CUADERNO DE EJERCICIOS DEL CURSO DE ESTADISTICA II

DEPARTAMENTO ACADEMICO DE ESTADISTICA

ITAM

ÍNDICE

Tema 1	Introducción	1
Tema 2	Distribuciones muestrales y Teorema Central del Límite	
	Introducción	6
	Construcción de distribuciones muestrales	8
	Distribución de muestreo de $\overline{X}, \hat{p}, S^2, \overline{X} + \overline{Y}, \overline{X} - \overline{Y}, \hat{p}_1 - \hat{p}_2, S_1^2/S_2^2$	14
	Uso de tablas	27
Tema 3	Propiedades de estimadores	31
	Estimación puntual	38
Tema 4	Intervalos de confianza	41
	Determinación de tamaño de muestra	49
Tema 5	Pruebas de hipótesis paramétricas	
	Teoría	51
	Problemas	53
Tema 6	Pruebas no paramétricas	64
Tema 7	Muestreo	
	Teoría	76
	Problemas	78
Tema 8	Ejercicios con bases de datos	82
	Respuestas	92
	ANEXO – EXÁMENES FINALES ANTERIORES	

ESTADÍSTICA II

INTRODUCCIÓN

1.1. ¿Cuáles son las ventajas y las desventajas de realizar un muestreo con respecto al levantamiento de un censo a la población entera.?

1.2. ¿Qué es población estadística?

1.3. ¿Cuál es la diferencia entre un parámetro y un estadístico?

1.4. Diga si cada una de las siguientes afirmaciones es de tipo inferencial o descriptivo:

a) Uno de los cuatro candidatos a gobernar un Estado predijo que sería un ganador al momento de conocer los resultados de, 15 de los 100 distritos electorales.
b) El valor promedio diario del índice metropolitano de contaminación ha sido de 115 en lo que va de este año.
c) En el periodo de 1954 a 1986, los salarios mínimos en el D.F. se incrementaron en 30 ocasiones. El incremento promedio fue de 25.8%.
d) Si la demanda de tarjetas BANAMEX-ITAM sigue comportándose como hasta ahora, se espera que al final del año se expidan 3000 tarjetas.

1.5. Diga si cada una de las siguientes afirmaciones es de tipo inferencial o descriptivo:

a) La Bolsa Mexicana de Valores tuvo un incremento promedio del 5.3% durante la tercera semana de enero de este año.
b) De acuerdo a los acontecimientos ocurridos en Bosnia-Herzegovina, varios analistas predicen una nueva conformación de poderes en Europa oriental.
c) El Índice Nacional de Precios al consumidor en la ciudad de México creció 1.7% en agosto de 1990, con lo que la inflación acumulada en los primeros ocho meses de ese año llegó a 19.3%.

d) En una encuesta sobre el Sistema de Transporte Público en el D.F. se eligió una muestra de 600 personas, de las cuales 307 fueron mujeres y 293 hombres. El 52.4% de los encuestados señaló que los servicios de transporte son regularmente eficientes y el 41.7% afirmó que tarda dos horas o más en llegar de su casa a su destino regular.

1.6. Describa un problema, relacionado con su carrera, cuya solución dependa de la aplicación de la metodología estadística. Para ello, especifique la(s) variable(s) de interés, la población bajo estudio, el(los) parámetros, objetivo(s), el marco muestral y el(los) estimador(es) correspondiente(s).

1.7. Responda de manera concisa y clara a las siguientes preguntas:

a) ¿Qué es un estadístico y para qué sirve?
b) ¿Cuál es el problema que pretende resolver la inferencia estadística?
c) ¿Qué es una muestra aleatoria?

1.8. Explique la diferencia entre un estadístico, un parámetro, un estimador y una estimación.

1.9. Diga si cada una de las siguientes afirmaciones es de tipo inferencial o no. Justifique su respuesta y establezca, en cada caso, la población y la(s) variable(s) de interés:

a) Se realizó una encuesta a nivel nacional y se entrevistaron a 1614 personas en 55 localidades del país. Una de las preguntas fue: ¿simpatiza usted con algún partido político?. Las respuestas fueron PRI: 28%, PAN: 6%, PRD: 6%, OTROS: 4%, ninguno: 56%. Y se concluyó, que más de la mitad de las personas del país no simpatizan con algún partido.
b) Un dirigente de la oposición afirma que, de acuerdo a lo observado en el pasado, el índice de abstencionismo para futuras votaciones seguirá disminuyendo.

c) Al preguntarle a cada uno de los alumnos que cursan Estadística II con qué probabilidad pensaban aprobar el curso, resultó que el 90% de los alumnos asignaron una probabilidad mayor a 0.85.

d) De acuerdo a lo afirmado en el inciso anterior, la mayoría de los alumnos que toman un curso relacionado con Estadística se sienten relativamente seguros de aprobar.

1.10. Una firma de mercadotecnia envía un cuestionario a 1000 residentes de una colonia para determinar sus preferencias como compradores. De los 1000 residentes, 80 responden al cuestionario. ¿Constituye la anterior una muestra aleatoria?

1.11. Suponga que usted forma parte de un equipo interdisciplinario al que se le encargó realizar un estudio para pronosticar el resultado de unas votaciones locales que se llevarán a cabo este año. Si el reporte del estudio debe entregarse antes de julio de 1997, elabore una propuesta en la que se describan los siguientes elementos:

i) Población Objetivo
ii) Variables a considerar
iii) Parámetros de Interés
iv) Marco de muestreo
v) Tipo de resultados que se obtendrían
vi) Alcances y limitaciones del estudio.

1.12. Imagine que, con el propósito de evaluar el funcionamiento de una universidad privada, un comité asesor recomienda la realización de un estudio con el cual se pueda conocer el tiempo que tardan los egresados en encontrar un puesto de trabajo remunerado en su área profesional, a partir del momento en que concluyen sus estudios. Suponga que usted forma parte del equipo que debe realizar el estudio. Elabore una propuesta que incluya una breve descripción de:

i) Población bajo estudio.
ii) Variables a considerar.
iii) Parámetros de Interés.

iv) Marco de muestreo.
v) Tipo de resultados que se obtendrían.
vi) Alcances y limitaciones del estudio.

1.13. ¿Qué importancia tiene tomar una muestra aleatoria de una población en el proceso de inferencia estadística?

1.14. La empresa automotriz Chrysler, para efectos de control de calidad, revisa uno de cada veinte vehículos que se van produciendo.

a) ¿Es este procedimiento un muestreo aleatorio?. (Justifique su respuesta).
b) ¿Es un muestreo aleatorio simple? ¿Por qué?
c) ¿Qué probabilidad tiene cada vehículo de ser incluido en la muestra?
d) Mencione algún parámetro con el que se pueda conocer alguna característica de la población.

1.15. Una firma industrial, dedicada a la producción de artículos metálicos para la industria automovilística, produce tornillos de 5 mm. de diámetro. Como parte del proceso de control de calidad, se selecciona cierto número de tornillos de la producción diaria y se mide cada uno de éstos.

i. Los tornillos seleccionados para el estudio constituyen:
a) la población.
b) el estimador.
c) el parámetro.
d) la muestra.

ii. Los diámetros de estos tornillos constituyen:
a) el estimador.
b) la variable de interés.
c) el parámetro.
d) la muestra.

iii. La medida del diámetro de cada tornillo constituye:

a) un dato.
b) la muestra.
c) un estimador.
d) la población.

iv. El diámetro promedio de los tornillos de la muestra constituye:

a) un parámetro.
b) una estimación.
c) una variable.
d) un dato.

v.Del resultado obtenido de la muestra se desea generalizar a:

a) el dato.
b) la variable.
c) el estimador.
d) la población.

vi. El diámetro medio de todos los tornillos que se producen en un día constituye:

a) el parámetro.
b) la variable.
c) el dato.
d) la población.

DISTRIBUCIONES MUESTRALES

Y EL TEOREMA CENTRAL DEL LIMITE

INTRODUCCION

2.1.1. Diga si las siguientes aseveraciones son falsas o verdaderas y justifique su respuesta

a) En el proceso de inferencia estadística se selecciona una muestra aleatoria para garantizar que ésta sea representativa de la población.
b) Los métodos de muestreo probabilístico que se utilizan para seleccionar una muestra de una población permiten medir la confiabilidad de las inferencias.
c) Los datos de dos muestras aleatorias, de igual tamaño, tomadas de la misma población, no son necesariamente iguales.
d) La dispersión de la distribución de muestreo de la media muestral es siempre menor que la varianza de los elementos de la población.

2.1.2. En un estudio de mercado, que se realizó en el D. F., 30 de las 50 amas de casa entrevistadas, prefirieron una determinada marca de detergente. Diga:

a) ¿Cuál es la muestra utilizada?
b) ¿Cuál es la población de interés?
c) ¿Cuál es la estadística muestral y cuál es el parámetro de la población?
d) Diga si cada una de las siguientes afirmaciones es de tipo descriptivo o inferencial:.
i) La proporción muestral $\hat{p}$ es 0.60 .
ii) Con una determinada probabilidad, la proporción poblacional p es menor que 0.8 y mayor a 0.4.

2.1.3. ¿Para qué estadísticos se puede aproximar su distribución muestral a la NORMAL?

2.1.4. ¿Qué efecto tiene el aumento del tamaño de muestra en la distribución muestral de algunos estadísticos?

2.1.5. ¿Cuál es el nombre de la desviación estándar de la distribución de muestreo de un estadístico?

2.1.6. ¿Puede considerarse como la realización de una variable aleatoria a un estadístico calculado a partir de una muestra, seleccionada mediante un procedimiento probabilístico?

2.1.7. Explicar la relación que existe entre:

Distribución de muestreo.
Teorema Central del Límite.
Inferencia Estadística.

2.1.8. ¿Qué es una distribución de muestreo y para qué sirve?

2.1.9. Diga cuál es la importancia del Teorema Central del Límite.

CONSTRUCCIÓN DE DISTRIBUCIONES MUESTRALES.

2.2.1. En una despensa encontramos 3 cajas de galletas: la C_1, la C_2 y la C_3. La C_1 , C_2 y C_3 contienen 30, 35 y 50 galletas, respectivamente.

Se selecciona al azar una caja de galletas. Sea X una variable aleatoria que indica si la caja contiene más de 40 galletas o no.

a) Obtener la distribución de probabilidad de X. Calcular su media y su varianza.
b) Encuentra la distribución de muestreo de la proporción muestral $\hat{p}$ de cajas que contienen más de 40 galletas. Para ello, considera todas las muestras posibles de tamaño 2 con reemplazo.
c) Calcula la media y varianza de la distribución de muestreo de $\hat{p}$. Compara tus resultados con los obtenidos en el inciso a.

2.2.2. Supóngase que una población está dada por los siguientes datos:

{ 2, 4, 6, 8, 10, 12 }

a) Determine el número posible de muestras de tamaño cinco que se pueden obtener sin reemplazo y lístelas.
b) Obtenga la mediana para cada una de las muestras posibles.
c) Obtenga la distribución de muestreo de la mediana muestral.
d) Obtenga la media y la varianza de la distribución de muestreo para la mediana muestral.

2.2.3. Se tienen 12 radiotransmisores de policía trabajando en un día normal. El uso de los radiotransmisores por día es como sigue:

No. de llamadas x	No. de radio transmisores con ese número de llamadas
10	2
12	6
14	4
	12

Se seleccionan al azar dos radiotransmisores con reemplazo. Encuentre:

a) La distribución de muestreo del número promedio de llamadas por radiotransmisor en la muestra de tamaño dos.

b) El valor esperado y la varianza de la distribución del número promedio de llamadas por radiotransmisor en la muestra de tamaño dos.

2.2.4. Una población consta de seis elementos. Supóngase que para cada uno de los elementos de esta población se registra el valor de una variable X y se obtiene el conjunto:

$$\Omega_x = \{ 0, 0, 1, 2, 5, 16 \}$$

a) Calcule la media, mediana y varianza de X en la población

b) Encuentre todas las muestras de tamaño 3, sin reemplazo, que se pueden obtener de $\Omega_{x.}$

c) Calcule la distribución de muestreo para la media muestral. Encuentre el valor esperado y la varianza de esta distribución. Compárela con la media poblacional.

d) ¿Con qué probabilidad la media muestral difiere de la media poblacional en más de dos?

e) Obtenga la distribución de muestreo de la mediana. Encuentre el valor esperado y la varianza de esta distribución.

f) ¿Cuál de las dos distribuciones muestrales obtenidas presenta mayor variabilidad y por qué?

2.2.5. Un director de la Secretaría de Hacienda desea investigar el número de veces por año que los contribuyentes de una cierta comunidad acuden a la oficina local a solicitar asesoría fiscal.

Supóngase que el número de habitantes en la localidad es de 20,000 y que se define a la variable aleatoria X como el número de veces que un contribuyente acude a solicitar asesoría. Si la distribución de probabilidad de X es la siguiente:

x	0	1	2	3
P(X=x)	0.1	0.6	0.2	0.1

y se toma una muestra aleatoria de tamaño 2, con reemplazo:

a) Calcule la distribución conjunta de (X_1,X_2), siendo X_1 y X_2 variables independientes.

b) Encuentre la distribución muestral de $\overline{X}$, donde ésta se define como:

$$\overline{X} = \frac{X_1 + X_2}{2}$$

c) Calcule $E(\overline{X})$ y $Var(\overline{X})$.

2.2.6. En una librería, cada uno de los libros se vende a $3.00. Las probabilidades de que un cliente compre uno o dos o tres libros son 1/3, 1/2 y 1/6, respectivamente. Si X denota el monto de la venta y, si se seleccionan aleatoriamente 3 personas de un conjunto muy grande de clientes que están comprando libros en la librería:

a) Construya la distribución de muestreo de la media muestral para el monto de la venta.

b) Construya la distribución de muestreo de la mediana muestral para el monto de la venta.

c) Calcule las desviaciones estándar de la media muestral y de la mediana muestral.

2.2.7. Para los 1000 empleados de un edificio de oficinas, la variable aleatoria X denota el peso (en Kgs.) por persona, se distribuye como sigue:

X	50	60	80	100
No. de personas	60	300	480	160

a) Encuentre E(X) y V(X).

b) Obtenga la distribución de muestreo para $\overline{X}$ considerando muestras aleatorias con reemplazo de tamaño 2. Calcule $E(\overline{X})$ y $V(\overline{X})$.

2.2.8. En una urna hay cinco bolas blancas, tres negras y dos rojas. Se eligen dos de éstas mediante muestreo aleatorio sin reemplazo. Sea X una variable aleatoria que atribuye

el valor de 1 ó 2 ó 3 a los resultados pelota blanca o negra o roja, respectivamente. Obtenga:

a) La distribución de muestreo de $\overline{X}$.
b) La media y la varianza de $\overline{X}$.

2.2.9. Se toman muestras aleatorias de tamaño 2 de una población finita formada por los siguientes números: 1, 3, 5, 7, 9 y 11.

a) Calcule la media y la desviación estándar de la población.
b) Construya la distribución muestral de la media en el caso en el que se selecciona la muestra sin reemplazo.
c) Utilice el inciso b) para calcular la probabilidad de que se obtenga una media mayor o igual que 10.
d) Construya un histograma para las medias de todas las muestras posibles de tamaño 2 que se pueden seleccionar en un muestreo con reemplazo.
e) Utilice el inciso d) para calcular la probabilidad de que se obtenga una media mayor o igual que 10?

2.2.10. Suponga que se tienen seis máquinas de escribir: A, B, C, D, E y F y que las máquinas C y F necesitan reparación, mientras que las otras no.

a) Obtenga la distribución de muestreo de la proporción de máquinas que necesitan reparación, si se toman muestras de dos máquinas sin reemplazo.
b) Calcule el valor esperado de la distribución de muestreo de $\hat{p}$ y compárela con la verdadera proporción de máquinas que necesitan reparación en la población.
c) Calcule el error estándar $\sigma_{\hat{p}}$ de la distribución muestral de proporciones.

2.2.11. En el ITAM hay 3 vehículos (V1, V2 y V3) para el transporte de empleados y alumnos. El automóvil V1 necesita afinación y los otros dos no. Si se selecciona al azar un auto, sea Y la variable aleatoria que indica si el vehículo necesita afinación.

a) Obtenga la distribución de probabilidad de Y.

b) Calcule el valor esperado y la varianza de Y.

c) Si se toman todas las muestras posibles de tamaño 2 (vehículos) con reemplazo, encuentre la distribución de muestreo de la proporción muestral $\hat{p}$ de vehículos que requieren afinación.

d) Obtenga el valor esperado y la varianza de la distribución muestral de $\hat{p}$.

e) Compare los resultados de los incisos b y d y comente.

