientes datos:

Empleado	1	2	3	4	5	6	7	8	9	10
Distancia a la fábrica	2	6	10	14	1	18	3	7	14	12
Ausencias anuales	5	2	6	8	2	8	4	3	7	5

La siguiente es la salida generada con Excel:



Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,8144
Coeficiente de determinación R^2	0,6632
R^2 ajustado	0,6211
Error típico	1,39165
Observaciones	10

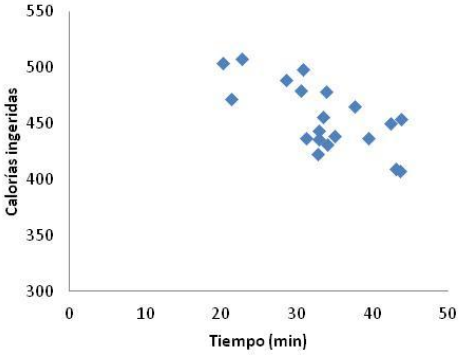
	Coefi-cientes	Error Típico	Estadístic o t	Proba-bilidad	Inferior 95%	Superior 95%
Intercepción	2,2354	0,8240	2,7130	0,0265	0,3353	4,1354
Variable X 1	0,3178	0,0801	3,9689	0,0041	0,1331	0,5024

Datos útiles: $Sxx = 302.1$; $Syy = 46$, $Sxy = 96$; $Se = 1,392$

a) Calcule e interprete la recta de mínimos cuadrados en función de las variables del problema.
b) A un nivel de significación del 5%, ¿cuál de los dos directivos tiene razón?
c) Utilice la ecuación estimada de regresión para estimar con una confianza del 95% la cantidad esperada de días de ausencia anuales para empleados que viven a 5 km de la fábrica.

RESPUESTAS: a) $\hat{y} = 2,24 + 0,32$. Si un empleado vive a 0 km de la fábrica, se estima que faltará 2,24 veces por año (nuevamente aquí extrapolamos en nuestra conclusión), mientras que por cada km que se aleje su domicilio de la fábrica, faltará a su trabajo 0,32 días.
b) El modelo lineal es válido, los datos le dan la razón al Lic. Gómez (t_{calc} =3,97>tcrít =2,31)
c) [2,598 ; 5,05]

Problema 5: El tiempo que los niños permanecen sentados en la mesa durante la comida, ¿puede ayudar a predecir cuánto comen? Se efectuó una investigación efectuada sobre 20 niños de tres años, que fueron observados durante varios meses en el jardín de infantes. Se registró el tiempo medio en el que cada niño permaneció en la mesa durante la comida, así como la cantidad media de calorías ingeridas, calculadas a partir de una detallada observación de lo que el niño comió cada día. El tiempo de permanencia en la mesa observado osciló entre 20,2 minutos y 43,7 minutos.
Para los datos obtenidos, se aplicó un modelo de regresión con los siguientes resultados:



Coeficiente de correlación	-0,70
Coef. de determinación	0,49
Error típico (Se, Sy/x)	22,03

	Coeficientes	*Error típico*
Ordenada al origen	556,67	24,85
Pendiente	-3,00	0,73 (Sb)

Sxx= 921,588

a) Escriba la ecuación de la recta estimada por mínimos cuadrados para predecir las calorías consumidas a partir de los tiempos en la mesa. Interprete los coeficientes. Describa brevemente lo que muestran los datos sobre el comportamiento de los chicos.
b) Determine si la regresión es significativa, con $\alpha = 0,05$. Interprete R^2 en términos del problema.
c) Calcule un intervalo de confianza del 95% para la verdadera pendiente de la recta de regresión.
d) Estime las calorías que espera ingieran niños de tres años que permanecen media hora en la mesa. ¿Podría usar la ecuación para predecir la ingesta de niños que permanecen 10 min en la mesa?

RESPUESTAS:

a) CALORÍAS = 556,67 – 3,00 TIEMPO. La ordenada no tiene interpretación, ya que ningún niño permaneció 0 min en la mesa, y por lo tanto constituye una extrapolación. Pendiente: Por cada minuto en la mesa, los niños ingieren, en promedio, 3 calorías menos.
b) Sí, ya que t_{calc} (-4,13) < $t_{crít}$ (-2,10). R^2: el 49% de la variabilidad en la cantidad de calorías consumidas se explica linealmente por los tiempos en la mesa.
c) [-4,525; -1,475]
d) 466,66 calorías; no es correcto efectuar la otra estimación ya que no se cuenta con datos en ese rango de tiempo.

Problema 6: Se realizó un ensayo a fin de encontrar una relación entre el porcentaje de rotura de granos de arroz durante el proceso de pelado respecto a las distintas

temperaturas a las que se sometió el grano en el proceso de secado. Se obtuvieron los siguientes resultados:

Temperatura (°C)	80	76	74	73	72	70	69
% rotura	33.9	29.9	31.4	25.8	24.2	19.7	21.3

Datos útiles: $\Sigma y = 186{,}2$ $\Sigma y^2 = 5122{,}24$ $\Sigma x = 514$ $\Sigma x^2 = 37826$
$\Sigma xy = 13782{,}5$ $Sxx = 83{,}71$ $Syy = 169{,}32$ $Sxy = 110{,}1$ $Se = 2{,}214$

a) Calcule, interprete y grafique la recta en función de las variables del problema.
b) Determine con un 10% de riesgo la significación de la regresión. ¿Qué implica?
c) ¿Cuál sería el porcentaje de granos de arroz rotos si la temperatura de secado a la que son sometidos es de 74°C? ¿Por qué no coincide exactamente con el valor obtenido durante el ensayo?
d) ¿Cuál sería el porcentaje de granos de arroz rotos si la temperatura de secado a la que son sometidos es de 45°C? ¿Qué opina del resultado?