2.2.12. Supóngase que los únicos valores que puede tomar una variable de interés son 1, 4 y 7. Considere muestras de tamaño dos con reemplazo.

a) Obtenga la distribución de muestro para S^2.

b) Compare la media de la distribución obtenida en a) con la verdadera varianza.

c) Calcule la probabilidad de que S^2 tome valores entre 4 y 10.

2.2.13. En la siguiente tabla se presenta la función de probabilidades de la variable aleatoria: X = número de accidentes de tránsito al día en la delegación A de la ciudad.

x	1	2	3
$P(X = x)$	0.1	0.4	0.5

Si se toman muestras aleatorias de tamaño 2 con reemplazo:

a) Construya la tabla para la distribución conjunta de (X_1, X_2), siendo X_1 y X_2 variables aleatorias independientes.

b) Encuentre la distribución muestral del estadístico:

$$W = \frac{(X_1 - X_2)^2}{2}$$

c) Calcule la media y la varianza de la distribución muestral de W.

d) Considérese que el número de viajes Y, que realizan las grúas de la delegación por día, únicamente está relacionado con el número de accidentes, de acuerdo a:

$$Y = 5 + 2X$$

Calcule el valor esperado y la varianza de Y.

2.2.14. Se realiza una investigación de mercado con el propósito de conocer la factibilidad de abrir una nueva tienda de autoservicio en cierta zona de la ciudad. Como parte del estudio se consideró la variable Z, que representa el número de veces al mes que cada una de las amas de casa asiste al supermercado. La distribución de Z está dada por

z	3	4	5	6
Número de amas de casa.	30	50	15	5

a) Calcule E(Z) y V(Z).

b) Obtenga la distribución de muestreo de S^2 para el caso en que se toman muestras aleatorias con reemplazo de tamaño dos.

c) Calcule la media y la varianza de la distribución de muestreo del inciso b).

DISTRIBUCIÓN DE MUESTREO DE: $\overline{X}$, $\hat{p}$, S^2, $\overline{X}+\overline{Y}$, $\overline{X}-\overline{Y}$, $\hat{p}_1-\hat{p}_2$, S_1^2/S_2^2

Identifique la variable de interés en cada uno de los siguientes ejercicios:

2.3.1. En el ejercicio 2.2.7, si se tomaran muestras aleatorias de tamaño 36, ¿cuál sería la probabilidad aproximada de que, la media muestral difiera de la verdadera media poblacional en no más de medio kilo?

2.3.2. En el ejercicio 2.2.13, si se tomaran muestras aleatorias de 35 días de la función de probabilidades de Y, calcule la probabilidad de que la media muestral de Y difiera de la media poblacional en no más de 0.5 unidades.

2.3.3. Se realizó un estudio sobre el número de accidentes automovilísticos por día en las carreteras del país. Sean X y Y el número de accidentes automovilísticos en las carreteras del norte y del sur, respectivamente. Las variables aleatorias X y Y se distribuyen como Poisson con medias 15 y 20 respectivamente. Se selecciona una muestra de 50 días para X y de 40 para Y.

La variable de interés es $\overline{X}+\overline{Y}$:

a) Diga qué modelo probabilístico es adecuado para esta nueva variable. Especifique su media y su varianza.

b) Calcule: $P\left(\overline{X}+\overline{Y}>37\right)$

2.3.4. Un productor de pinturas asegura que la duración promedio de las mismas es de 5 años. Se sabe por experiencia que la desviación estándar de la duración es de medio año y que la distribución de la duración de la pintura es aproximadamente Normal.

a) Si se pinta una casa con dicha pintura y ésta sólo dura 4 años y medio, ¿consideraría usted que este hecho sería evidencia suficiente de que lo que asegura el productor es falso? (Sugerencia: suponga que $\overline{X}$ = 4.5 con n = 1 y obtenga la probabilidad de que esto pase suponiendo que lo que dice el productor es cierto).

b) Si se pintan 10 casas con dicha pintura, y éstas sólo duran en promedio 4.5 años ¿habría razón para suponer que lo que dice el productor es falso?

c) Repita el inciso (b) si la duración promedio de la pintura es de 4.9 años para las 10 casas pintadas.

2.3.5. En un banco, el tiempo que transcurre entre la llegada de dos clientes a la unifila sigue una distribución exponencial con media de 3 minutos.
Calcule la probabilidad de que, el tiempo promedio de espera en una muestra aleatoria de 36 clientes, sea mayor a 3.8 min.

2.3.6. Para cierta prueba de aptitud se sabe con base en la experiencia que el número medio de aciertos es 1000 con una desviación estándar de 125. Si se aplica la prueba a 100 personas seleccionadas al azar, aproximar las siguientes probabilidades:

a) $P(\overline{X} > 1020)$.
b) $P(\overline{X} < 975)$.
c) $P(960 < \overline{X} < 1040)$.
d) $P(985 < \overline{X} < 1015)$.

2.3.7. Un banco reportó al Gobierno Federal que sus cuentas de ahorro tienen un saldo promedio de $1,890 y una desviación estándar de $264. Los auditores del Gobierno seleccionan aleatoriamente 144 cuentas de ahorro para comprobar la confiabilidad del reporte dado por el banco. Los auditores van a certificar el reporte del banco solo si el saldo medio de la muestra difiere a lo más en $50 del saldo medio reportado. ¿Cuál es la probabilidad de que los auditores no certifiquen el informe del banco?

2.3.8. Una encuesta de agentes de compras de 250 compañías industriales señaló que 25% de los compradores reportaron en enero de 1985 niveles más altos de nuevos pedidos que en meses anteriores (Wall Street Journal, 4 de febrero, 1985). Supóngase que los 250 agentes de la muestra representan una muestra aleatoria de los agentes de compras de compañías en todo Estados Unidos.

a) Describa la distribución de muestreo de $\hat{p}$, la proporción de compradores en Estados Unidos con niveles más elevados de nuevos pedidos en enero. Justifique su respuesta.

b) ¿Cuál es la probabilidad de que $\hat{p}$ difiera en más de 0.01 de p?

2.3.9. Muchos insumos de producción, como el mineral de hierro, el carbón y el azúcar sin refinar, se muestrean para determinar su calidad por un método que implica la toma periódica de muchas pequeñas muestras cuando el material se mueve sobre una banda transportadora. Posteriormente las muestras pequeñas se juntan y mezclan para formar lo que se denomina una muestra compuesta. Sea Y_i el volumen promedio de la i-ésima muestra pequeña de un lote particular y suponga que Y_1, Y_2,.. Y_n es una muestra aleatoria, en donde cada Y_i tiene media μ (en pulgadas cúbicas) y varianza σ^2. El volumen promedio de las muestras, μ , se puede regular ajustando el tamaño del equipo que se utiliza para el muestreo. Suponga que la varianza de los volúmenes de las muestras, σ^2 , es aproximadamente 4. Se requiere que el volumen total de la muestra compuesta (es decir ΣY_i) exceda las 200 pulgadas cúbicas con una probabilidad aproximada de 0.95 cuando se seleccionan n=50 muestras pequeñas. Determine el valor de μ que permitirá satisfacer los requerimientos del muestreo.

2.3.10. En el ejercicio 2.2.14, considere otra variable X definida como X = 3Z-4. Si se tomara una muestra aleatoria de tamaño 50, ¿cuál es la probabilidad de que $\overline{X}$ difiera de la media poblacional μ_x en al menos 0.5 unidades?

2.3.11. Se procede a detener el funcionamiento de una máquina para repararla, si en una muestra aleatoria de 100 artículos de la producción diaria se encuentran por lo menos 15% de artículos defectuosos. (Diariamente se produce un gran número de artículos). Si realmente la máquina produce sólo 10% de artículos defectuosos, encuentre la probabilidad de que se pare la máquina en un día dado.

2.3.12. Supóngase que el 45% de las personas en el D.F. está de acuerdo con el programa de un día sin auto, ¿cuál es la probabilidad de que, en una muestra de 2400 personas menos de 1000 estén de acuerdo con el programa?

2.3.13. Un departamento estatal de carreteras vigiló una determinada autopista durante 35 períodos de una hora seleccionados al azar en un mes. Se contó el número de trailers que pasaban durante cada período de una hora y se calculó la media muestral. Supóngase que la distribución del número de trailers por hora tiene media $\mu = 50$ y desviación estándar $\sigma = 7$.

a) ¿Cuál es la probabilidad de que la media muestral para n = 35 períodos de una hora sea mayor de 55?
b) Suponga que se cuentan los trailers para cada uno de n = 4 períodos de una hora, seleccionados al azar. ¿Cuál es la probabilidad de que $\overline{X}$ sea mayor que 55? ¿Qué se requiere conocer sobre la distribución del número de trailers que circulan por hora en la autopista?
c) ¿Cuál es la probabilidad de que el número total de trailers, para un período de cuatro horas, sea mayor que 180?

2.3.14. Un antropólogo quiere estimar la estatura promedio de los hombres de cierta raza. Si se supone que la desviación estándar de la población es de 2.5 pulgadas y se selecciona al azar a 100 hombres:
a) Encuentre la probabilidad de que, la diferencia entre la media de la muestra y la media verdadera de la población no exceda de 0.5 pulgadas.
b) Suponga que el antropólogo quiere que la diferencia entre la media de la muestra y la media de la población sea menor de 0.4 pulgadas con una probabilidad de 0.95. ¿Cuántos hombres tendrá que seleccionar para alcanzar su objetivo?

2.3.15. Para determinar la base gravable de los contribuyentes, la SHCP sabe que las cantidades declaradas por concepto de deducciones médicas (X_1), contribuciones caritativas (X_2) y funerales (X_3) son variables aleatorias independientes y normalmente distribuidas con medias \$400, \$800 y \$100 y desviaciones estándar de \$100, \$250 y \$40 respectivamente.

¿Cuál es la probabilidad de que la cantidad total declarada por concepto de estas tres deducciones no exceda a $1600?

2.3.16. Una cadena de supermercados va a lanzar la campaña promocional: "Saque boleto blanco y no pague". Por estudios anteriores, se sabe que el gasto promedio por cliente es de $ 391.70 con una desviación estándar de $ 138. Si se van a distribuir 150 boletos premiados diariamente y si la cadena de supermercados está dispuesta a "regalar" máximo $60,000 por día, ¿cuál es la probabilidad de que en un día determinado se rebase el límite establecido?

2.3.17. El tiempo requerido por unos trabajadores para completar un trabajo de armado tiene una media de 50 minutos y una desviación estándar de 8 minutos. Para revisar el rendimiento de los trabajadores, su supervisor registra el tiempo que, 60 empleados tardan en completar su trabajo.

a) ¿Cuál es la probabilidad de que la media muestral sea mayor a 52 minutos?

b) Si la media muestral resulta ser de 53 minutos, ¿debería considerar el supervisor que el proceso de armado está siendo más lento?. (Suponga que la desviación estándar no cambió).

2.3.18. Suponga que se toma una muestra aleatoria de tamaño 100 de una distribución Poisson con parámetro $\lambda = 0.1$. Sean $X_1, X_2, .. X_{100}$ elementos de la muestra y considere:

$$Y = \Sigma X_i \qquad y \qquad \overline{X} = Y/100$$

a) Calcule el valor esperado y la varianza de $\overline{X}$

b) Calcule $P(Y \leq 5)$ utilizando el Teorema Central del Límite.

c) Ahora calcule exactamente $P(Y \leq 5)$ y compárela con la obtenida en b).

(Use el resultado siguiente: Si $X_1, X_2, .. X_n$ es una muestra aleatoria de una Poisson con parámetro λ, entonces Y tiene una distribución Poisson con parámetro $n\lambda$).

2.3.19. Después de haber observado muchas generaciones de estudiantes, se sabe que el 75% de los alumnos aprueban la materia "Estadística Básica" de una Universidad. Si los alumnos de una muestra de 40 estudiantes inscritos en dicho curso pueden ser considerados como similares a aquellos que han sido observados en cuanto a preparación académica, edades e intereses:

a) ¿Cuál es la probabilidad de que al menos 50% de ellos aprueben el curso?

b) ¿Cuál es la probabilidad de que haya un máximo de 10% de reprobados?

c) ¿Cuál es la probabilidad de que haya a lo más un reprobado?

2.3.20. Después de analizar los resultados de las clínicas de reducción de peso se obtuvo que el número medio de kilos que bajó una persona en un mes, es de 8 kilogramos con una varianza de 9. Una de estas clínicas llamadas "Baja Kilitos, S.A." tiene 20 pacientes y el dueño quiere lanzar una campaña motivacional en la que ofrecerá regalar un viaje a Huatulco a todos los pacientes si bajan en promedio 10 kgs en un mes. Usted es asesor del dueño, ¿le recomendaría o no que haga la campaña?

2.3.21. El conmutador de una empresa cuenta con 45 líneas telefónicas. El número de llamadas que se reciben en cada una de éstas, en un período de tres minutos, se puede modelar como una variable aleatoria Poisson. El parámetro de esta función de densidad representa el promedio de llamadas por línea durante un período de tres minutos. En la compañía se ha determinado que, de acuerdo a la experiencia, se reciben 0.85 llamadas en tres minutos, en un día hábil entre las 11 y las 12 hrs.

¿Cuál es la probabilidad de que en un día hábil entre las 11 y las 12 hrs. el conmutador se sature? (Es decir, ¿cuál es la probabilidad de que el número de llamadas que recibe el conmutador en un momento dado sea mayor que 45?). A partir de su resultado, comente la factibilidad de pensar en la adquisición de un conmutador con mayor número de líneas.

2.3.22. Los pesos de los paquetes que se almacenan en una tienda tienen una media de 300 kgs. con una desviación estándar de 50 kgs. ¿Cuál es la probabilidad de que, el peso de 25 paquetes recibidos al azar supere los 8000 kgs.? Establezca los supuestos necesarios.

2.3.23. Considere a la variable aleatoria X con función de densidad

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

y a $\overline{X}$ como la media de una muestra aleatoria de tamaño 75.
Calcule: $P(0.45 < \overline{X} < 0.55)$.

2.3.24. Se sabe que la distribución del espesor de un material plástico es normal, con una desviación estándar de 0.01 cm. Para una muestra aleatoria de 25 piezas de este material da como resultado una desviación estardar muestral de 0.015 cm. La varianza de la población es de .0001.

a) ¿Cuál es la probabilidad aproximada de que, la varianza muestral sea igual o mayor que $(0.015)^2$cm.^{2}?

b) ¿Qué puede concluir con respecto a la variación de este proceso?

2.3.25. Las variaciones en el número de unidades manejadas por los operarios A y B son las mismas. Se obtuvo el número promedio de unidades manejadas por día y la varianza muestral para el operario A y B, después de haberlos observado durante 16 y 21 días, respectivamente.

¿Cuál es la probabilidad de que el cociente de varianzas muestrales exceda de 0.52 suponiendo que, los números de unidades manejadas por día por los dos operarios, son dos variables aleatorias independientes que siguen una distribución aproximadamente normal?

2.3.26. El presidente de una fabrica está molesto por el número de teléfonos defectuosos que se producen en ésta. En promedio, diariamente se devuelven 120 teléfonos con una desviación estándar de 81. A causa de ese problema, él ha decidido que, a menos que en los próximos 40 días, tenga una confianza del 85% de que no serán devueltos más de 135 teléfonos por día, ordenará revisar el proceso de producción. ¿Tomará la medida el presidente?

2.3.27. Una gran cadena de supermercados desea comprar una compañía de alimentos que tiene 122 tiendas de autoservicio. Antes de cerrar el trato, el dueño de la cadena ha decidido analizar los registros financieros de 40 de las tiendas de la empresa que va a adquirir. La gerencia de ésta afirma que las utilidades de cada establecimiento tienen una distribución normal con la misma media y una desviación estándar de $1000. ¿Qué probabilidad hay de que la media muestral de las 40 tiendas esté a lo más a $ 150 de la media real?

2.3.28. La duración de un componente eléctrico se encuentra exponencialmente distribuida con una vida media de 100 horas. Si del proceso de producción se toma una muestra aleatoria de 40 componentes, ¿cuál es la probabilidad de que la vida media muestral sea mayor de 120 horas?

2.3.29. Determine el número de unidades que debe asegurar una compañía de seguros de tal forma que, el monto total de reclamos no sea mayor a 22,000 con una probabilidad de 0.97. Supóngase que, el reclamo individual tiene un valor esperado de 12 y una varianza de 9.61 y además que existe independencia entre los reclamos.

2.3.30. Se sabe que, el 30% de las compras en una tienda son por cantidades superiores a $500. Si se seleccionan muestras aleatorias de 100 compras, ¿qué proporción de éstas, es posible que tengan entre 20% y 30% de compras mayores a $500?

2.3.31. En una universidad hay un movimiento para formar un sindicato de profesores. Aproximadamente 40% de toda la institución favorece la idea. Un empleado a favor del movimiento, toma una muestra aleatoria de 100 profesores y calcula $\hat{p}_1$: la proporción muestral a favor del sindicato. Una persona que no aprueba la idea toma otra muestra aleatoria independiente del mismo tamaño y obtiene $\hat{p}_2$ la proporción de profesores a favor en esta muestra. Calcule la probabilidad de que $\hat{p}_1$ exceda a $\hat{p}_2$ por más de 0.1.

2.3.32. Una población muy grande tiene una media de 20 y una desviación estándar de 1.4. Si se toman muestras de 49 observaciones:

a) ¿Cuál es el valor esperado de la distribución de la media muestral?

b) ¿Cuál es la desviación estándar de la distribución de muestreo anterior?

c) ¿Qué porcentaje de posibles valores de $\overline{X}$ diferirán de la media de la población por más de 0.2?

2.3.33. Los amperímetros producidos por una compañía se venden en el mercado con la especificación de que la desviación estándar de las lecturas es menor que 0.2 amp. Se utilizó uno de esos amperímetros para efectuar 10 lecturas independientes en un circuito de prueba con corriente constante. Si la varianza de estas 10 mediciones es 0.065, ¿indica este resultado que el amperímetro que se utilizó no satisface las especificaciones del mercado? ¿Qué supuestos son necesarios al establecer su resultado?

2.3.34. Si $\overline{X}_1$ y $\overline{X}_2$ representan las medias de muestras aleatorias tomadas de dos poblaciones discretas con las siguientes distribuciones de probabilidad:

x_1	2	3	7
$P(X_1 = x_1)$	1/3	1/3	1/3

x_2	1	3
$P(X_2 = x_2)$	2/3	1/3

Obtenga la probabilidad de que $\overline{X}_1$ - $\overline{X}_2$ sea mayor que 1.84 cuando se toman muestras aleatorias independientes con reemplazo de tamaños $n_1 = 125$ y $n_2 = 100$, respectivamente.

2.3.35. Muestre que, si la media de una muestra aleatoria de tamaño n se utiliza para estimar la media de una población con desviación estándar σ, hay 50% de posibilidades de que el error de estimación sea menor que $0.6745\ \sigma/\sqrt{n}$.

2.3.36. a) Si X~ χ^2_m y Y~ χ^2_k ¿Cuál es la distribución de probabilidad de W, donde W está definida por $W = \dfrac{kX}{mY}$?

b) Si $Z_1, Z_2, ..., Z_n$ son variable aleatorias normales estandarizadas e independientes, ¿Cuál es la distribución de probabilidad de W, donde W está definida por:

$$W = \sum_{i=1}^{n} Z_i^2$$

2.3.37. Demuestre que el cuadrado de una variable *t* con *k* grados de libertad es una variable $F_{1,k}$.

2.3.38. Un médico quiere estimar la estatura promedio de los pacientes adolescentes. Si se sabe que, la desviación estándar de la población es de 2.5 cms. y selecciona al azar a 100 pacientes, encuentre la probabilidad de que la diferencia entre la media de la muestra y la media de la población no exceda 0.5 cms.

2.3.39. Como parte de un estudio sobre el uso de agua potable, un grupo de consultores ha establecido que en las zonas habitacionales la cantidad de metros cúbicos de agua consumidos por semana y por hogar es una variable aleatoria con media 1.8 m^3 y desviación estándar de 0.9 m^3.

Se ha decidido verificar esto en base al consumo semanal de agua en una muestra aleatoria en 50 hogares. Se aceptará lo que los consultores dicen si el valor de la media muestral está entre 1.6 y 2.0 m^3. Calcule la probabilidad de que no se acepte lo que ellos establecen.

2.3.40. Se tomó una muestra aleatoria de países y se investigó el nivel de inflación de cada uno de estos en 1990. A continuación se presentan los datos obtenidos:

Suecia	10.9%	EU	6.1%	Japón	3.8%
Inglaterra	9.3%	Australia	6.0%	Bélgica	3.5%
España	6.5%	Suiza	5.3%	Francia	3.3%
Italia	6.4%	Canadá	5.0%	Alemania	2.8%

a) Si la desviación estándar mundial fue de 1.7%, ¿cuál es la probabilidad de que la media muestral difiera a lo más en 0.02 de la media poblacional? Establezca los supuestos necesarios.
b) Si la proporción real de países con inflación mayor o igual a 5% es de .60, ¿cuál es la probabilidad de observar una proporción muestral igual o mayor que 0.67?

2.3.41. Una agencia de empleos aplica pruebas de aptitudes a todos los que solicitan trabajo. La empresa ha reunido datos durante años y ha llegado a la conclusión de que, la distribución de las puntuaciones no es normal, sino que está sesgada a la izquierda con una media de 86 y una desviación estándar de 16.
¿Cuál es la probabilidad de que, en una muestra de 75 solicitantes, la puntuación promedio sea a lo más de 84 o mayor que 90?

2.3.42. Se sabe que la demanda de ambulancias de un gran hospital se distribuye como Poisson con media de 6 por día. Calcule la probabilidad de que, en un período de 40 días, la demanda promedio por día sea al menos de 7.

2.3.43. Un fabricante garantiza que la duración media de sus fusibles es de 12.4 minutos con una desviación estándar de 4 minutos. Determine el valor de “a” tal que, el 1% de las veces al tomar muestras aleatorias de 36 fusibles la media muestral no sea mayor de “a” minutos.

2.3.44. Se sabe que, el consumo de café anual por familia tiene una distribución normal con media desconocida μ y una desviación estándar de 1.25 libras.

a) Si se toma una muestra de 36 familias y se registra su consumo de café. ¿Cuál es la probabilidad de que la media muestral se encuentre, como máximo, a media libra de la media de la población?
b) ¿De qué tamaño debe ser la muestra a fin de que tenga una seguridad del 95% de que, la media muestral se encuentre, como máximo, a media libra de la media de la población?

2.3.45. El peso de un producto, se distribuye en forma normal con media 8.8 kgs. y desviación estándar de 2.64.

a) ¿Cuál es la probabilidad de que un producto, seleccionado al azar pese más de 6.6 kgs?
b) Se toma una muestra de tamaño 9 de los productos. ¿Cuál es la probabilidad de que la varianza de la muestra sea menor de 3?

2.3.46. En un proceso de producción de varillas de acero se sabe que la longitud Y es aleatoria y que Y^2 sigue una distribución $\chi^2_{(27)}$. Si se mide la longitud de cada varilla con un instrumento cuyo error de medición X es aproximadamente Normal con media $\mu = 0$ y desviación estándar $\sigma = 0.02$ metros, ¿cuál es la probabilidad de que el error relativo en la medición exceda a 0.01?. (Considere que, el error relativo está dado por X/Y).

2.3.47. Suponga que con base en la experiencia, se sabe que el tiempo que los clientes esperan para ser atendidos en la caja de un banco, es una variable aleatoria que sigue una distribución Normal con media de 15 minutos y desviación estándar de 4 minutos.

Si se toma una muestra aleatoria de 10 clientes, ¿cuál es la probabilidad de que la varianza muestral sea mayor que 30? Indique explícitamente todos los supuestos que considere necesarios.

2.3.48. Suponga que, en el mismo contexto del problema anterior, se obtiene una segunda muestra, ahora de tamaño 11, independiente de la primera. ¿Cuál es la probabilidad de que la varianza muestral S_1^2 sea menor que tres veces la varianza muestral S_2^2 ?

2.3.49. Con el objeto de seleccionar a una de dos secretarias (A o B) para un trabajo, a cada una de ellas, en 50 pruebas, se les toma el tiempo que les lleva realizar un cierto trabajo en máquina.

Si las medias muestrales para las 50 pruebas difieren en más de 1 segundo, la secretaria con la media más pequeña obtiene el trabajo. De otra forma, se considera que hay empate. Si las desviaciones estándar de los tiempos para ambas secretarias son de 2 segundos, ¿cuál es la probabilidad de que la secretaria A obtenga el trabajo, aún cuando las dos sean igualmente hábiles ?

2.3.50. El tiempo de espera en cola, en una tintorería, tiene generalmente una distribución de probabilidad sesgada a la derecha, ya que raramente ocurren períodos de tiempo largos de espera.

Supóngase que la variable aleatoria tiene media de 3.8 min. con una desviación estándar de 0.75 min.

a) Especifique la población, la población estadística, la muestra y la variable de interés.

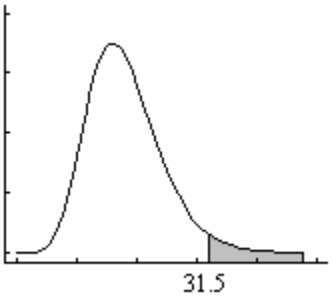
b) Para una muestra de tamaño 48, ¿cuál es la probabilidad de que el tiempo medio de espera muestral sea menor a 3.5 min.? (Justifique su respuesta).

c) Suponga que el dueño de la tintorería desea que la diferencia entre la media de la muestra y la media de la población sea menor de 0.5 min. con una probabilidad de .99. ¿De cuántos clientes debe ser la muestra para alcanzar su objetivo?

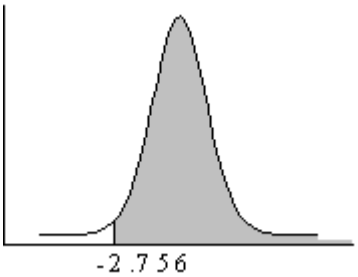
USO DE TABLAS

I.Obtenga la probabilidad sombreada en cada una de las siguientes figuras:

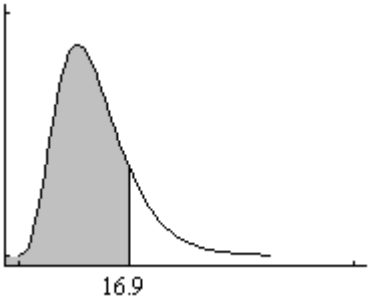
2.4.1 χ^2_{18}



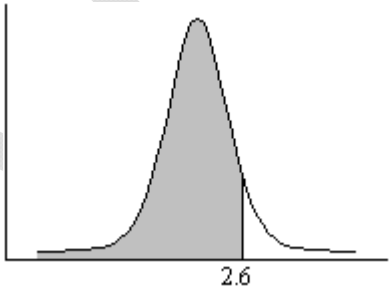
2.4.2 t_{29}



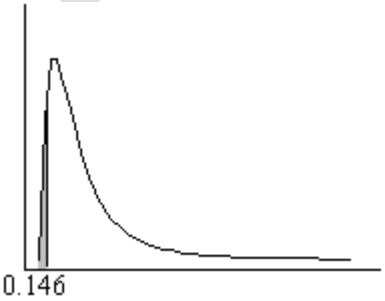
2.4.3 χ^2_9



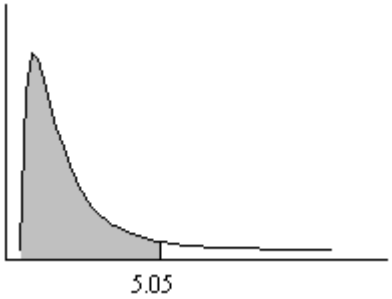
2.4.4 t_{20}



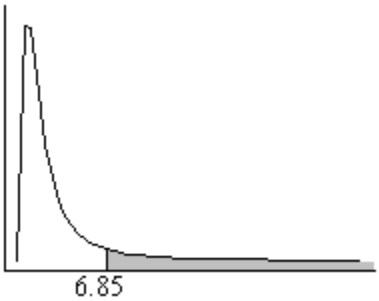
2.4.5 $F_{(5,7)}$



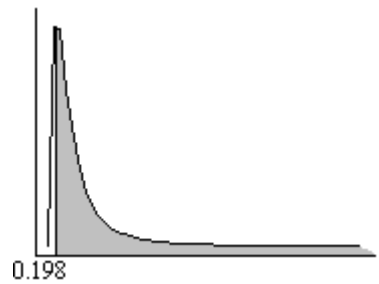
2.4.6 $F_{(10,20)}$



2.4.7 $F_{(7,5)}$



2.4.8 $F_{(8,4)}$



2.4.9 $P\left(.71 \le \chi^2_{(4)} \le 7.7\right)$

2.4.10 $P(17.5 \le \chi^2_{(11)} \le 19.6)$

2.4.11 $P(13.5 \le \chi^2_{(30)} \le 15.1)$

2.4.12 $P(1.15 \le \chi^2_{(5)} < 15.1)$

2.4.13 $P(-2.093 < t_{19} < 2.861)$

2.4.14 $P(-2.528 \le t_{20} < -1.725)$

2.4.15 $P(-2.179 \le t_{12} \le 2.681)$

2.4.16 $P(1.313 < t_{28} < 2.763)$

2.4.17 $P(.37 \le F_{(10,30)} \le 3.34)$

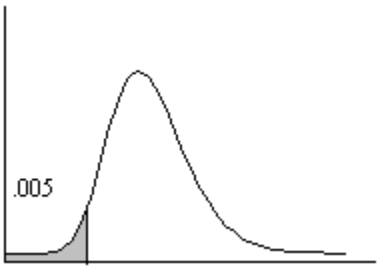
2.4.18 $P(.426 < F_{(60,15)} < 2.16)$

2.4.19 $P(.299 < F_{(30,10)} < 2.7)$

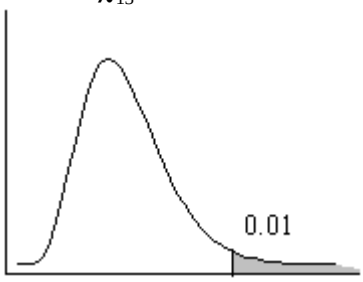
I.T.A.M.

II. Obtenga el o los valores cuantiles

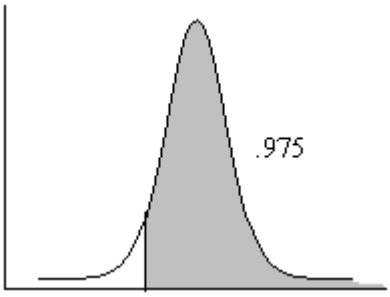
2.4.20 χ^2_{23}



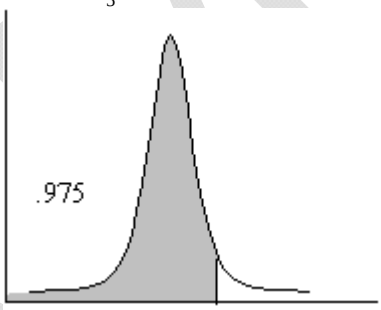
2.4.21 χ^2_{13}



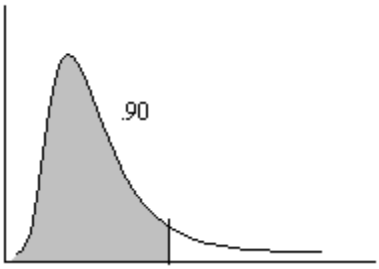
2.4.22 t_{20}



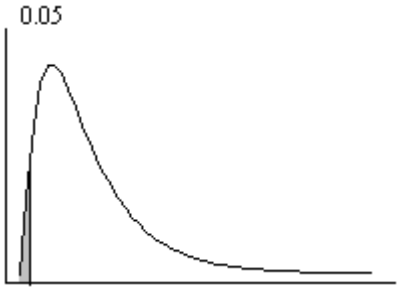
2.4.23 t_5



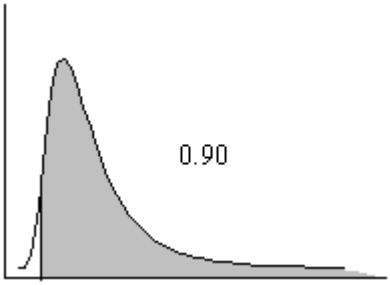
2.4.24 $F_{(10,20)}$



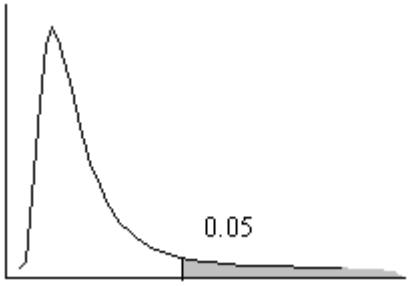
2.4.25 $F_{(5,15)}$



2.4.26 $F_{(20,10)}$



2.4.27 $F_{(15,8)}$



2.4.28 $P(7.26 < \chi^2_{15} < b) = .925$

2.4.29 $P(-1.721 < t_{21} < b) = .045$

2.4.30 $P(a < F_{2,4} < 4.32) = .875$

2.4.31 $P(a < F_{60,60} < 1.67) = .875$

En los siguientes ejercicios, considere la misma probabilidad en cada una de las colas de la distribución.