RESPUESTAS: a) $\hat{y} = -69{,}972+1{,}315x$ b) El modelo lineal es válido (t_{calc} =5.43>2.015)
c) 27,35% d) –10.79%. Es un valor absurdo, ya que se extrapoló.

Problema 7: Si un test es confiable se espera que arroje puntuaciones más o menos similares cuando es evaluado en un mismo individuo, siempre y cuando el individuo no haya cambiado. Se midió la confiabilidad de una prueba para evaluar el razonamiento lógico. A tal efecto 8 adultos fueron sometidos a la prueba en dos oportunidades, separadas entre sí por 2 meses como mínimo. Los resultados fueron:

Individuo	Puntaje 1° vez	Puntaje 2° vez
1	9	7
2	4	4
3	7	6
4	8	9
5	7	5
6	6	7
7	5	5
8	8	7

a) Dibuje un diagrama de dispersión para estos datos. ¿Parece plausible la existencia de una relación lineal entre ellos?
b) Calcule el coeficiente de correlación y e interprételo en términos del ejercicio

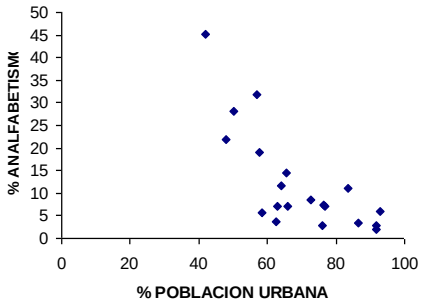
RESPUESTA: $r = 0.7308$; $t_{calc} = 2.62$; Se concluye que la prueba es confiable.

Problema 8: Los siguientes datos fueron extraídos del Anuario estadístico de América Latina y el Caribe 2006 editado por la Comisión Económica para América Latina y el Caribe (CEPAL) (http://www.eclac.org). Corresponden a el porcentaje de población urbana,

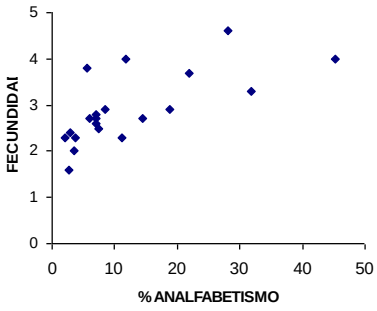
es decir residente en localidades de al menos 2500 habitantes, el % de analfabetismo en individuos mayores de 15 años y la tasa global de fecundidad, que mide el promedio de hijos por mujer, para los países de Latinoamérica, año 2006.

País	% población urbana	% Analfabetismo	Fecundidad
Argentina	91,8	2,8	2,4
Bolivia	64,2	11,7	4,0
Brasil	83,4	11,1	2,3
Chile	86,6	3,5	2,0
Colombia	76,6	7,1	2,6
Costa Rica	62,6	3,8	2,3
Cuba	76,1	2,7	1,6
Ecuador	62,8	7,0	2,8
El Salvador	57,8	18,9	2,9
Guatemala	50,0	28,2	4,6
Haití	41,8	45,2	4,0
Honduras	47,9	22,0	3,7
México	76,5	7,4	2,5
Nicaragua	56,9	31,9	3,3
Panamá	65,8	7,0	2,7
Paraguay	58,4	5,6	3,8
Perú	72,6	8,4	2,9
Rep. Dominicana	65,6	14,5	2,7
Uruguay	91,9	2,0	2,3
Venezuela	92,8	6,0	2,7

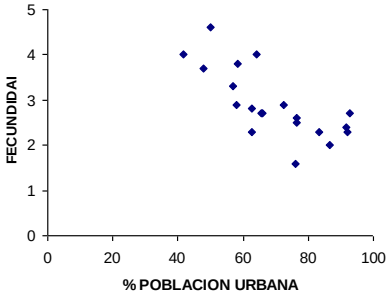
Estudiar la asociación entre las tres variables. Interpretar el valor de los coeficientes utilizando los gráficos de dispersión y los coeficientes de correlación generados en Excel.



r = -0.748



r = 0.689



r = -0.735

Revisión conceptual

Indique si las siguientes afirmaciones son verdaderas o falsas, justificando su respuesta

1) Si al calcular el coeficiente de correlación de dos variables X e Y, se tiene r=-0.20 ocurre que:
2) La pendiente de la recta de regresión es pequeña.
3) La pendiente de la recta de regresión es grande.
4) X e Y están poco relacionadas, aunque cuando X decrece, Y tiene tendencia a crecer.
5) El modelo lineal de regresión explica el 20% de la variabilidad de una variable cualquiera en función de la otra.
6) El modelo lineal de regresión explica el 80% de la variabilidad de una variable cualquiera en función de la otra.

2) Se observa que al disminuir el consumo de comida rápida, disminuye el nivel de colesterol en sangre. Se usa un modelo de regresión entre ambas que ofrece una bondad de ajuste del 36%. Entonces:

a) El 36% de las predicciones del modelo son correctas.
b) El 36% del nivel de colesterol se explica por el consumo de comida rápida
c) r= +0.60
d) r= +0.36
e) r= -0.60
f) r= -0.36

PRACTICO 7. SERIES DE TIEMPO

Consideraciones generales:

El estudio de las series de tiempo es fundamental en diversos campos de aplicación, ya que la mayoría de los fenómenos que se refieren a producción, comercialización, consumo, trabajo, son procesos que se desarrollan a lo largo del tiempo.

Como ejemplos de una serie de tiempo podemos mencionar las ventas de una empresa metalúrgica desde el año 2000, la producción anual de cierta materia prima de una empresa desde 2001 o la inscripción anual a una Universidad, entre otros.

El análisis de una serie de tiempo será de utilidad para que, por ejemplo, la administración de determinada empresa evalúe decisiones actuales y logre planificar con base en una predicción de largo plazo, ya que se sospecha, en general que los patrones pasados se extenderán en el futuro.