2.4.32 $P(a < \chi^2_{40} < b) = .80$

2.4.33 $P(a < \chi^2_{70} < b) = .99$

2.4.34 $P(a < t_9 < b) = .80$

2.4.35 $P(a < t_{11} < b) = .98$

2.4.36 $P(a < F_{60,60} < b) = .95$

2.4.37 $P(a < F_{10,20} < b) = .98$

2.4.38 $P(a < F_{20,10} < b) = .98$

I.T.A.M.

PROPIEDADES DE ESTIMADORES

3.1.1. Responda falso o verdadero a las siguientes afirmaciones. Justifique su respuesta.

a) Un estimador es cualquier característica numérica de la población.
b) Es posible obtener estimadores diferentes para un mismo parámetro poblacional.
c) El error de estimación y el sesgo son conceptos que se refieren a lo mismo.
d) Se dice que un estadístico es un estimador eficiente del parámetro de una población si es insesgado y su varianza es mínima.
e) El que un estimador $\hat{\theta}$ sea insesgado, significa que cualquier estimación de θ siempre va a ser igual al parámetro poblacional.
f) Si ECM($\hat{\theta}_1$) < ECM($\hat{\theta}_2$) eso significa que V($\hat{\theta}_1$) < V($\hat{\theta}_2$).
g) Si ECM($\hat{\theta}_1$) < ECM($\hat{\theta}_2$) eso significa que $\hat{\theta}_1$ es insesgado y $\hat{\theta}_2$ no lo es.

3.1.2. De una población con media μ y varianza σ^2 se extraen dos muestras aleatorias simples e independientes de tamaños n_1 y $n_2 = n_1/2$. Sus medias muestrales son $\overline{X}_1$ y $\overline{X}_2$, respectivamente. Para estimar a μ se proponen tres estimadores:

$$\hat{\mu}_1 = \overline{X}_1 \qquad \hat{\mu}_2 = \overline{X}_2 \qquad \text{y} \qquad \hat{\mu}_3 = \frac{\overline{X}_1 + \overline{X}_2}{2}$$

a) Diga si son insesgados.
b) Encuentre su varianza.
c) ¿Cuál de los tres estimadores es mejor?

3.1.3. Si $X_1, X_2, ..., X_n$ son variables aleatorias Bernoulli independientes con el mismo parámetro p, considérense los siguientes estimadores de dicho parámetro.

$$\hat{p}_1 = \frac{\sum_{i=1}^{n} X_i - 1}{n-1}, \qquad \hat{p}_2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

¿Son estimadores insesgados de p?

3.1.4 Si X_1 y X_2 son dos variables aleatorias independientes tales que $E(X_1) = E(X_2) = \mu$ y $Var(X_1) = Var(X_2) = \sigma^2$, determine si el estimador.

$$\hat{\sigma}_1^2 = \frac{(X_1 - X_2)^2}{2}$$

es un estimador insesgado para σ^2.

3.1.5. Con base en una muestra aleatoria de 100 observaciones, tomadas de una población con media μ y varianza σ^2, considere los siguientes estimadores de μ:

$$\hat{\mu}_1 = \frac{X_1 + X_2 + X_3 + ... + X_{100}}{100}$$
$$\hat{\mu}_2 = \frac{X_1 + X_2 + X_3 + ... + X_{90}}{90}$$

a) ¿Cuál es la eficiencia relativa de $\hat{\mu}_2$ con respecto a $\hat{\mu}_1$?

b) ¿Cuál es la causa de dicha ineficiencia?

3.1.6. Se elige una muestra aleatoria independiente de 3 observaciones de una población con una función de distribución uniforme:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{si } x \notin [a,b] \end{cases}$$

a) Encuentre la media y la varianza de la función de densidad.

Se definen los siguientes estimadores para la media de la distribución:

$$\hat{\mu}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$$
$$\hat{\mu}_2 = \frac{1}{6}X_1 + \frac{1}{6}X_2 + \frac{2}{3}X_3$$
$$\hat{\mu}_3 = \frac{1}{3}X_1 + \frac{1}{6}X_2 + \frac{1}{6}X_3$$

b) Obtenga la esperanza de estos estimadores.

c) ¿Cuáles de estos estimadores son insesgados?

d) Determine las varianzas de cada estimador.

e) ¿Cuál de éstos es el mejor estimador?

3.1.7. Sea $X_1; X_2; ...; X_n$ una muestra aleatoria de una población con media μ y varianza σ^2. Considere los tres estimadores siguientes para μ:

$$\hat{\mu}_1 = \frac{1}{2}(X_1 + X_2), \qquad \hat{\mu}_2 = \frac{1}{4}X_1 + \frac{X_2 + ... + X_{n-1}}{2(n-2)} + \frac{1}{4}X_n, \qquad \hat{\mu}_3 = \overline{X}$$

a) Determine si son insesgados.

b) Encuentre la varianza de cada estimador e identifique cuál es el más eficiente.

c) Determine la eficiencia relativa de $\hat{\mu}_3$ con respecto a $\hat{\mu}_2$ y $\hat{\mu}_1$, respectivamente.

3.1.8. Un gerente de producción supone que el peso de un objeto se distribuye normalmente con varianza conocida, pero con media μ es desconocida. Se toma una muestra aleatoria de cuatro observaciones independientes X_1, X_2, X_3, X_4. Considere los siguientes estimadores de μ.

$$\hat{\mu}_1 = 2X_1$$

$$\hat{\mu}_2 = \frac{4X_1 + 3X_2 + 2X_3 + X_4}{10}$$

$$\hat{\mu}_3 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

a) Determine cuáles estimadores son insesgados.

b) ¿Cuál de los estimadores insesgados es el de mayor eficiencia relativa?

c) ¿Cuál de los estimadores tiene el menor error cuadrático medio?

3.1.9. Sea $X_1, X_2, X_3,$ y X_4 una muestra aleatoria de una distribución exponencial con parámetro θ desconocido.

De los siguientes estadísticos:

$$\hat{\theta}_1 = \frac{X_1 + 2X_2 + 3X_3 + 4X_4}{5}$$
$$\hat{\theta}_2 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

a) Determine si son estimadores insesgados de θ.
b) ¿Cuál es el error cuadrático medio de cada estimador?
c) De estos estimadores, ¿cuál es el mejor?

3.1.10. ¿Para qué valor de la constante k es $\hat{\theta} = kX$ un estimador insesgado del parámetro θ de la población, cuya función de densidad está dada por?

$$f(x) = \begin{cases} \dfrac{3}{\theta} & 0 < x < \theta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

3.1.11. Si $\overline{X}_1$ y $\overline{X}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 tomadas de una población normal con media μ y varianza σ^2:

a) Demuestre que $\hat{\mu} = W\overline{X}_1 + (1-W)\overline{X}_2$ es un estimador insesgado de μ para cualquier W tal que $0 \leq W \leq 1$.
b) ¿Cuál es el valor que debe tomar W para que la varianza del estimador $\hat{\mu}$ sea mínima?

3.1.12. Si X_1, X_2, X_3 es una muestra aleatoria tomada de una población normal con media μ y varianza σ^2. ¿Cuál es la eficiencia del estimador $\hat{\mu} = \dfrac{X_1 + 2X_2 + X_3}{4}$ en relación con $\overline{X}$?

3.1.13. Siendo $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con media μ y varianza σ^2, se tienen los siguientes estimadores de σ^2:

$$S_1^2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2, \qquad S_2^2 = \frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2$$

¿Son insesgados estos estimadores?

3.1.14. Para estudiar la proporción θ de personas de México que consideran adecuada la continuación de PECE, se eligió una muestra aleatoria de n personas y se consideró al siguiente estimador de la proporción:

$$\hat{\theta} = \frac{Y+1}{n+2}$$

en donde Y es el número de personas en la muestra que sí están de acuerdo en la continuación del PECE.

Obtenga el error cuadrático medio de $\hat{\theta}$.

3.1.15. Sea Y una variable aleatoria de Poisson con media λ. Se obtiene una muestra aletoria de tamaño n y se desea estimar λ. ¿Es $\hat{\lambda} = \sum_{i=1}^{n} Y_i(Y_i - 1)$ estimador insesgado de λ?

3.1.16. El tiempo que tarda un tren del metro en salir de una estación terminal tiene la siguiente distribución $f(x;\theta) = \frac{2x}{\theta^2}, \quad 0 \le x \le \theta, \quad \theta > 0.$

Se midió el tiempo de salida del metro en una muestra aleatoria de tres terminales del sistema y se registró el tiempo transcurrido X_1, X_2, X_3. Se desea tener una estimación del parámetro θ para poder programar adecuadamente las salidas y que no existan problemas de retraso.

Para tal propósito se definen dos estimadores, a saber:

$$\hat{\theta}_1 = \frac{X_1 + X_2 + X_3}{3}, \hat{\theta}_2 = \frac{X_1 + X_2}{2}$$

a) Obtenga el mejor estimador de θ que podría usarse.

b) Si la muestra observada da los siguientes valores: 4.3, 5.1, 2.2 minutos, obtenga la estimación de θ.

3.1.17. Sea X una variable aleatoria Poisson. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria. Considere los siguientes estimadores de la media:

$$\hat{\lambda}_1 = \overline{X} \qquad \text{y} \qquad \hat{\lambda}_2 = \frac{X_n - X_1}{2}$$

a) ¿Son insesgados estos estimadores?. Calcule el sesgo de cada estimador.
b) Encuentre el ECM de $\hat{\lambda}_1$ y $\hat{\lambda}_2$
c) ¿Cuál de los dos es mejor estimador de la media?

3.1.18. El ingreso mensual percibido por los empleados de nivel ejecutivo de una empresa tiene la distribución:

$$f(x; \theta) = 3\theta^3 / x^4 \qquad \text{para } x > \theta .$$

Para la contratación de nuevos solicitantes la empresa desea estimar el ingreso mínimo θ, por lo que selecciona una muestra aleatoria de tres empleados de ese nivel y registra su ingreso $X_1, X_2,$ y X_3.
Se propone a $\hat{\theta} = b\ \overline{X}$. Determine el valor de la constante b para que $\hat{\theta}$ sea un estimador insesgado.

3.1.19. Considere la población constituida por las industrias medianas en México. Es de interés estimar la proporción de éstas industrias que han adquirido algún financiamiento para seguir trabajando. Para ésto se define una variable aleatoria Y tal que toma el valor de uno si la industria adquirió algún financiamiento, y el valor de cero en caso contrario. Si θ es la proporción que se desea estimar, entonces Y se distribuye como una Bernoulli con parámetro θ.
Se ha decidido seleccionar una muestra aleatoria de n industrias. Considere $Y_1, Y_2, ..., Y_n$ en dicha muestra.

a) Se propone que un estimador de θ sea $\overline{Y}$.

¿Es este estimador insesgado?

b) Como la varianza del estimador $\hat{\theta} = \overline{Y}$ depende de θ, y θ no se conoce, se desea también estimar dicha varianza. Se propone:

$$\hat{V}(\overline{Y}) = \frac{\overline{Y}(1-\overline{Y})}{n}.$$

Demuestre que este estimador es sesgado. ¿Cuál es el sesgo?

3.1.20. Sean X y Y las variables aleatorias que representan el gasto mensual por familia en dos regiones del país. Dos muestras aleatorias independientes del mismo tamaño son obtenidas, una para cada región, y lo que se desea estimar es la diferencia de medias $\mu_x - \mu_y$. Las muestras fueron $X_1, X_2, ..., X_n$ y $Y_1, Y_2, ..., Y_n$.

a) El estimador $\hat{\theta}_1 = \frac{X_1 + X_n}{2} - \frac{Y_1 + Y_n}{2}$, ¿es insesgado?

b) Ahora sea $\hat{\theta}_2 = \overline{X} - \overline{Y}$, ¿es insesgado?

c) Sean ahora $\sigma_x = 23$ y $\sigma_y = 12$. ¿Cuál debería ser el número mínimo de observaciones que asegura un error estándar menor o igual a 4 para el estimador del inciso anterior?

3.1.21. Se tiene una muestra aleatoria $X_1, X_2, ..., X_n$ de una población Poisson con parámetro λ. Calcule el sesgo de los siguientes estimadores:

$$\hat{\lambda}_1 = \frac{2\sum_{i=1}^{n} iX_i}{n(n+1)}$$

$$\hat{\lambda}_2 = (X_1 + X_n)/2n$$

ESTIMACION PUNTUAL

3.2.1. Los datos que a continuación se presentan son los pesos, en gramos, del contenido de 16 cajas de cereal que se seleccionaron al azar de un proceso:
506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496.
Obtenga la estimación puntual del peso promedio.

3.2.2. A 12 personas se les tomó el tiempo (en min.) que tardaron en ser atendidas en un supermercado. Los resultados fueron los siguientes:
3.01, 3.05, 2.99, 2.99, 3.00, 3.02, 2.98. 2.99, 2.97, 2.97, 3.02, 3.01.
Obtenga la estimación puntual del tiempo promedio.
Obtenga la estimación puntual de la varianza.

3.2.3. Cierto metal se produce por lo común mediante un proceso estándar. Se desarrolla un nuevo proceso en el que se añade una aleación a la producción del metal. Los fabricantes se encuentran interesados en estimar la verdadera diferencia entre las tensiones de ruptura promedio de los metales producidos por los dos procesos. Para cada metal se seleccionan 12 especímenes y cada uno de estos se somete a una tensión hasta que se rompe. La siguiente tabla muestra las tensiones de ruptura de los especímenes en kilogramos por centímetro cuadrado:

Proceso Estándar	428	419	458	439	441	456	463	429	438	445	441	463
Proceso Nuevo	462	448	435	465	429	472	453	459	427	468	452	447

Estime puntualmente la tensión de ruptura promedio en el proceso nuevo y en el proceso estándar.

3.2.4. Una máquina produce piezas metálicas en forma cilíndrica. Se toma una muestra aleatoria de éstas y los diámetros resultan de 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Estime la varianza de la dimensión de los diámetros.

3.2.5. La Cámara de Comercio de una ciudad está interesada en estimar la cantidad promedio de dinero que gasta la gente que asiste a convenciones, calculando comida, alojamiento y entretenimiento por día. De las distintas convenciones que se llevan a cabo en la ciudad se seleccionaron 16 personas y se les preguntó la cantidad que gastaban por día. Se obtuvo la siguiente información en dólares.

150, 175, 163, 148, 142, 189, 135, 174,
168, 152, 158, 184, 134, 146, 155, 160.

a) Estime la cantidad promedio real.
b) Estime la varianza de la cantidad gastada por día.

3.2.6. En cierta ciudad industrial se tomó una muestra aleatoria del número de empleados que hay por empresa. La muestra es:

75, 92, 517, 3200, 428, 117, 92, 50, 714.

a) Estime el número promedio de empleados.
b) Estime la varianza del número de empleados.
c) Estime la proporción de empresas que tienen más de 100 empleados.

3.2.7. Se espera tener una cierta variación aleatoria en el espesor de las láminas de plástico que produce una máquina. Para determinar cuándo la variación en el espesor se encuentra dentro de ciertos límites, cada día se seleccionan en forma aleatoria 12 láminas de plástico y se mide su espesor (en milímetros). Los datos que se obtuvieron en el día de hoy son los siguientes:
12.6, 11.9, 12.3, 12.8, 11.8, 11.7, 12.4, 12.1, 12.3, 12.0, 12.5, 12.9
Estime la varianza del espesor.