Las proyecciones a largo plazo son esenciales a fin de dar tiempo suficiente para que los departamentos de compras, manufacturas, ventas, finanzas y otros sectores de una empresa confeccionen planes para financiamiento, mejora e impulsos de nuevos productos

Conceptos generales:

Una serie de tiempo es un grupo de datos registrados en un determinado período: semanal, trimestral o anual que tiene por objetivo analizar el comportamiento de un atributo cuantitativo a través del tiempo.

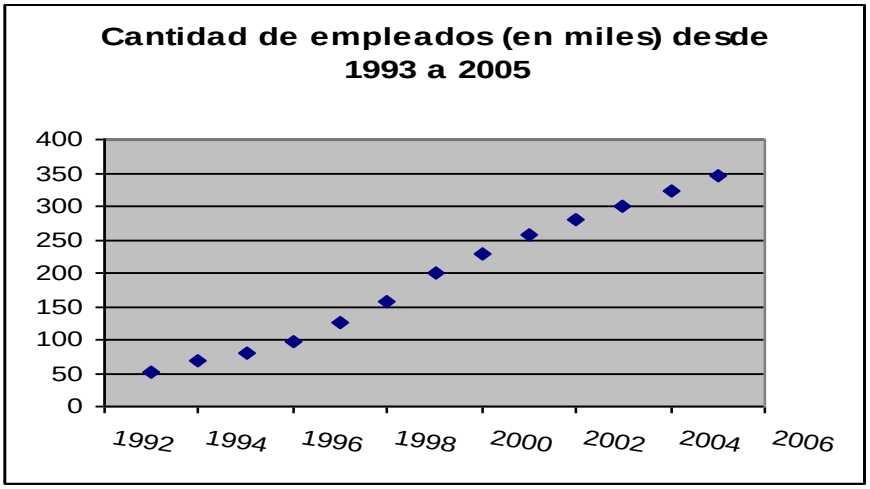
Para poder efectuar el análisis de una serie de tiempo, deberemos identificar los componentes que han provocado las fluctuaciones en los datos que se han observado.

Una serie de tiempo consta de cuatro componentes que son:

- ***Tendencia:*** Es la dirección uniforme de una serie de tiempo de largo plazo. Si a lo largo de un período prolongado de tiempo la serie marca una trayectoria ascendente, diremos que los datos muestran una tendencia positiva. Si el nivel de los datos va descendiendo con el tiempo, nos encontramos ante una tendencia negativa. Los datos se consideran estacionarios si no se advierte en ellos ninguna tendencia, ni positiva ni negativa.

Las tendencias de largo plazo de las ventas, el empleo, los precios y de otras series de mercados siguen variados patrones; muchas de ellas se mueven hacia arriba en forma uniforme, otras decaen y otras persisten iguales con el paso del tiempo.

Veamos en el próximo ejemplo un gráfico en el cual podemos observar cómo la cantidad de empleados de una empresa aumentó rápidamente en los últimos años; se observa que en 1993 había más de 50000 empleados y en 2005 aumentó a más de 340000.



Entonces la componente de tendencia, para el caso analizado, es ascendente.

- ***Variación cíclica***: Es el aumento y disminución de una serie de tiempo durante períodos mayores de un año.
- ***Movimiento estacional***: Son patrones de cambio en una serie de tiempo en un año. Estos patrones tienden a repetirse cada año.

En casi todos los negocios se observan patrones estacionales recurrentes, los casos en los cuales este patrón se presenta habitualmente es en aquellos comercios que se dedican a la venta de juguetes o a la venta de helados, por ejemplo, presentando fuertes picos durante diferentes periodos, como ser para el último caso, el trimestre de verano.

- ***Movimientos Irregulares***: Esta componente contiene las fluctuaciones que no forman parte de las otras tres componentes. Representa fluctuaciones aleatorias o provocadas por eventos casuales como: huelgas laborales, embargos, fallas en los equipos u otros hechos que perturban de manera favorable o desfavorable el valor de la variable que se considera.

Técnicas de suavización. Promedio móvil.

Muchas veces los directores o gerentes de una empresa deben efectuar previsiones de los stocks de muchos de sus productos. El directivo y/o gerente tal vez utilice alguna forma de suavización de series cronológicas con el objetivo de poder evaluar la tendencia de la misma.

Existen varias técnicas de suavizamiento como ser: el promedio móvil, el suavizamiento exponencial simple, el suavizamiento de Holt, el suavizamiento de Winters.

En nuestra asignatura nos dedicaremos especialmente a la técnica de promedio móvil (el lector interesado podrá consultar el resto de las técnicas mencionadas en la bibliografía recomendada)

Promedio móvil

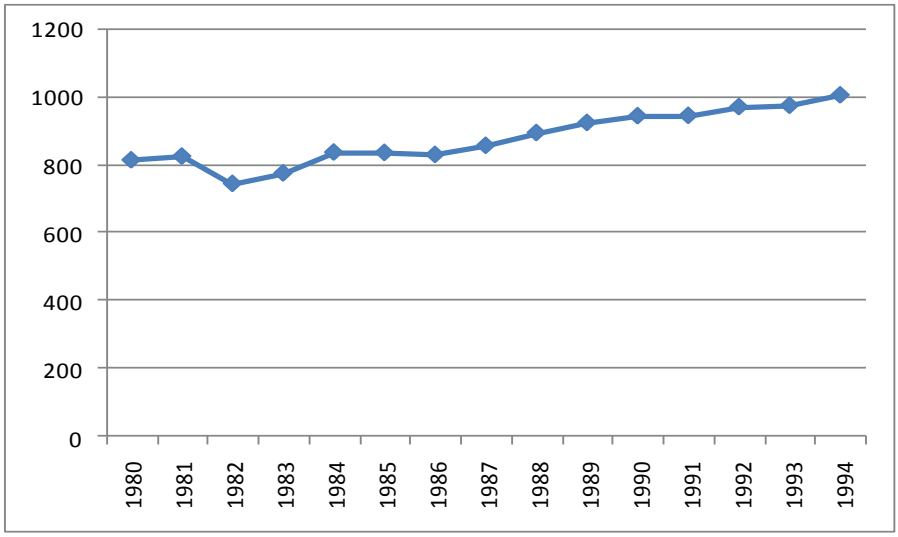
El método del promedio móvil suaviza las fluctuaciones de los datos, y esto se logra al desplazar los valores medios aritméticos en la serie de tiempo.

Un promedio móvil es útil para suavizar una serie de tiempo y poder así apreciar su tendencia. Es el método esencial para medir la fluctuación estacional.

Explicaremos la técnica con el análisis de una serie de tiempo para el consumo de energía eléctrica en una industria metalúrgica.

En la siguiente tabla se muestran los consumos de electricidad desde el año 1980 a 1984, y se observan que son muy variables por lo que utilizaremos un promedio móvil de tres años para suavizar estas fluctuaciones y luego uno de 5 años:

Año	Consumo de electricidad
1980	815
1981	826
1982	745
1983	776
1984	838
1985	837
1986	831
1987	858
1988	896
1989	926
1990	946
1991	947
1992	973
1993	977
1994	1008



Mostraremos como obtener los promedios móviles centrados de 3 años considerando si sólo tuviéramos los primeros años desde 1980 a 1984.

En primer lugar consideraremos al primer punto del promedio móvil centrado, que es el promedio de los tres primeros años donde 1981 está en el centro, por lo que se obtiene:

$$\frac{815+826+745}{3}=795,3$$

Luego para continuar, el segundo punto incluye el consumo desde 1981 a 1983, tomando como centro al año 1982, resultando:

$$\frac{826+745+776}{3}=782,3$$

De esta manera continuamos y armamos la siguiente tabla

Año	Consumo de electricidad	Promedio móvil centrado de 3 años
1980	815	
1981	826	795,333333
1982	745	782,333333
1983	776	786,333333
1984	838	

Observemos que no se calcularon valores para el año 1980 debido a que se encuentra en el centro del período 1979-1981 y no contamos con esos datos. Lo mismo sucede con el valor para el año 1984.

Si ahora continuamos completando la tabla con el procedimiento mencionado resulta la siguiente tabla:

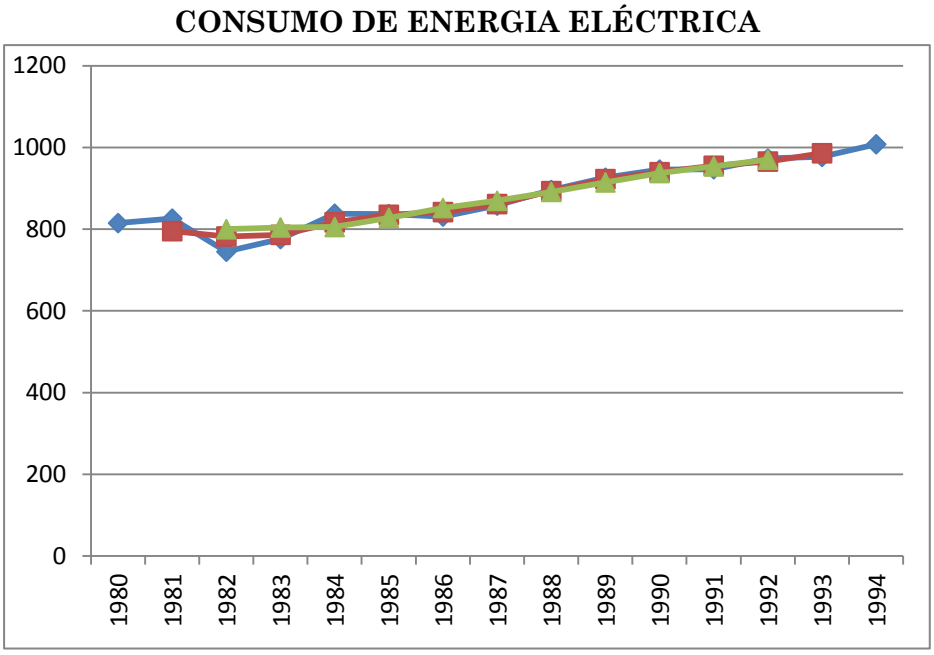
Año	Consumo de electricidad	Promedio móvil centrado de 3 años	Promedio Móvil centrado de 5 años
1980	815		
1981	826	795,333333	
1982	745	782,333333	800
1983	776	786,333333	804,4
1984	838	817	805,4
1985	837	835,333333	828
1986	831	842	852
1987	858	861,666667	869,6
1988	896	893,333333	891,4
1989	926	922,666667	914,6
1990	946	939,666667	937,6
1991	947	955,333333	953,8
1992	973	965,666667	970,2
1993	977	986	
1994	1008		

Podrán observar que también hemos incluido el cálculo del promedio móvil centrado de 5 años. Por ejemplo, para el primer valor de promedio móvil centrado en 5 años se realizó:

$$\frac{815+826+745+776+838}{5}=800$$

Análogamente pueden obtener el resto de los valores.

Gráficamente pueden observarse las diferentes líneas de tendencia utilizando el método de los promedios móviles. La línea de tendencia de color azul muestra los datos originales, en rojo aparece la gráfica correspondiente al promedio móvil de 3 años y en verde al gráfico correspondiente al promedio móvil centrado de cinco años.