3.2.8. Una cadena de supermercados desea hacer eficiente su servicio de cajas. Para ello se han obtenido datos del número de clientes que llegan a una de sus tiendas de autoservicio entre las 11 y 12 hrs. del día. El registro de dichos datos es:

90, 48, 110, 75, 83, 94, 78, 93, 84, 108, 103, 79, 81, 93, 97, 89, 101, 87, 91, 79, 100, 93, 102, 90, 121, 78, 87, 91, 94, 112, 81, 78, 85, 94, 98, 113, 80, 91, 69, 100, 93, 97, 79, 77, 88, 103, 79, 91, 90, 83.

Determine la estimación puntual para la verdadera media y la varianza del número de clientes que llegan al supermercado.

3.2.9. Los siguientes datos se extrajeron de una muestra aleatoria de instituciones privadas de educación superior en México. Corresponden al número de alumnos de licenciatura inscritos en cada una de ellas.

Institución	Alumnos	Institución	Alumnos
UIC	2959	U. las Américas	4276
La Salle AC	5000	UIA	529
UIC	55	U. Cuauhtémoc	1618
UAG	12889	ITESMC (Sinaloa)	291
U.Anáhuac	5014	U.Mex.Nort.	149
CEUM	3118	ICEST	39

a) Estime el número promedio de estudiantes en las instituciones privadas.
b) Estime la varianza del número de alumnos en universidades privadas.

3.2.10. Una compañía publicitaria desea determinar si el número de anuncios comerciales de televisión (X) está relacionado con el número de ventas (Y) de cierto producto. Se obtuvieron los datos siguientes de una muestra aleatoria:

x	10	12	15	5	7	5	5	15	8	1	13	15
y	30	60	50	12	10	25	10	60	25	10	75	95

Estime el coeficiente de correlación.

INTERVALOS DE CONFIANZA

4.1.1. En el caso del intervalo de confianza para la media cuando σ^2 es conocida, ¿qué relación existe entre la longitud del intervalo de confianza y:

a) el coeficiente de confianza.
b) el tamaño de muestra.

4.1.2. Diga si cada una de las siguientes aseveraciones es verdadera o falsa. Justifique su respuesta:

a) A mayor tamaño de muestra, menor longitud del intervalo de confianza.
b) Un intervalo de confianza del $(1-\alpha)$ 100% para un parámetro está contenido en el correspondiente intervalo de confianza al $(1-\alpha')$ 100% con $\alpha > \alpha'$.
c) El intervalo es una gama de valores que se usan para estimar la forma de la distribución de una población.
d) La probabilidad de que un intervalo de confianza contenga al verdadero valor del parámetro recibe el nombre de nivel de confianza.
e) No es recomendable emplear altos niveles de confianza, pues producen intervalos de confianza más amplios.

4.1.3. Después de seleccionar una muestra y calcular el intervalo de confianza para μ, una persona dice: "Tengo una confianza del 88% de que la media de la población fluctúa entre 106 y 122".
¿Qué es lo que realmente está diciendo?

a) Hay una probabilidad de .88 de que μ fluctúe ente 106 y 122.
b) Hay una probabilidad de .88 de que el valor real de μ sea 114 (el punto medio del intervalo).
c) 88% de los intervalos calculados de las muestras de este tamaño contendrán la media de la población.
d) Todos los anteriores.

4.1.4. Una muestra de 100 votantes elegidos al azar de, entre todos los de un distrito dado, indicó que el 55% de ellos estaban a favor de un determinado candidato. Hallar los límites de confianza del 95% y 99%, para estimar la proporción de todos los votantes que están a favor de este candidato.

4.1.5. En el ejercicio 3.2.2, supóngase que el tiempo de espera es una variable aleatoria normalmente distribuida.
Calcule el intervalo de confianza del 99% para estimar la varianza.

4.1.6. Las pelotas de tenis que se utilizan en torneos profesionales deben pasar pruebas rigurosas para demostrar que la variación del rebote alrededor de un valor específico es mínima. Cuando éstas se dejan caer de una altura determinada, deben rebotar a una altura promedio de 4 pies. En una muestra aleatoria de 91 pelotas, la altura media en los rebotes fue de 4 pies y la varianza de 0.36.

a) Encuentre un intervalo de confianza del 90% para la varianza poblacional.
b) Encuentre un intervalo de confianza del 95% para la varianza poblacional.
c) Si la varianza máxima que se permite es de 0.5, ¿sería razonable decir que las pelotas de esta marca se pueden utilizar en torneos profesionales?

4.1.7. Se recibe un lote muy grande de artículos proveniente de un fabricante que asegura que el porcentaje de artículos defectuosos en la producción es del 1%. Al seleccionar una muestra aleatoria de 500 artículos, se encuentran 8 defectuosos. Obtenga los intervalos de confianza del 90% y 99% para la verdadera proporción de artículos defectuosos en el proceso de manufactura del fabricante. Con base en estos resultados, ¿qué se puede concluir con respecto a la información del fabricante? Compare los dos intervalos.

4.1.8. Una máquina produce piezas metálicas en forma cilíndrica. Se toma una muestra de éstas y los diámetros resultan de 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Obtenga un intervalo de confianza de 99% para el diámetro medio de las piezas producidas por la máquina.

Establezca los supuestos necesarios.

4.1.9. Un instrumento de precisión tiene como garantía el leer con un error máximo de una unidad. Una muestra de cuatro lecturas del mismo objeto dio como mediciones 353, 351, 351 y 355.

Calcule un intervalo de confianza del 90% para la varianza de la población. ¿Qué supuestos deben establecerse? ¿Es adecuada la garantía?

4.1.10. La Cámara de Comercio de una ciudad está interesada en estimar la cantidad promedio de dinero que gasta la gente que asiste a convenciones, calculando comida, alojamiento y entretenimiento por día. De las distintas convenciones que se llevan a cabo en la ciudad se seleccionaron 16 personas aleatoriamente y se les preguntó la cantidad que gastaban por día. Se obtuvo la siguiente información en dólares.
150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184, 134, 146, 155, 160.

a) Obtenga un intervalo de confianza del 98% para la cantidad promedio real. Establezca los supuestos necesarios.
b) Obtenga un intervalo de confianza del 95% para la varianza desconocida e interprete el resultado.

4.1.11. El fabricante de cierto tipo de cigarrillos asegura que éstos contienen en promedio 14 mg. de alquitrán. Se seleccionan 5 cigarrillos aleatoriamente y se obtiene que contienen 14.5, 14.2, 14.4, 14.3 y 14.6 mg. de alquitrán.

a) Construya un intervalo del 95% de confianza para la media del contenido de alquitrán por cigarrillo. Establezca los supuestos necesarios.
b) ¿Cuál es el límite máximo para el error de estimación del inciso anterior? Comente.

4.1.12. Considere la población constituida por los estudiantes de licenciatura del ITAM. Sean P, P_M y P_H la proporción total, la proporción de mujeres y la proporción de hombres respectivamente que, piensan que sí serán limpias las próximas elecciones.

Seleccione una muestra aleatoria simple de 40 mujeres y otra de 40 hombres.

a) Construya los intervalos del 90% y 95% de confianza para cada uno de los parámetros P, P_M y P_H.
b) ¿Podría decirse que el 80% de los estudiantes de licenciatura del ITAM piensan que sí serán limpias?

4.1.13. Se dice que un bolichista es consistente en su juego, si la desviación estándar en los puntos que obtiene es a lo más de 3.4. En una muestra tomada al azar de 41 juegos, un bolichista logró un promedio de 158 puntos con una varianza de 30.96.

a) Construya un intervalo del 98% de confianza para la varianza del bolichista.
b) ¿Cuáles son los requisitos necesarios para obtener el intervalo del inciso anterior?
c) ¿Se puede concluir que al 98% de confianza el bolichista no es consistente en su juego?

4.1.14. Un médico investigador piensa seleccionar una muestra aleatoria de personas que, hayan fumado por lo menos una cajetilla diaria durante 20 años, para saber cuántos llegan a desarrollar cáncer pulmonar en el transcurso de los próximos cinco años.

¿Cuál debe ser el tamaño de muestra que el investigador debe seleccionar de manera tal que, con una probabilidad de 0.95, la proporción muestral se encuentre a no más de 0.02 unidades de la proporción verdadera?

4.1.15. Al gerente de producción de una compañía procesadora de cítricos le interesa saber si, en los últimos tres años las heladas han dañado los 2500 naranjos que posee la empresa. A fin de averiguar el daño causado a los árboles, él mandó contar el número de naranjas producidas por árbol en una muestra de 42 de estos y obtuvo

una media muestral de 525 naranjas por árbol. La población de naranjas que ha venido manejando la empresa tiene una desviación estándar de 30 naranjas por árbol.

a) Construya un intervalo de confianza del 98% para la producción media por árbol de los 2500 naranjos. Interprete.

b) Si el rendimiento promedio de naranjas por árbol fue de 600 naranjas hace 5 años, ¿qué puede decir el gerente de producción sobre la posible existencia de un daño en este momento?

4.1.16. Una papelería recibe un embarque de cierta marca de bolígrafos. El propietario desea estimar la proporción de defectuosos que éste contiene y para ello, toma una muestra aleatoria de 300 plumas, de las cuales 30 tienen algún defecto. Establezca una estimación por intervalo, con 90% de confianza, para la proporción de bolígrafos defectuosos en el embarque.
Supóngase que se puede devolver el embarque si éste contiene más de 5% de defectuosos. Con base en los resultados de la muestra, ¿le recomendaría al dueño de la papelería regresar todos los bolígrafos?

4.1.17. Un lote de azulejos de cerámica es inspeccionado justo antes de ser embarcado. El gerente de control de calidad manda revisar 6 cajas completas, cada una con un contenido de 144 azulejos. Se encuentran los siguientes defectuosos: 2, 3, 6, 0, 4 y 9. Si los datos pueden considerarse como una muestra aleatoria de una población, construya un intervalo de confianza del 98% para el número promedio de unidades defectuosas por caja.
Establezca los supuestos pertinentes para ello.

4.1.18. El dueño de una inmobiliaria revisó aleatoriamente 150 de las 3000 cuentas de la Compañía y determinó que, el 60% de éstas estaban al corriente.

a) Encuentre un intervalo de confianza del 95% para la proporción de cuentas de la compañía que están al corriente.

b) En base al resultado del inciso a), estime por intervalo, el número total de cuentas de la inmobiliaria que están al corriente. (Utilice el mismo nivel de confianza de 95%).

4.1.19. Se seleccionó una muestra aleatoria de 21 ingenieros de un grupo más grande que labora para un fabricante de equipo electrónico y se les preguntó sobre el número de horas que trabajan semanalmente. La desviación estándar muestral fue de 7 horas.

Determine un intervalo de confianza del 90% para la varianza de las horas de trabajo para todos los ingenieros que laboran para el fabricante. (Supóngase que estas mediciones tienen una distribución normal).

4.1.20. Los empleados que trabajan para una compañía de mudanzas llevan las cajas al sitio, las ensamblan y empacan. La Compañía quiere una estimación del tiempo requerido para ensamblar un tipo especial de caja.

Se eligen aleatoriamente 16 situaciones en las cuales se toma el tiempo de ensamblaje de este tipo de cajas. Los resultados son $\sum X_i = 336$ segundos y $\sum X_i^2 = 7116$ segundos^2.

Determine un intervalo de confianza del 99% para el tiempo medio de ensamblaje.

4.1.21. Se sospecha que los anticonceptivos orales reducen el nivel de hemoglobina del usuario. En un estudio de sangre de una muestra de 37 mujeres que, reportaron estar tomando anticonceptivos orales, se obtuvo una media de 14.491 y una desviación estándar de 1.079 en el nivel de hemoglobina. Construya un intervalo de confianza del 95% para el nivel medio poblacional de hemoglobina de mujeres que toman anticonceptivos orales.

Si el nivel medio normal es de 14.9, ¿hay razón para sustentar la sospecha? Justifique su respuesta.

4.1.22. Se desea realizar un estudio sobre las utilidades que tienen las pequeñas industrias en México. En una primera etapa se decidió seleccionar una muestra aleatoria de

17 industrias en Guadalajara y otra del mismo tamaño en la Ciudad de México. Los datos fueron:

Utilidad mensual (millones)	
Guadalajara	Cd. de México
11.2	16.7
9.4	17.9
8.7	17.4
12.2	14.9
7.7	16.2
11.5	14.0
7.3	15.1
11.8	18.3
15.1	11.3
10.5	18.3
17.	17.1
15.6	12.0
13.2	15.2
17.2	16.2
14.6	14.9
12.2	10.5
9.8	16.5

a) Estime el valor medio y el coeficiente de variación de las utilidades para las industrias en Guadalajara y también para las de la Ciudad de México. Interprete.
b) Calcule los intervalos de confianza para el valor medio (98%) y la varianza (95%) de las utilidades tanto para las industrias de Guadalajara como para las de la Ciudad de México.
c) ¿Qué supuestos consideró en cada uno de los incisos anteriores?

4.1.23. Con el objeto de estimar la proporción de televidentes que han visto el anuncio de un producto, se entrevistó a 2300 telespectadores y resultó que, 1974 de ellos sí lo habían visto.

a) Encuentre un intervalo de confianza de 95% para la proporción de todos los espectadores que han visto la publicidad del producto.
b) Obtenga el tamaño de muestra indispensable para que el intervalo del inciso (a) tenga una longitud máxima de 10% con la misma confianza. Comente.

4.1.24. Los costos variables, principalmente de la mano de obra, hacen que los precios de la construcción de casas varíen de una edificación a otra. Un constructor de casas necesita tener una ganancia media por encima de $ 8,500 (dólares) por casa para alcanzar una ganancia anual establecida como meta. Las ganancias por casa para las cinco edificaciones más recientes del constructor son $8,760, $6,370, $9,620, $8,200 y $10,350, respectivamente. (Considere estas ganancias como una muestra aleatoria).

a) Encuentre un intervalo de confianza de 95% para el promedio de la ganancia del constructor por edificación. Interprete el intervalo.
b) Especifique el supuesto indispensable utilizado en (a).
c) Con la información del intervalo de (a), ¿sería razonable pensar que el constructor está trabajando al nivel de ganancia deseado?

4.1.25. En una muestra aleatoria de 100 estudiantes de cierto colegio, 60 de ellos opinaron que estaban a favor del pase automático a las universidades. Obtenga un intervalo de confianza del 95% para la proporción poblacional. Interprete este intervalo.

4.1.26. Un hombre de negocios está interesado en invertir en un instrumento que piensa le puede dar altos rendimientos. No obstante, sabe que, en general, a mayor rendimiento se tiene mayor riesgo. Al considerar instrumentos similares se observaron los siguientes rendimientos (%):

8.7, 15.5, 21, 18, 17.3, 22.1, 25

El inversionista considera que si se tiene un riesgo mayor a 10 (desviación estándar), no le conviene invertir.

Construya el intervalo de confianza del 95% adecuado y diga si este hombre hará la inversión o no, justificando su respuesta.

¿Qué supuestos fueron necesarios para resolver este problema?

DETERMINACIÓN DEL TAMAÑO DE MUESTRA

4.2.1. El gerente de un supermercado desea recabar información sobre la proporción de clientes a los que no les agrada una nueva política sobre aceptación de cheques. ¿Cuántos clientes se tendrían que incluir en una muestra si, se desea que la fracción de la muestra esté a lo más a 0.15 de la verdadera fracción con una probabilidad de 0.98?

4.2.2. Un contratista de pintores desea saber el tiempo promedio que tarda un grupo de trabajadores en pintar un departamento. Suponga una desviación estándar de 4 horas. ¿De qué tamaño debe ser la muestra para poder afirmar, con una probabilidad de 0.90, que la estimación tendrá un error de cuando mucho 3 horas? ¿Se necesitaría algún supuesto adicional para la población?

4.2.3. Se desea estimar el número promedio de horas por semana que los alumnos de nivel universitario dedican a estudiar. Suponga una desviación estándar de 4.5 horas. ¿Cuál es el tamaño de muestra necesario para afirmar, con una probabilidad de 0.90, que el tiempo promedio estimado dista del tiempo medio real en menos de media hora?.