Si observamos el comportamiento de las líneas de tendencia puede verse que los dos promedios móviles, centrados en el tercer y quinto año, amortiguan las fluctuaciones de la serie original, pero la curva del promedio móvil de cinco años (color verde) está un poco más suavizada que la del promedio de tres años. **La inclusión de más períodos en el promedio móvil "amortigua" aún más las fluctuaciones originales.** Esto se debe a que una base más grande reduce el impacto de cualquier punto singular de los datos.

En Excel podemos obtener el promedio móvil, ingresando dentro de la opción de datos, análisis de datos y seleccionando la opción de **media móvi**l. Al desplegarse el cuadro de diálogo de la media móvil, les pedirá el rango de entrada; allí deberán indicar sólo la columna correspondiente a los valores de la variable analizada. Una de las opciones disponibles es la de **intervalos** que pueden utilizarla para incorporar el lapso para el cual desean calcular el promedio, es decir, si será un promedio de 3 años, de 5 años o algún otro periodo. Vale aclarar que si no se indica el periodo en cuestión la planilla de cálculo asume que el promedio móvil se calculará por un periodo de tres años.

Componente de tendencia. Método lineal.

La tendencia de largo plazo de muchas series de negocios, como ventas, envíos y producción, con frecuencia se aproximan a una recta. En este caso la **ecuación de tendencia lineal** para describir este crecimiento es:

$$\hat{Y} = a + b\,t$$

Donde

- $\hat{Y}$ es el valor proyectado de la variable para un valor seleccionado de t.
- a es la intersección con el eje y. Es el valor estimado de Y cuando $t = 0$.
- b es la pendiente de la recta, o cambio promedio de $\hat{Y}$ por cada aumento de una unidad en t
- t es un valor de tiempo seleccionado.

En los métodos de proyección, el tiempo es la variable independiente y el valor de la serie de tiempo. Con frecuencia se codifica la variable independiente para facilitar la interpretación, es decir se designa con $t = 1$ para el primer año, $t = 2$ para el segundo año y así sucesivamente.

Para obtener el valor de la pendiente de la recta utilizaremos la siguiente fórmula, que se obtiene utilizando el método de los mínimos cuadrados[2]:

$$b = \frac{\left(\sum t\,y\right) - n\overline{t}\;\overline{y}}{\left(\sum t^2\right) - n\overline{t}^{\,2}}$$

Para obtener el valor de a:

$$a = \overline{y} - b\,\overline{t}$$

Ejemplo

Se muestran a continuación las ventas de una importante cadena de supermercados desde el año 2002, registrada en millones de pesos.

Año	Ventas (millones de pesos)
2002	7
2003	10
2004	9
2005	11
2006	13

Se pide: Determinar la ecuación de tendencia lineal, el incremento anual de las ventas y la proyección de las ventas para 2009

[2] Puede consultarse la bibliografía recomendada para profundizar los conceptos acerca de la obtención de los coeficientes a y b, mediante el método de los mínimos cuadrados.

Procedimiento

Para determinar la ecuación de tendencia lineal utilizaremos la fórmula vista previamente, por lo que debemos en primer lugar hallar los valores de a y de b. Codificamos los valores de t y aplicamos las fórmulas correspondientes:

Año	t	y (Ventas)	t cuadrado	t.y
2002	1	7	1	7
2003	2	10	4	20
2004	3	9	9	27
2005	4	11	16	44
2006	5	13	25	65

De manera que obtenemos sumando los valores de las columnas construidas, lo siguiente:

$$\sum ty = 163 \quad \sum t^2 = 55 \quad \bar{t} = 3 \quad \bar{y} = 10$$

$$b = \frac{163 - 5.(30)}{55 - 5(3)^2} = 1,3 \qquad a = 10 - 1,3.(3) = 6,1$$

Por lo tanto la ecuación de tendencia lineal resulta:

$$\hat{Y} = 6,1 + 1,3\,t$$

¿Cómo se interpreta esta ecuación? El valor de la pendiente 1,3 indica que las ventas aumentaron con una tasa de 1,3 millones de pesos por año. En términos generales el coeficiente b mide la variación que se obtiene en el valor de la variable por cada año adicional analizado. El valor 6,1 es el valor estimado de las ventas en el año 0, es decir el estimado para 2001.

Podemos usar esta ecuación de tendencia lineal obtenida para estimar valores futuros. Por ejemplo si queremos estimar el valor para el año 2009, sabemos que de acuerdo a la codificación efectuada corresponde a $t = 8$, entonces remplazando en la ecuación resulta:

$$\hat{Y} = 6,1 + 1,3.(8)$$

$$\hat{Y} = 16,5$$

De manera que el estimado para el año 2009 es de 16,5 millones de pesos.

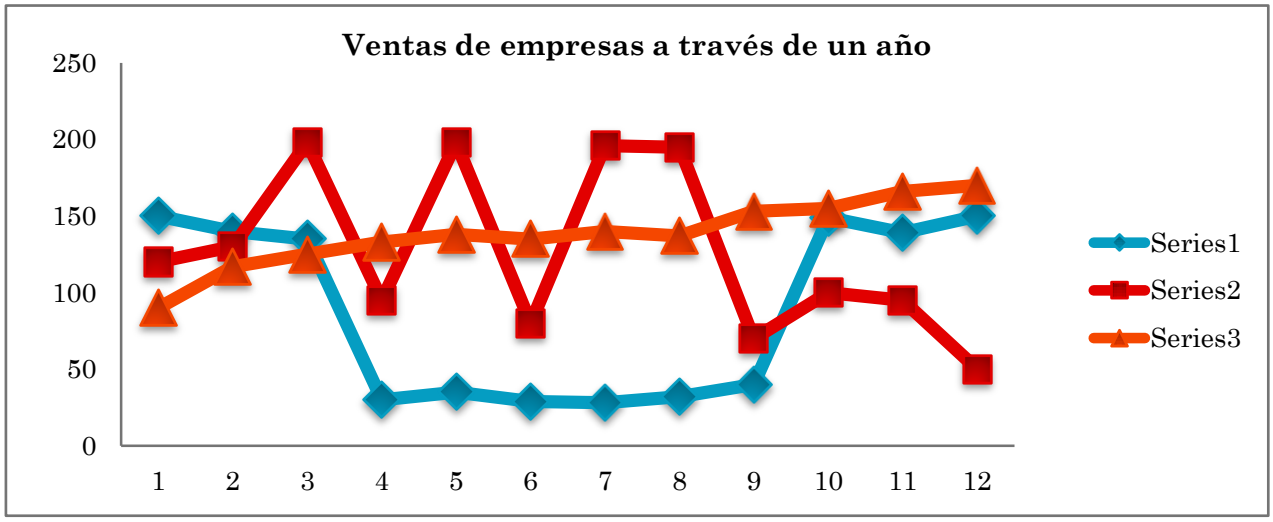
En este ejemplo de serie de tiempo, había 5 años de datos de ventas. Con base en estas cinco cifras de ventas, se estimaron las ventas de 2009, sin embargo hay que tener en cuenta que no es aconsejable proyectar ventas, producciones u otras series de negocios o económicas más que $n/2$ períodos a futuro, donde n es el número de puntos de datos. Por ejemplo si hay 10 datos sólo se estiman hasta 5 años a futuro. Para el ejemplo tenemos 5 puntos de datos que dividido dos nos da 2,5, es decir que no es aconsejable proyectar a mas de dos años y medio del último años evaluado.

En Excel podemos obtener la ecuación lineal, ingresando dentro de la opción de datos, análisis de datos y seleccionando la opción de **Regresión**. Al desplegarse el cuadro de diálogo de Regresión, les pedirá el rango de entrada; allí deberán indicar las columnas correspondientes a los valores de ambas variables que siempre deberán estar expresadas en columnas contiguas.

Cabe destacar que, si bien se ha visto el ajuste a una tendencia lineal, las variables analizadas podrán poseer un mejor ajuste a otro tipos de modelos como el exponencial, logarítmico, etc. Este fenómeno dependerá de la variable en cuestión. Debido a que dentro de esta materia sólo se estudia el ajuste lineal, se aconseja al lector que, ante situaciones en las cuales el ajuste lineal no sea el más adecuado, profundizar este tema con la bibliografía recomendada.

Ejercitación:

Problema 1: En el siguiente gráfico se muestran las ventas mensuales (en miles de unidades) de empresas, de diversos ramos, a lo largo de un año. Determinar cuál es la componente que afecta a cada una de las series involucradas. Justifique su respuesta.



Problema 2: Una compañía global basada en ciencia para el cuidado de la salud y pionera en biotecnología vende sus productos terapéuticos En la siguiente tabla aparecen las ventas netas de 1997 a 2004. Las ventas netas se dan en millones de dólares

Año	1997	1998	1999	2000	2001	2002	2003	2004
Ventas Netas	6714	7991	9075	9775	9762	10180	8334	8272

Ajuste una ecuación de tendencia lineal ¿Cuáles son las ventas estimadas para 2005?

Respuestas: $\hat{y} = 7909,86 + 189,56\,t$. El estimado para 2005 es de 9615,89

Problema 3: Las cantidades de vidrio de desecho producido por una empresa especializada en cristalería de laboratorio y tubos de cristal se muestran a continuación

Año	2002	2003	2004	2005	2006
Desecho (Toneladas)	2	4	3	5	6

Ajuste una ecuación de tendencia lineal y estimar la cantidad de desecho para 2008.

Respuestas: $\hat{y} = 1,30 + 0,90\,t$. El estimado para 2008 es de 7,6 toneladas

Problema 4: Determinar un promedio móvil de tres años para las ventas de una empresa dedicada al diseño de piezas originales para diferentes tipos de máquinas.

Año	Número producido en miles
2000	2
2001	6
2002	4
2003	5
2004	3
2005	10

Respuestas:

Año	Total móvil de tres años	Promedio móvil de tres años
2000		
2001	12	4
2002	15	5
2003	12	4
2004	18	6
2005		

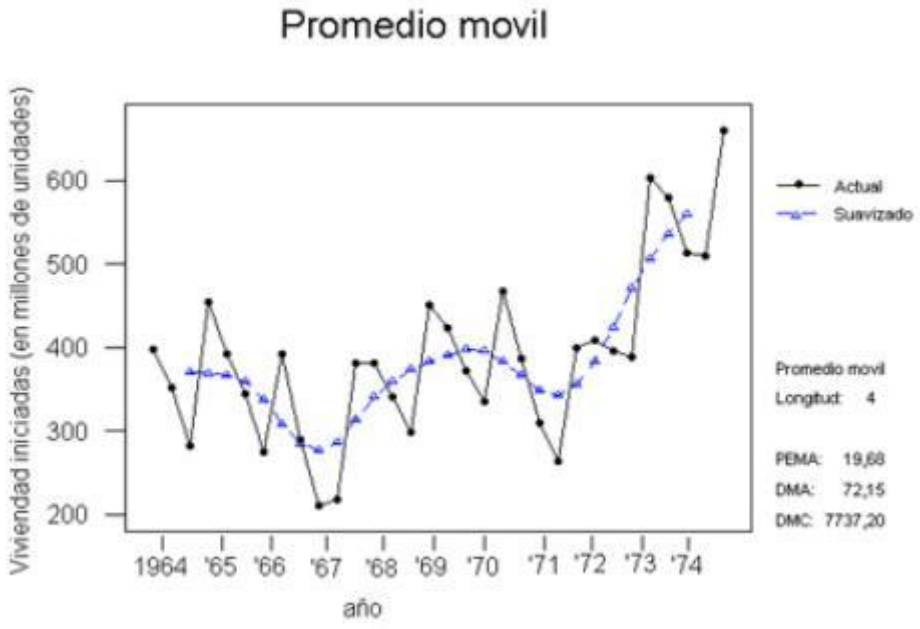
Problema 5: Calcular un promedio móvil ponderado en cuatro trimestres para el número de suscriptores de determinada revista técnica durante los nueve trimestres que abarcan los datos. Éstos se reportan en miles. Aplicar ponderaciones de 0,1-0,2-0,3-0,4