4.2.4. Un productor de semillas que envasa sus productos de manera automática ha venido recibiendo quejas respecto al peso de los paquetes que vende. Para evitar posibles multas de las autoridades correspondientes decide tratar de verificar cuál es el verdadero peso medio del producto envasado por su máquina. La documentación de dicho aparato garantiza una desviación estándar de 0.2 Kg. ¿De qué tamaño debe ser la muestra que permita afirmar al productor, con una probabilidad de 0.95, que la estimación del valor promedio del peso de los paquetes dista del verdadero valor de dicho parámetro menos de 0.05 Kg?

4.2.5. Un guardabosques desea estimar el área promedio de la base de los pinos. Después de varios años de estudio, se observó que dichas áreas tienen una distribución normal con una desviación estándar de aproximadamente 4 pulgadas cuadradas.

a) Si se selecciona una muestra de 9 árboles, encuentre la probabilidad de que la media muestral difiera a lo más 2 pulgadas cuadradas de la media de la población.

b) Suponga que el guardabosques quisiera que la media de la muestra difiera a lo más 1 pulgada cuadrada de la media de la población con una probabilidad de 0.90.

¿Cuántos árboles tendría que medir para estar seguros de obtener tal grado de exactitud?

4.2.6. Se desea realizar un estudio para conocer la proporción de estudiantes que están en contra de un cambio en los reglamentos. ¿Cuántas personas se deberían incluir en una muestra si se desea que, con una probabilidad de 0.96, la proporción muestral no diste de la verdadera proporción en más de 0.15?

4.2.7. Un investigador social desea estimar la media de cierta población. Para ello quiere usar una muestra lo suficientemente grande para que la probabilidad de que la media muestral no difiera de la media poblacional por más del 25% de la desviación estándar, sea de .95. ¿De qué tamaño debe elegir la muestra este investigador?

4.2.8. El director administrativo de un colegio desea usar la media de una muestra aleatoria para estimar el tiempo que en promedio tardan los alumnos en ir de una clase a la siguiente y además quiere asegurar, con una confianza del 99%, que el error sea a lo más de 0.25 minutos. Si puede suponer por experiencia que $\sigma = 1.40$ minutos. ¿De qué tamaño debe seleccionar su muestra?

PRUEBAS DE HIPÓTESIS PARAMÉTRICAS

TEORÍA

5.1.1. Responda verdadero o falso y justifique su respuesta.

a) En una decisión se comete o el error tipo I o el error tipo II.

b) Es más fácil rechazar una hipótesis con una prueba de una cola que, con una prueba de dos colas.

c) Los tipos de errores son complementarios.

d) Si una hipótesis es rechazada con un nivel de significancia del 0.04, también será rechazada con 0.05 .

e) En una prueba de hipótesis de dos colas, la zona de no rechazo es equivalente al intervalo de confianza para el parámetro de interés.

f) β es el error tipo II .

g) (1- α) es conocida como el nivel de significancia de la prueba de hipótesis.

h) La única manera de disminuir simultáneamente la probabilidad de cometer los errores tipo I y tipo II es aumentando el tamaño de la muestra.

i) A mayor tamaño de muestra, mayor potencia de la prueba.

j) La potencia de la prueba mide la probabilidad de rechazar correctamente la hipótesis nula.

k) $\alpha + \beta = 1$.

l) α debe ser siempre más pequeña que β .

m) Con un menor nivel de significancia la probabilidad de rechazar una hipótesis nula que en realidad sea verdadera disminuye.

n) En la prueba de hipótesis se supone que algún parámetro de la población asume un valor particular antes de extraer la muestra. Esta suposición que debe probarse se llama hipótesis alternativa.

ñ) En una prueba de hipótesis, la estadística de prueba sigue siempre una distribución normal

o) Si se comete él error tipo I, se rechaza una hipótesis nula que es realmente verdadera.

p) Si las pruebas de hipótesis se efectuaran con un nivel de significancia de 0.60, la hipótesis nula generalmente no se rechazaría aunque no fuera verdadera.

q) Si las muestras son dependientes, entonces no se puede realizar una prueba de hipótesis para la diferencia de medias.

5.1.2. Defina los siguientes conceptos y establezca la relación que existe entre éstos:

a) Error tipo I

b) Error tipo II

c) α

d) β

5.1.3. Suponga que en un juicio la hipótesis nula es que un individuo es inocente. Analice los dos tipos de errores y comente sobre la importancia de ambos.

5.1.4. α es igual a .10 para una prueba de hipótesis, entonces se dice que:

a) 10% es el estándar mínimo de la probabilidad aceptable.

b) 10% es el riesgo que corremos de rechazar una hipótesis verdadera.

c) 10% es el riesgo que corremos de aceptar una hipótesis falsa.

PRUEBAS DE HIPÓTESIS PARAMETRICAS

Identifique en cada problema la variable de interés y el o los parámetros objetivo.

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

a) Calcule la probabilidad de cometer el error tipo I.
b) Calcule la probabilidad de cometer el error tipo II para p=0.6 y p=0.4.
c) Si la región de rechazo $\{Y \leq c\}$ encuentre el valor de c, tal que $\alpha = .01$.
d) Para región de rechazo $\{Y \leq 11\}$, encuentre β si p es igual a 0.6.

5.2.2. En una muestra aleatoria de 30 compañías manufactureras, con activos fijos por debajo de $10,000, se obtuvo una utilidad promedio del 1.8% con una desviación estándar de 0.4%.
En otra muestra seleccionada al azar de 40 compañías manufactureras con activos fijos entre $10,000 y $50,000, la utilidad promedio y la desviación estándar fueron de 2.4% y 0.6% respectivamente.
a) Con, $\alpha = .01$, pruebe que las varianzas son iguales.
b) ¿Con una significancia de 0.05, ¿se puede decir que la diferencia entre las utilidades medias de los dos tipos de compañía son diferentes?
¿Qué se requiere suponer sobre el comportamiento de la variable de interés?

5.2.3. La resistencia al rompimiento de cierta cuerda de alambre tiene un valor medio de μ=48000 libras. Al investigar la efectividad de una nueva aleación de acero, el fabricante desea probar la hipótesis nula de que μ = 48000 libras contra una alternativa adecuada.

a) ¿Qué hipótesis alternativa debe plantear el fabricante si no desea emplear la nueva aleación de acero a menos que sea definitivamente más resistente?

b) ¿Qué hipótesis alternativa debe emplear el fabricante si el nuevo proceso es en realidad más económico y el fabricante desea utilizar la nueva aleación de acero a menos que ésta disminuya en realidad la resistencia de la cuerda?

5.2.4. Un sociólogo registró el número de años cumplidos por una mujer al momento de su primer parto. Se clasificó a las mujeres de acuerdo al medio en que vivían, rural o urbano, y se obtuvieron los siguientes resultados:

		Años cumplidos al momento del primer parto	
Medio	Número de mujeres	Media	Desv. estándar
Rural	120	18.3	2.1
Urbano	60	21.2	4.3

El investigador sostiene que en el medio rural, debido a la mayor uniformidad en las costumbres, la variabilidad en la edad al momento del primer parto es menor que la que se tiene en el medio urbano. Establezca los supuestos necesarios y utilice la información para probar la hipótesis del investigador. Establezca claramente las hipótesis, encuentre el nivel de significación descriptivo aproximado y tome una decisión.

5.2.5. La experiencia ha mostrado que la resistencia de ciertos cables de acero es una variable aleatoria normal con $\sigma^2 = 1$. En una fábrica se desea probar que el la resistencia media poblacional es mayor que 10, en base a una muestra aleatoria de 25 cables,

a) Calcule la probabilidad de cometer el error tipo II cuando los cables, en realidad tienen una resistencia media de 11 toneladas. (Considere α =0.025).

b) Determine el tamaño de muestra tales que α=0.025 y β=0.05 cuando en realidad la resistencia media es de 11 toneladas.

5.2.6. En un estudio, 68 de 160 niños considerados en la muestra recibieron un tratamiento de fluoruro y tuvieron caries. Otro tratamiento químico dio como resultado que 38 de 110 niños, seleccionados al azar, tuvieran caries.

a) ¿Se puede afirmar, con un nivel de significancia igual a 0.05, que el tratamiento con fluoruro dio peor resultado que el tratamiento químico? Establezca claramente las hipótesis.

b) ¿Cuál es el nivel de significancia descriptivo o valor p?.

5.2.7. Un fabricante desea comparar el proceso actual de armado para uno de sus productos con un método propuesto que supuestamente reduce el tiempo de armado. Se les pidió a ocho trabajadores de la planta que armaran las unidades con ambos procesos.

Enseguida se presentan los tiempos observados en minutos.

trabajador	proceso actual	proceso propuesto
1	38	30
2	32	32
3	41	34
4	35	37
5	42	35
6	32	26
7	45	38
8	37	32

a) Con α =0.05, ¿existe alguna razón para creer que el tiempo de armado para el proceso actual es mayor que el del método propuesto por más de 2 minutos?.

b) ¿Cuál es el nivel de significancia descriptivo?

c) ¿Cuáles fueron los supuestos necesarios para resolver los incisos anteriores?.

5.2.8. En diversos anuncios se afirma que el contenido promedio de nicotina de cierto cigarrillo es 0.3 miligramos. Al sospechar que esta cifra es demasiado baja, una agencia de protección al consumidor toma una muestra al azar de 15 cigarrillos de diversos lotes de producción y advierte que su contenido de nicotina fue en promedio 0.33 miligramos, con una desviación estándar de 0.018.

a) Utilice un nivel de significación 0.05 para probar la hipótesis de que el nivel medio de nicotina es mayor a 0.3 miligramos.

b) Con $\alpha = 0.05$, obtenga β para $\mu = 0.32$

c) Contraste la hipótesis nula σ =0.01 contra la hipótesis alternativa de que esta cifra es demasiado baja. (Use α =0.05)

5.2.9. Una muestra aleatoria de tamaño 27 de una distribución normal bivariada, produjo un coeficiente de correlación muestral $\hat{\rho}$ = -0.45. ¿La hipótesis H_0 : ρ =0 puede ser rechazada a favor de H_1: $\rho \neq 0$ con un nivel de significancia de α =.05?.

5.2.10. Completa los espacios en blanco:

Un gerente de mercadotecnia, ante un política de reducción de gastos, tiene que decidir si continúa o no una campaña, considerando el impacto que ésta ha tenido en el público consumidor, Establece entonces las siguientes hipótesis:

H_0: La propaganda no afecta el nivel de ventas, por lo tanto, se puede prescindir de esta campaña.

H_1: El consumidor es muy sensible a la propaganda.

En este caso, el error tipo _____ consiste en un excesivo gasto publicitario y el error tipo _____ repercute en una disminución de las ventas.

5.2.11. Un establecimiento para lavado en seco afirma que un nuevo removedor puede eliminar más del 70% de las manchas a las cuales se les aplique. Para verificar esta aseveración, el producto se utiliza en 12 manchas que se escogieron al azar. En caso de que se eliminen menos de 11 manchas, no se rechazará la hipótesis nula; de lo contrario, se concluirá que si puede eliminar más del 70%.

a) Evalúe α

b) Evalúe β para la alternativa p=0.9

Si ahora, en lugar de utilizar el producto en una muestra aleatoria de 12 manchas, se prueba el producto en 100 manchas y se determina que, la región critica es X > 82 donde X es la cantidad de manchas removidas

c) Evalúe α

d) Evalúe β para la alternativa p=0.9

5.2.12. Cuando se juega boliche es a menudo posible que se tire bien en el primer juego y se tire pobremente en el segundo o viceversa.

Las siguientes seis parejas de números representan las puntuaciones del primer y segundo juego de la misma persona en seis lunes seleccionados aleatoriamente. Suponiendo que las puntuaciones siguen una distribución normal bivariada, pruebe que los resultados de los juegos están relacionados linealmente (Utilice $\alpha = .10$).

Juego 1	170	190	200	183	187	178
Juego 2	197	178	150	176	205	153

5.2.13. En un estudio de la efectividad del ejercicio físico en la reducción de peso, se seleccionaron 16 personas y se les sometió a un programa particular de ejercicio físico durante un mes, con los siguientes resultados:

Peso antes (libras)	Peso después (libras)
209	196
178	171
169	170
212	207
180	177
192	190
158	159
180	180
170	164
153	152
183	179
165	162
201	199
179	173
243	231
144	140

¿Es el programa de ejercicio físico efectivo para reducir el peso? Use $\alpha = 0.01$.

5.2.14. Se sabe que la proporción de artículos defectuosos en un proceso de manufactura es de 0.15. El proceso se vigila en forma periódica tomando muestras aleatorias de tamaño 20 e inspeccionando las unidades. Si se encuentran cuatro o más unidades defectuosas en la muestra, el proceso se detiene y se considera como "fuera de control".

a) Enuncie las hipótesis nula y alternativa apropiadas.
b) Obtenga la probabilidad del error del tipo I.
c) Obtenga la potencia de la prueba para los siguientes valores alternativos de la proporción de artículos defectuosos: 0.20, 0.25
d) Obtenga la potencia de la prueba para p = .20, p = .25 cuando la muestra inspeccionada es de 100 artículos y el proceso se detiene si hay 20 o más unidades defectuosas.

5.2.15. En un estudio reciente, se investigó el efecto de la ingestión de una cierta sustancia en el desarrollo de cáncer pulmonar. Se encontró que de 488 hombres que habían ingerido una baja cantidad de esta sustancia, 14 desarrollaron cáncer pulmonar. En un grupo del mismo tamaño en el que el consumo de esta sustancia era mayor, sólo 5 personas desarrollaron cáncer pulmonar. Bajo las suposiciones apropiadas.

a) ¿Pruebe si la ingestión alta de esa sustancia reduce el riesgo de cáncer pulmonar en los hombres? Emplee $\alpha = 0.01$.
b) Obtenga el nivel de significancia descriptivo o valor p

5.2.16. Supóngase que, un psicólogo piensa que la edad influye en el coeficiente de inteligencia (IQ). Se toma una muestra aleatoria de 100 personas de mediana edad, de quienes se conoce su IQ a la edad de 16 años y actualmente. De restar, los coeficientes de su juventud de los coeficientes actuales, se obtuvo una diferencia promedio de 6 puntos, con una desviación estándar muestral de 7 puntos.
Utilice $\alpha = .01$ para probar la hipótesis de que el IQ aumenta con la edad.

5.2.17. ¿Afecta la música de fondo la conducta de los compradores en los supermercados? Para contestar esta pregunta se realizó un experimento en un supermercado durante los meses de verano, que son relativamente estables en lo que se refiere a compras. Se seleccionaron dos días a media semana y se escogió aleatoriamente un día para suprimir la música de fondo. En el otro día escogido se tocó música a compás lento. Las ventas diarias (en millones de pesos) en 12 semanas se muestran a continuación:

Música	Semana											
	1	2	3	4	5	6	7	8	9	10	11	12
Suprimida	14172	15485	13922	12204	15501	15106	14608	13946	15002	14670	16202	13286
De compás lento	15917	16110	13818	14709	13982	16416	14727	14823	14825	15949	15488	14855

¿Proporcionan los datos evidencia suficiente que indique que el promedio de las ventas diarias de los días con música de compás lento es superior al promedio de las ventas diarias de los días sin música de fondo?.

a) Establezca las hipótesis y realice la prueba con un $\alpha = 0.10$.

b) Si con música de compás lento las ventas promedio superan en 1000 millones de pesos a las ventas promedio en los días en que no hay música; calcule la potencia de la prueba. (Considere $\alpha = 0.10$)

5.2.18. La estabilidad de las mediciones de las características de un producto manufacturado es importante para mantener su calidad. En realidad, es mejor tener, a veces, una pequeña variación en el valor medio de alguna característica importante de un producto y tener una media del proceso un poco fuera del "blanco", que tener una amplia variación con una media que se ajuste perfectamente a las necesidades. La última situación puede producir un porcentaje mayor de productos defectuosos que la primera. Un fabricante de lámparas eléctricas sospechó que una de sus líneas de producción estaba fabricando bombillas con una alta variación en su vida útil. Para probar su teoría comparó la vida útil de $n = 40$ lámparas, muestreadas al azar, de la línea que supuestamente no trabajaba bien, y $m = 40$ de la línea que parecía estar "bajo control". Las medias y las varianzas de las muestras para las dos líneas eran:

Línea sospechosa	$\bar{x}_1 = 1520$	$s_1^2 = 92000$
Línea "bajo control"	$\bar{x}_2 = 1476$	$s_2^2 = 37000$

a) ¿Proporcionan los datos evidencia suficiente para indicar que las bombillas (o focos), producidos por la línea que supuestamente trabaja mal poseen una mayor varianza en la vida útil que las lámparas producidas por la línea que supuestamente estaba bajo control? Utilice $\alpha = 0.05$.

b) Encuentre el nivel de significancia descriptivo que se ha observado para la prueba e interprete su valor.