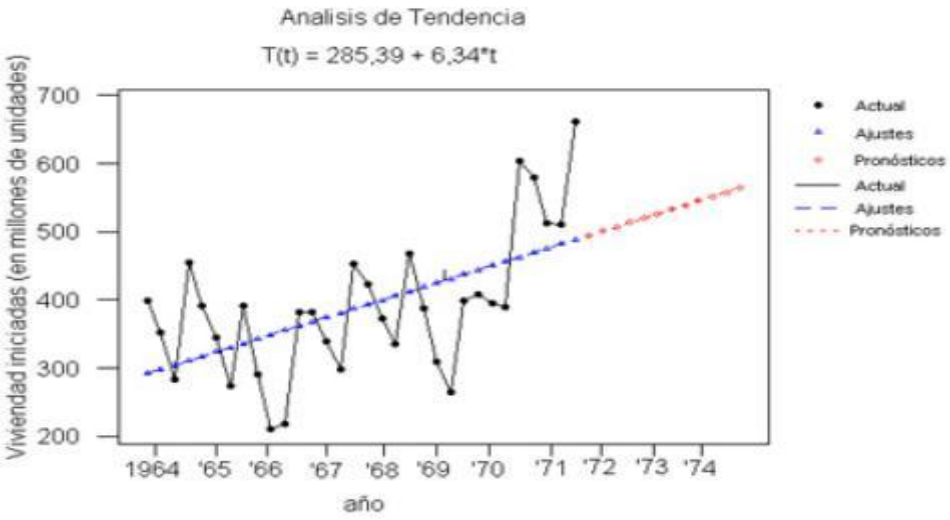
respectivamente, a los trimestres, es decir describir la tendencia en el número de suscriptores

31-Mar-04	28766
30-Jun-04	30057
30-Sep-04	31336
30-Dic-04	33240
31-Mar-05	34610
30-Jun-05	35102
30-Sep-05	35308
31-Dic-05	35203
31-Mar-06	34386

Respuestas: Los promedios móviles ponderados son: 31584,8 33088,9 34205,4 34899,8 35155,0 34887,1

Problema 6: En los siguientes gráficos se presenta una serie de tiempos para la cantidad de viviendas que se han iniciado durante el periodo de 1964 a 1975. En uno de los gráficos se ha aplicado la técnica de los promedios móviles para periodos de 4 años y en el otro se ha aplicado una tendencia lineal. A su parecer, ¿Cuál de los dos métodos representa más fehacientemente la serie analizada y por qué? Discuta.




Analisis de Tendencia
T(t) = 285,39 + 6,34*t
Vivienda iniciadas (en millones de unidades)
700
600
500
400
300
200
1964
'65
'66
'67
'68
'69
'70
'71
'72
'73
'74
año
Actual
Ajustes
Pronósticos
Actual
Ajustes
Pronósticos

EJERCICIOS INTEGRADORES

Problema 1: Una de las claves del éxito en una empresa es que los trabajadores estén alineados con la política de la misma. Se realizó un estudio para indagar si existe una vinculación entre el apoyo por parte de los empleados a las decisiones tomadas por la empresa con su participación en la toma de decisiones. Se entrevistó a 200 trabajadores, clasificándolos según su acuerdo con las decisiones tomadas por la empresa y su participación o no en esas decisiones:

Toma de decisiones	Participa	No participa
Aprueban las decisiones	73	51
No aprueban las decisiones	27	49

a) ¿Proporcionan los datos evidencia suficiente para indicar que la aprobación o no de decisiones de la empresa depende de si los trabajadores participan en la toma de decisiones? Utilice un riesgo de 0,05.
b) De los que participan, se cree que más del 25% no aprueban las decisiones, en cuyo caso la empresa debería comunicar las razones de tales decisiones en forma más detallada. ¿Se comprueba la sospecha? ¿Qué recomendaría con un 10% de riesgo de equivocarse?

RESPUESTAS:

a) Como χ^2calc (10,272) > χ^2crít (3,841), se concluye con un riesgo de 0.05 de que existen evidencias de que la aprobación depende de la participación en la toma de decisiones. Obsérvese que el 73% de los que participan aprueban las decisiones contra solo un 51% en el caso de los que no participan, sugiriendo las ventajas de incluir a los empleados en la toma de decisiones. Alternativamente podría haberse resuelto el ejercicio mediante una comparación de 2 p, siendo p_1 = proporción de empleados que aprueban las decisiones dentro de los que participan en la toma de las mismas (prueba unilateral derecha)
b) Como $\hat{p}calc$ (0,27) < $\hat{p}crít$ (0,306), no rechazo Ho. Por lo tanto no se recomienda comunicar las decisiones en forma más detallada.

Problema 2: Una empresa que produce galletitas desea controlar el funcionamiento de una máquina empaquetadora que, en condiciones normales, opera con un peso medio de los paquetes de 250g. Se tomó una muestra de 20 paquetes obteniéndose un peso medio de 280 g. y una desviación estándar de 15 g., estableciéndose en un 1% la probabilidad máxima de detener y revisar innecesariamente la máquina. Luego de evaluar los resultados de la muestra se decide detener la máquina y revisarla.

a) ¿Considera Ud. que la decisión tomada es correcta? Justifique la respuesta.
b) Estime con 99% de confianza el peso medio de los paquetes.
c) ¿Cómo disminuiría el error de la estimación anterior en un 30%?
d) En una segunda máquina que también produce las galletitas en esa empresa se tomó una muestra de 20 paquetes obteniéndose un peso medio de 255 g. y una

desviación estándar de 12 g. Con un nivel de significación del 1% ¿podría concluir que existe una diferencia significativa en los pesos promedio de las dos máquinas?