5.2.19. El supervisor de control de calidad de una empresa automotriz quiere determinar la uniformidad del número de defectos en los automóviles que se producen en dos líneas de montaje. Si una línea presenta significativamente más variabilidad que la otra en el número de defectos, habrá que hacer cambios. El supervisor recabó los siguientes datos:

	Número de defectos	
	Línea de montaje A	Línea de montaje B
tamaño de la muestra	20	16
media	10	11
varianza	9	25

a) Pruebe si cada una de las varianzas poblacionales es distinta de 10

b) Si $\alpha = .05$, ¿Tiene la línea de montaje B más variabilidad en el número de defectos? Indique claramente si es necesario algún supuesto para realizar la prueba.

5.2.20. En una empresa se realizó un cambio en el plan de incentivos de los empleados y se desea saber si éste fue exitoso. Se seleccionaron 12 empleados y los siguientes datos muestran las tasas de trabajo defectuoso de los empleados antes y después del cambio en el plan.

Antes	8	7	6	9	7	10	8	6	5	8	10	8
Después	6	5	8	6	9	8	10	7	5	6	9	5

a) ¿Se puede afirmar que con el cambio disminuyeron las tasas de unidades defectuosas producidas?

b) Indique si es necesario algún supuesto para realizar la prueba.

5.2.21. La dirección del Banco del Ahorro desea saber si existe o no diferencia estadística en los promedios de ahorros mensuales de dos de sus sucursales (A y B). Para tal efecto se seleccionaron aleatoriamente 12 y 16 cuentahabientes de las sucursales A y B respectivamente. Cada persona entrevistada declaró el monto de sus ahorros en el último mes. Los datos obtenidos son los siguientes (en miles de nuevos pesos):

Sucursal A:

4.51, 1.4, 2.8, 0.75, 1.85, 5.4, 3.2, 2.75, 3.3, 0.9, 1.1, 1.9

Sucursal B:

3.2, 4.1, 3.8, 2.7, 3.1, 4.2, 2.3, 1.7, 1.9, 0.7, 3.3, 1.4, 4.0, 3.8, 4.3, 2.31

Suponiendo que ambas muestras provienen de poblaciones normales probar las siguientes hipótesis:

a) Las varianzas σ_A^2 y σ_B^2 de las poblaciones normales muestreadas son iguales (hipótesis nula) contra que dichas varianzas son distintas

b) El promedio de ahorro mensual de la Sucursal A es igual al de los ahorradores de la sucursal B (hipótesis nula) contra que el promedio en la sucursal B es mayor que el de la sucursal A. ¿Es necesario hacer alguna hipótesis particular acerca de las varianzas de cada población? Comente.

5.2.22. Un instructor hace un cuestionario formado por 10 preguntas falso-verdadero. Para probar la hipótesis de que el estudiante acierta por casualidad, adopta la siguiente regla de decisión: si el estudiante acierta 7 o más respuestas concluirá que no las atina por casualidad; en caso contrario, concluirá que el estudiante está adivinando.

a) Plantee las hipótesis adecuadas.

b) Hallar la probabilidad de rechazar la hipótesis nula siendo correcta.
c) Hallar la probabilidad de no rechazar la hipótesis nula cuando realmente:

i) $p = 0.6$ ii) $p = 0.7$ iii) $p = 0.8$ iv) $p = 0.9$

(Nota: np y nq < 5 por lo que NO se puede usar la aproximación Normal y debe usarse la distribución Binomial para calcular las probabilidades).

5.2.23. La Policía Federal de Caminos desea probar un radar detector de velocidades. Para esto hace circular frente al radar 61 patrullas a una velocidad de 110 Km/h. según su correspondiente velocímetro. De las 61 lecturas del radar obtiene $s^2 = 1.69$. Según la PFC el radar es aceptable si $\sigma < 2$km/h.
Adopte el punto de vista que el error grave es comprar el radar cuando en realidad el error de medición está por arriba del límite.

a) Formule las hipótesis H_0 y H_1 apropiadas a este problema.
b) Pruebe las hipótesis de a) con un nivel de significancia del 1%.
c) De una cota superior para el valor de P.
d) Verifique que la decisión de b) concuerda con la estimación de c).

5.2.24. Los fabricantes de un conocido refresco embotellado decidieron cambiar la fórmula de su producto. En vista de que un número importante de personas se han manifestado en contra del cambio, la gerencia de la embotelladora decidió llevar a cabo una encuesta. Por otra parte, la misma gerencia ha manifestado que "era normal que una actitud así se presente y que, sólo sería alarmante la situación, si más de la mitad de los consumidores están en contra del cambio".
De 937 personas entrevistadas, 531 se manifestaron en contra de la nueva fórmula.
Plantear el problema como una prueba de hipótesis y usar $\alpha = 0.05$.

5.2.25. Con respecto a una población muy grande de longitudes de barras metálicas cuya desviación estándar se supone es de 0.10 pulgada, un comprador desea demostrar (sobre la base de una muestra tomada al azar de barras de un lote) la hipótesis nula de que la longitud media real de la barra es de 4 pulgadas contra la hipótesis alterna de que, es menor a 4 pulgadas.

Las consecuencias de rechazar la hipótesis nula si ésta es verdadera y de no rechazarla si la longitud media es en realidad 3.95 pulgadas se consideran igualmente graves y sus riesgos se fijan en 0.02. (Es decir $\alpha = \beta = .02$) Encontrar el valor crítico y el tamaño de muestra necesario para que esto se cumpla.

5.2.26. El actual sistema aleatorio de control aduanero para usuarios del Aeropuerto Internacional de la Ciudad de México está basado en la suposición de que "únicamente el 10% de los viajeros provenientes del extranjero deberían pagar derechos por artículos que ingresan al país". Con objeto de verificar tal supuesto, regularmente se hace muestreo entre pasajeros.

En uno de tales muestreos, de 12 personas se detectaron 3 personas que estaban en posesión de objetos que debían pagar los derechos correspondientes.

a) ¿Es ésta una evidencia que permita pensar que actualmente la proporción de viajeros que deban pagar derechos aduanales es mayor del 10%? Utiliza un nivel de significancia aproximado de 0.02 .

b) Evalúa la potencia de la prueba cuando la proporción verdadera fuera de 20% y cuando fuera de 30%.

(Sugerencia: Observa que no es posible utilizar la aproximación a la normal, por lo que es necesario utilizar la distribución exacta de $\sum X_i$.

De manera intuitiva, si H_A: $p > 0.1$ fuera verdadera, el número de pasajeros debiendo pagar derechos debería ser "grande". Es decir, para rechazar H_0 se necesita que $\sum X_i$ exceda un cierto valor - digamos K_α que depende del nivel de significancia α).

PRUEBAS NO PARAMÉTRICAS

6.1. Luego de recabar datos sobre el grado de contaminación atmosférica en la Cd. de México, la SEMARNAP decidió emitir nuevas y estrictas normas para limitar la cantidad de hidrocarburos en el aire. Para el siguiente año se realizaron mediciones mensuales de este contaminante y se compararon con las del año precedente durante esos mismos meses. Basándose en los datos de la tabla siguiente, ¿tiene la SEMARNAP suficiente evidencia para afirmar con 95% de confianza que las nuevas normas lograron reducir la contaminación?

	E	F	M	A	M	J	J	A	S	O	N	D
X: Año anterior	7.0	6.0	5.4	5.9	3.9	5.7	6.9	7.6	6.3	5.8	5.1	5.9
Y: Año actual	5.3	6.0	5.6	5.7	3.7	4.7	6.1	7.2	6.4	5.7	4.9	5.8

* Partes por millón de hidrocarburos

6.2. Se probaron dos tipos de plásticos, cada uno producido mediante un procedimiento diferente para encontrar su resistencia de ruptura. Las mediciones en la tabla siguiente representan las cargas de rupturas en unidades de 1000 libras por pulgada cuadrada.

X: Plástico 1	15.3	18.7	22.3	17.6	19.1	14.8
Y: Plástico 2	21.2	22.4	18.3	19.3	17.1	27.7

a) ¿Presentan los datos evidencia de una diferencia entre las medianas de las distribuciones de las resistencias de ruptura para los dos plásticos? (Utilice $\alpha = .05$)

b) ¿Presentan los datos evidencia de una diferencia en las dispersiones? (Utilice $\alpha = .05$)

6.3. Un fabricante de coches encuesta a 120 distribuidores sobre qué hacer para promover las ventas. 60 opinaron que había que aumentar la publicidad, 40 opinaron que había que reducir precios y 20 fueron indiferentes. Pruebe, con $\alpha = .05$ si hay preferencia entre una proposición y la otra.

6.4. Un grupo de analistas de inversión clasificaron 10 compañías de acuerdo con su crecimiento potencial y el valor de sus acciones de la siguiente manera:

Compañía	1	2	3	4	5	6	7	8	9	10
Crecimiento	4	8	6	5	9	3	7	1	10	2
Valor de sus acciones	8	3	10	1	6	2	5	7	4	9

Ellos afirman que existe una relación evidente entre el valor de las acciones de la compañía y su crecimiento potencial. ¿Que diría usted a los analistas?

6.5. Para aumentar las ventas, durante los días de gran demanda, una cadena de tiendas que vende queso en centros comerciales obsequia muestras a la entrada del local. La gerencia de la cadena selecciona aleatoriamente los días en que se regalaron muestras. En la siguiente tabla se presentan las ventas en miles de pesos de los días de promoción y días sin promoción.

X: días de promoción	18	21	23	15	19	26	17	18	22	20	18	21
Y: días sin promoción	22	17	15	23	25	20	26	24	16	17	23	21

a) Pruebe que, regalar muestras en la entrada de las tiendas produce mayores ventas. Utilice una prueba no paramétrica.

b) Utilice una prueba no paramétrica para verificar si las dispersiones son iguales. (Considere $\alpha = .05$).

c) Verifique si las ventas en cada caso siguen una distribución normal utilizando papel normal.

6.6. A fin de determinar si existe relación entre el tipo de sangre y la severidad de los resfriados del invierno, se realizó un estudio y se obtuvieron los siguientes resultados:

	Tipo de Sangre			
	A	B	AB	O
resfriado severo	34	57	82	55
resfriado moderado	53	45	37	57
sin resfriado	213	198	181	188

¿Qué se puede concluir de los datos anteriores con respecto a la relación propuesta?

6.7. En una escuela se tomó una muestra aleatoria de tamaño 45, usando un padrón que incluye profesores y estudiantes. A cada individuo en la muestra se le preguntó si es profesor o estudiante y si considera si el nivel académico de la escuela es **deficiente (D), regular (R) o excelente (E).** Se desea dar respuesta a la pregunta ¿la opinión sobre el nivel académico es independiente de la categoria de los entrevistados?. A continuación se presentan los resultados:

OPINION

Profesor: D,D,D,D,D,D,D,R,R,R,R,R,R,E,E,E,E,E,E,E,E,E

Estudiante: D,D,D,D,D,R,R,R,R,R,R,R,R,E,E,E,E,E,E,E,E,E,E

a) Plantee la pregunta como un problema de prueba de hipótesis. Escriba en el contexto del problema las hipótesis nula y alterna.

b) Pruebe las hipótesis usando un nivel de significancia del 5%. Escriba su conclusión en el contexto del problema.

6.8. Se sospecha que una compañía lleva a cabo una política de discriminación con sus empleados la cual se refleja en los salarios percibidos. Se seleccionaron aleatoriamente, entre ellos, a 12 hombres y 12 mujeres. Sus salarios anuales, en miles de dólares, fueron los siguientes:

Hombres	22.5	19.8	20.6	24.7	23.2	19.2	18.7	20.9	21.6	23.6	20.7	21.6
Mujeres	21.9	21.6	22.4	24.0	24.1	23.4	21.2	23.9	20.5	24.5	22.3	23.6

a) ¿Presentan los datos suficiente evidencia para apoyar la sospecha? Utilice α =0.05.

b) Utilice una prueba no paramétrica para probar que, las dispersiones son iguales.

6.9. Se seleccionaron al azar once estudiantes de un grupo de estadística y se registraron sus calificaciones en dos exámenes sucesivos.

	Calificación	
Estudiante	Examen 1	Examen 2
1	94	85
2	78	65
3	89	92
4	62	56
5	49	52
6	78	74
7	80	79
8	82	84
9	62	48
10	83	71
11	79	82

a) Utilice alguna prueba no paramétrica para determinar si, el segundo examen fue más difícil que el primero.

b) Utilice ahora una prueba paramétrica. (Establezca los supuestos necesarios)

c) ¿Existe alguna relación entre las dos calificaciones?

6.10. Un psicólogo industrial desea probar si los tiempos de reacción de los operarios de una línea de ensamble son o no equivalentes con dos diferentes métodos de entrenamiento.

Nueve operarios son asignados aleatoriamente al método A y otros nueve al método B. Los siguientes datos representan la calificación desde 1 (más rápido) hasta 25 (más lento) de los tiempos de reacción para completar una tarea.

Método	
A	B
(X)	(Y)
2	1
3	6
4	8
9	15
10	16
14	17
19	21
20	22
23	25

Utilice una prueba no paramétrica para probar si hay diferencia en los tiempos de reacción (α =.05)

6.11. Considere los siguientes datos obtenidos de Estados Unidos en el periodo de 1975 a 1982.

Ingreso personal disponible por año

	1975	1976	1977	1978	1979	1980	1981	1982
Ingreso	1096.1	1194.4	1314.0	1474.0	1650.2	1824.1	2029.1	2173.4
Ahorro	94.3	82.5	78.0	89.4	96.7	106.2	130.2	142.7

a) ¿Proporcionan los datos evidencia suficiente para decir que el ahorro aumenta al incrementarse el ingreso personal disponible? (Utilice α = .05).

b) Calcule el valor p.

6.12. Se cuenta con una muestra aleatoria de 18 grandes empresas en México, y cada una de ellas se ha evaluado en términos de su importancia en el mercado de capital. Dos son las evaluaciones que tiene cada empresa: una realizada por el Gobierno y otra por el Centro de Estudios Económicos del Sector Privado.

a) Si para 12 de las 18 empresas el Gobierno dio una evaluación mayor que el Centro y sólo en dos casos se tuvo la misma evaluación, ¿es posible decir que el Gobierno y el CEESP difieren en su opinión con respecto a la importancia de las grandes empresas?

b) Considere que las evaluaciones fueron las siguientes:

Empresa	1	2	3	4	5	6	7	8	9
Gobierno	0.8	2	1.3	1.1	0.9	1.4	0.5	2.1	1.3
Centro	0.3	1.5	1.3	1	0.8	1.6	0.8	1.7	1

Empresa	10	11	12	13	14	15	16	17	18
Gobierno	0.6	1.8	1.4	2.1	1.1	0.9	1	1.2	1.
Centro	0.6	1.7	0.9	1.3	0.8	0.8	1.7	1.6	0.3

¿Existe relación entre ambas evaluaciones?

6.13. En un estudio se investigó la relación entre niveles de histamina e irritantes inhalados y alergénicos. Para esta parte del estudio se consideraron 22 fumadores, de los cuales 9 tenían síntomas de alergia y los demás no.
El contenido de histamina se midió en microgramas por gramo de peso seco de saliva. Los datos resultaron ser:

Alérgicos:
165.1, 112.0, 102.4, 100.0, 67.6, 65.9, 64.7, 39.6, 31.0

No Alérgicos:
48.1, 48.0, 45.5, 41.7, 35.4, 34.3, 32.4, 29.1, 27.3, 18.9, 6.6, 5.2, 4.7

a) Haga el diagrama de puntos para ambas muestras.
b) Utilice un método no paramétrico para probar la hipótesis de que, fumadores alérgicos y no alérgicos tienen niveles distintos de histamina.