RESPUESTAS:

a) Sí, ya que $\bar{x}$ (280g) > $\bar{x}crít$ (259,596).
b) [270,404 ; 289,596]
c) Podría incrementarse el tamaño de la muestra a 31 paquetes, es decir incorporando 11 paquetes más a los 20 paquetes iniciales.
d) Como tcalc (5,821) > $tcrit_2$ (2,712), existe diferencia en los pesos promedio de los paquetes fabricados por ambas máquinas.

Problema 3: En la sucursal de un banco se toma una muestra de 250 operaciones y se observa que en ellas hubo 15 reclamos.

a) Estime el porcentaje de reclamos en la sucursal con un nivel riesgo de 0,05.
b) Calcule el nivel de confianza que podría asignar a una estimación I.C. = (0,035; 0,085), realizada sobre la base de la muestra mencionada.

RESPUESTAS:

a) [3,53 ; 8,47%]
b) 0,95254

Problema 4: Una empresa de copiado está probando dos plotters, marcas XEROX y HP. Se hicieron 50 copias de prueba de 15 tiradas cada una con ambos plotters. Con el Xerox ha obtenido un rendimiento medio de 3,2 ppm (ppm: páginas tamaño A3 por minuto) con un desvío de 0,6 ppm mientras que con el HP ha obtenido un rendimiento medio de 2,8 ppm con un desvío de 0,4 ppm. Los rendimientos, se ha probado, se distribuyen normalmente y no se ha demostrado previamente si la variabilidad en el rendimiento de ambas máquinas es igual o no.

a) Verifique al 5% si es que existen diferencias significativas de rendimiento promedio entre ambas marcas.
b) De existir diferencias ¿qué marca de plotter recomendaría? Estime con una confianza del 95% cuánto más veloz es en promedio la marca que recomendada con respecto a la otra.
c) Estime la velocidad promedio de la marca recomendada con una confianza del 95%.

RESPUESTAS:

a) Las varianzas en el rendimiento no difieren, ya que Fm (2,25) no es mayor que $Fcrit_2$ (2,98), por lo que se efectúa una prueba t para varianzas iguales. Se concluye que existen diferencias significativas de rendimiento entre ambas marcas, con un riesgo del 5% (2,15 > 2,05).
b) Recomendaría la marca Xerox, ya que su rendimiento promedio en la prueba fue mayor. Con una confianza del 95% se estima que el plotter marca Xerox produce en promedio entre 0.02 y 0.78 ppm más que el de marca HP.
c) Se estima que el rendimiento promedio del plotter marca Xerox se encuentra entre 2,89 y 3.53 ppm con una confianza del 95%.

Problema 6: Un club cuenta con la siguiente información obtenida de una muestra de socios:

Cantidad de socios por deporte

	Voley	Tenis	Natación
Varón	45	125	30
Mujer	100	150	50

Edades por deporte

	Cantidad	Promedio	Desvío
Voley	15	23	2
Tenis	25	40	5
Natación	20	25	4

De los socios que asisten a las clases de los dos profesores de natación se eligieron al azar 8 y se les pidió que calificaran de 1 a 10 las clases de ambos profesores. Se obtuvo:

Socio	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Prof Julio	9	8	10	8	7	9	5	8
Prof. Martín	7	5	7	9	8	6	6	8

En cuanto a la relación entre la cantidad de socios nuevos por mes y el valor de la cuota (en pesos) de 9 meses se tiene el siguiente registro:

Mes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Socios	100	90	70	75	70	60	58	55	40
Cuota	40	45	50	50	60	80	85	85	100

Para todas las preguntas usar riesgo 10%.

a) Determine si la elección de un deporte depende del género.
b) Verifique si la edad promedio de las personas que practican tenis es mayor que la de las personas que hacen natación
c) Estime la diferencia media de las edades de los socios que practican tenis y natación.
d) Determine un modelo lineal que relacione la cantidad de socios nuevos con el valor de la cuota e interprete el coeficiente de regresión.
e) Valide el modelo.
f) Estime la cantidad media de socios nuevos que se espera tener para una cuota de 90$.
g) Estime la variación en la cantidad de socios por cada peso que aumenta la cuota.

Estadísticas de la regresión	
Coeficiente de determinación R^2	0,9328
Error típico	7,0489
Observaciones	9

Sxx=3738.9

$\bar{x} = 66.11$

	Coeficientes	*Error típico*	*Estadístico t*
Intercepción	120,8581		
Variable X 1	-0,7824	0,1153	-6,85

RESPUETAS:

a) Como X^2calc = 8,47 > X^2crit= X^2(2;0,90)=4,605 Rechazo H0 el deporte no es independiente del género
b) Como F calc= 1,5625 < Fcrit= F(24;19;0,95)= 2,114 no Rechazo H0 Las variancias no son diferentes
   Como tcalc=10,89 > tcrit= t((43;0,90)=1,302 Rechazo H0 las personas que practicas tenis son más que las que practican natación
c) 12,869 ≤ μt - μn ≤ 17,32
d) $\hat{y} = 120{,}85 - 0{,}789x$. Por cada peso que aumente la cuota hay 0,789 socios menos
e) Como t calc= 6,81 < tcrit= t (7;0,95)= -1,895 Rechazo H0 es válido el modelo
f) 42,85 ≤ μ y/x= 90 ≤ 56,67
g) -1,00982 ≤ ß ≤ -0,57018