6.14. Un gerente decide comparar la opinión del público al presentarle dos opciones para adquirir un mismo servicio.
Una muestra aleatoria de 12 personas dio los siguientes resultados:

persona	1	2	3	4	5	6	7	8	9	10	11	12
X: opción 1	R	M	B	B	M	M	B	B	R	B	B	R
Y: opción 2	R	R	M	R	M	R	M	B	B	M	R	M

Cada opción es etiquetada con M, R ó B dependiendo si se considera mala, regular o buena, respectivamente. ¿Existe alguna preferencia entre las opciones?

6.15. Un grupo de estudiantes, seleccionado en forma aleatoria entre todos aquéllos que llevan un mismo curso, fue clasificado según su calificación en esa materia en “superior al promedio” (S) y “promedio o inferior al promedio” (P-I). A estos mismos

estudiantes se les preguntó cuántas veces consultaron a sus profesores en su cubículo durante el semestre. Los resultados se presentan a continuación.

ALUMNO	CALIFICACION	No. DE CONSULTAS
1	S	8
2	P-I	1
3	S	5
4	P-I	2
5	P-I	13
6	S	12
7	P-I	7
8	P-I	3
9	S	9
10	P-I	2
11	S	7
12	S	10
13	P-I	0
14	P-I	1
15	P-I	4
16	P-I	0
17	S	15
18	S	11
19	P-I	0
20	P-I	6
21	P-I	10
22	S	4
23	P-I	1
24	S	16
25	P-I	9

a) ¿Cuáles son las variables de interés y en qué escala están medias cada una? (Explique).
b) ¿Presentan los datos de la tabla suficiente evidencia que permita concluir que los alumnos clasificados como “superior al promedio” (S) tienden a consultar más a

sus profesores, en sus cubículos, que los alumnos clasificados como “promedio o inferior” Utilice un nivel de significancia $\alpha = 0.05$.

c) Con base en estos datos, el número promedio de consultas al mes es de 6.24. Clasifique a los alumnos de acuerdo a su posición arriba y abajo del número promedio de consultas y, por otro lado, de acuerdo a su calificación (S) o (P-I) y pruebe si hay dependencia entre la calificación del alumno y las consultas que le hace al profesor. (Utilice el nivel de significancia descriptivo).

d) ¿Se satisfacen todas las condiciones para realizar la prueba de indepencia del inciso anterior?

6.16. El Gerente de una Empresa de productos electrónicos ha reunido las siguientes Estadísticas de sueldos respecto a los ingresos de su fuerza de ventas.

Ganancias en Miles

	25-30	31-36	37-42	43-48	49-54	55-60	61-66
Frecuencia Observada	9	22	25	30	21	12	6

Utilizando papel normal investigue si la distribución de sueldos es normal.

6.17. Durante cinco años se llevó a cabo un estudio para determinar si existe alguna diferencia en el número de resfriados que sufren los fumadores y los no fumadores. Con base en muestras aleatorias de 10 no fumadores y 8 fumadores se observaron, a lo largo de los cinco años los siguientes datos.

No fumadores	1	0	2	7	3	1	2	2	4	3
Fumadores	4	2	6	5	8	10	8	7		

Determine si existe alguna razón para creer que el fumar influye en la propensión a resfriarse con $\alpha = 0.05$.

a) Mediante una prueba no paramétrica.

b) Mediante una prueba paramétrica.

(Establezca los supuestos necesarios en cada inciso).

6.18. Un almacén de calzado utiliza cuatro fabricantes de zapatos para caballeros. Cada par de zapatos es inspeccionado en el almacén antes de ser vendido. Hay tres defectos diferentes que causarían la devolución de los zapatos al fabricante. Los fabricantes alegan que los defectos son independientes de quien provee el producto.

Considerando que en los tres meses pasados se registraron las siguientes tabulaciones.

	Defectos		
Fabricante	B_1	B_2	B_3
A_1	17	10	13
A_2	10	10	10
A_3	18	15	17
A_4	15	5	10

a) Pruebe si existe independencia entre fabricante y defectos. Use α =0.01.

b) ¿En qué cola de la distribución está la región de rechazo? ¿Por qué?

6.19. Una planta manufacturera de ropa recibe tela de dos proveedores (A y B).

A una medida del desperdicio de tela se le denomina índice de desperdicio (ID). El ID se mide como porcentaje semanal de desperdicio relativo al desperdicio que se obtendría del despliegue de los patrones sobre la tela que hace la computadora. Como los cortadores de la planta algunas veces pueden ganarle a la computadora reduciendo el desperdicio desplegando los patrones a mano sobre la tela, el ID puede ser negativo.

Los siguientes datos corresponden a los valores observados de ID para los proveedores A y B en un cierto período.

Proveedor A		Proveedor B	
0.12	0.03	1.21	0.98
1.01	0.35	0.97	0.65
-0.20	-0.08	0.74	0.57
0.15	1.94	-0.21	0.51
-0.30	0.28	1.01	0.34
-0.07	1.30	0.47	-0.08
0.32	4.27	0.46	-0.39
0.27	0.14	0.39	0.09
-0.32	0.30	0.36	0.15

-0.17	0.24	0.96
0.24	0.13	

a) Utilice la gráfica normal para visualizar si los datos de cada proveedor provienen de distribuciones normales.

b) Compare las distribuciones de los ID de cada proveedor utilizando diagramas de punto y diagramas de caja.

c) Utilice una prueba no paramétrica para, analizar si hay diferencias en las locaciones de los ID de los proveedores (obtenga el valor *p*).

d) Compare los resultados de los niveles de significancia empíricos (valores *p*) de la prueba F y la prueba de la variante de Mann-Whitney para analizar si hay diferencias en las dispersiones de los ID de los proveedores.

6.20. Una compañía distribuidora de artículos de oficina efectúa una campaña de promoción de su mercancía. Los siguientes datos representan las ventas (en miles de nuevos pesos) correspondientes a una muestra de 23 clientes para el mes en el que se realizó la promoción y para el mes inmediato siguiente.

antes	después	Antes	después	antes	después
1.14	0.73	1.33	1.24	0.66	0.14
2.33	2.53	1.49	0.20	2.16	3.45
5.40	10.12	0.53	0.34	2.54	2.54
2.40	5.27	1.53	1.54	2.24	0.31
0.57	1.21	1.44	0.47	2.54	2.52
0.69	0.37	2.25	2.06		
1.58	1.10	1.20	0.96		
2.75	2.64	0.64	0.16		
0.90	0.42	2.13	2.11		

a) Utilice la gráfica normal para visualizar si los datos de antes y después de la campaña provienen de distribuciones normales.

b) Analice las diferencias de las ventas mediante un diagrama de caja.

c) Utilice una prueba no paramétrica para analizar si hay un efecto significativo de la campaña sobre las ventas (obtenga el valor *p*).

6.21. Se desea probar que a las sucursales con un número de órdenes grande les corresponden en general, mayores cantidades de volumen de facturación total.
Con ese fin considere los siguientes datos

Sucursal	número de órdenes	facturación (miles de $)
1	515	32.50
2	286	26.80
3	469	11.41
4	410	14.85
5	461	32.64
6	436	23.82
7	479	29.84
8	198	21.83
9	389	24.65
10	262	22.50
11	536	36.00

a) Construya un diagrama de dispersión que represente a estos datos.
b) Utilice una técnica no paramétrica y el nivel de significancia descriptivo para probar las hipótesis.

6.22. Un fabricante de computadoras introduce una modificación en la arquitectura del equipo que mayor número de ventas reporta. Para averiguar el impacto que tuvo la modificación, lleva a cabo un estudio consistente en preguntarle a una muestra aleatoria de sus clientes de tamaño 14, su opinión sobre el cambio que introdujo en las computadoras, antes y después de haberles mostrado el nuevo equipo. El fabricante espera que la modificación complazca a sus clientes.
El fabricante hace la siguiente afirmación y pide que se califique de acuerdo a la siguiente escala:

"LA ARQUITECTURA DEL EQUIPO ES SOBRESALIENTE"

De acuerdo con la afirmación (1)
Indiferente con la afirmación (0)
En desacuerdo con la afirmación (-1)

Los resultados se muestran a continuación:

CLIENTE

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ANTES	0	0	-1	-1	0	0	1	0	0	0	0	-1	0	-1
DESPUES	1	0	1	0	1	-1	-1	1	1	-1	1	1	-1	1

a) Identifique la variable respuesta e indique su escala de medición.

b) Plantee el problema como una prueba de hipótesis. Escriba la hipótesis nula y alterna en contexto del problema.

c) Calcule la estadística de prueba así como el "valor p" o valor de significancia descriptivo.

d) Escriba su conclusión en el contexto el problema si se eligió una $\alpha = 0.06$.

MUESTREO

TEORIA

7.1.1. Supóngase que se desea usar muestreo para estimar el número total de palabras en un libro que contiene ilustraciones.

a) ¿Hay algún problema en la definición de la población?
b) ¿Cuáles son las ventajas y desventajas de usar como unidad de muestreo la página?
c) ¿Cuáles son las ventajas y desventajas de usar como unidad de muestreo la línea?

7.1.2. ¿De qué clase de muestreo se trata en cada uno de los siguientes casos?

a) Un reportero se para enfrente de un supermercado y a cada persona que pasa le pregunta por cuál candidato va a votar en las siguientes elecciones.
b) Todos los alumnos del ITAM se listan por clave única. Se selecciona al azar uno de los primeros 20 nombres y después de éste, cada 20 se selecciona uno. Si hay 2000 alumnos, ¿de qué tamaño va a ser la muestra?
c) Se obtienen 150 números aleatorios de 5 cifras y se eligen aquellos alumnos con matrículas similares.
d) Para un estudio demográfico se seleccionan 60 manzanas del D.F. aleatoriamente.
e) Se seleccionan 50 personas de cada clase social en el Distrito Federal.

7.1.3. a) Comente las ventajas y desventajas de las preguntas abiertas y cerradas en un cuestionario.
b) Dé un ejemplo adecuado para cada uno de los siguientes métodos de recolección de datos: entrevista personal, entrevista telefónica, cuestionario por correo y observación directa.

7.1.4. Suponga que decide realizar una encuesta telefónica sobre opinión pública, muestreando aleatoriamente números telefónicos del directorio. La encuesta se

realiza de las 9:00 A.M. a las 5:00 P.M. ¿Representarán las respuestas resultantes una muestra aleatoria de la opinión del público adulto en la comunidad? Explique.

7.1.5. Si tuviera que realizar un estudio sobre cuál es el perfil que se desea de un estudiante recién egresado de la universidad con respecto a las necesidades del país:

a) ¿Es necesario el uso de técnicas estadísticas? ¿Por qué?
b) En el caso de realizar un muestreo, ¿qué diseño sería el adecuado y qué variables podrían ser de interés?
Justifica lo mejor posible.

7.1.6. Defina lo que es el muestreo, muestreo probabilístico, muestreo aleatorio simple, muestreo estratificado y muestreo por conglomerados.

MUESTREO

PROBLEMAS

7.2.1. La Cía. Nielsen ha instalado monitores electrónicos en aparatos de T.V. de aproximadamente 1200 de los 80 millones de hogares en USA. Los datos obtenidos de los monitores registran estimaciones de la proporción de hogares que sintonizaron un programa de TV en particular. Nielsen presenta la siguiente justificación para este tamaño de muestra:

"Mezcle 70,000 frijoles blancos y 30,000 frijoles rojos y luego extraiga una muestra de 1000 frijoles. Las probabilidades matemáticas del número de frijoles rojos de estar entre 270 y 330, o sea del 27% al 33% de la muestra (lo que corresponde a una clasificación de 30, más o menos tres) son de 95%. La regla estadística básica no cambiaría si el muestreo proviniera de 80 millones de frijoles en lugar de 100,000".

Interprete y justifique esta declaración en términos de los resultados estudiados en el capítulo de muestreo de su curso de Estadística II.

7.2.2. El gerente de un taller de maquinaria desea estimar el tiempo promedio que necesita para terminar una tarea sencilla. El taller tiene 98 operadores. Se seleccionaron aleatoriamente 8 operadores y se les tomó el tiempo. Los resultados son (en minutos):

4.2, 5.1, 7.9, 3.8, 5.3, 4.6, 5.1 y 4.1

a) Estime el tiempo promedio para terminar la tarea, y establezca un límite para el error de estimación. Interprete.

b) Si se desea estimar el tiempo promedio con un error máximo de 0.5 minutos, ¿de qué tamaño debería tomarse la muestra?

7.2.3. Un dentista está interesado en la efectividad de una nueva pasta dental. Un grupo de 1000 niños de escuela participan en el estudio. Los registros de un estudio anterior mostraron que había en promedio 2.2 caries cada seis meses para el grupo. Después de tres meses de iniciado el estudio, el dentista muestreó 10 niños para determinar cuánto habían progresado con la nueva pasta dental y los resultados fueron los siguientes:

niño	1	2	3	4	5	6	7	8	9	10
número de caries	0	4	2	3	2	0	3	4	1	1

a) Estime el número medio de caries para el grupo y obtenga un límite para el error de estimación.
b) Estime el número total de caries para toda la población estudiada. Encuentre una cota para el error de estimación.
c) Estime la proporción de niños con más de 2 caries. Establezca una cota para el error de estimación.
d) Determine el tamaño de muestra requerido para estimar el número medio de caries con una cota en el error de, 20% menos que la obtenida en el inciso a).
e) Determine el tamaño de muestra requerido para estimar la proporción de niños con más de dos caries, con una cota en el error de, 30% menos que la obtenida en c).

7.2.4. Los resultados de una encuesta establecen que el 36% de los norteamericanos registra al fútbol como su deporte favorito. Un artículo establece: "Con una muestra de ese tamaño (1091 adultos) se puede decir con 95% de seguridad que, los resultados están dentro de más menos 3% de lo que estarían si la población completa de adultos hubiera sido encuestada". ¿Está usted de acuerdo?

7.2.5. Una muestra irrestricta aleatoria de 100 medidores de agua es controlada dentro de una comunidad para estimar el promedio de consumo de agua diario por casa, durante un período estacional seco. La media y la varianza muestrales fueron $\overline{Y} = 12.5$ y $S^2 = 125.2$. Si se supone que hay 10,000 casas dentro de la comunidad.

a) Estime μ el promedio de consumo diario verdadero y establezca un límite para el error de estimación.
b) Estime el total de galones de agua τ usado diariamente durante el período seco y establezca un límite para el error de estimación. Interprete el resultado.

7.2.6. Se seleccionó una muestra aleatoria de n = 50 de una población. La media y la varianza de la muestra eran $\overline{Y} = 84.1$ y $S^2 = 22.44$.

a) Calcule un intervalo de confianza del 95% para μ , suponiendo que el número de elementos (N) en la población es muy grande.

b) Calcule un intervalo de confianza del 95% para μ , suponiendo que el número de elementos (N) en la población sea igual a 100.

c) Compare los dos intervalos de confianza de los incisos (a) y (b) y observe el efecto del factor de corrección para población finita sobre la longitud del intervalo del inciso (b).

7.2.7. Un psicólogo desea estimar el tiempo de reacción promedio para un estímulo entre 200 pacientes de un hospital especializado en trastornos nerviosos. Una muestra irrestricta aleatoria de 20 pacientes fue seleccionada y fueron medidos sus tiempos de reacción con los resultados siguientes:

$\bar{y} = 2.1$ segundos $\qquad$ $s = 0.4$ segundos

Estime la media con un intervalo de, aproximadamente el 95% de confianza. Interprete el resultado.

7.2.8. En una encuesta de opinión se interrogó a 1684 adultos en todo México. En la encuesta se reportó tener 6% de margen de error para los resultados referentes a las proporciones de entrevistados que favorecen ciertas decisiones. ¿Está usted de acuerdo con esa cifra del 6%? (Suponga muestreo aleatorio irrestricto).

7.2.9. Una población consiste de N=100 escuelas primarias rurales en Yucatán. Se seleccionó una muestra aleatoria de n=15 escuelas para conocer el número de alumnos en cada una de ellas. Los resultados obtenidos se muestran a continuación: 127, 206, 185, 170, 294, 175, 98, 130, 142, 104, 153, 149, 133, 214, 147.

a) Calcule el número medio de alumnos por escuela.

b) Calcule Var(X) y estime la cota del error.

c) Estime el total de alumnos de dicha población y la Var($\hat{\tau}$).

d) Estime la cota del error para el total de la población.

7.2.10. Los siguientes datos representan una muestra aleatoria de 5 empresas mexicanas tomadas de un total de 50 empresas:

EMPRESA	PERSONAL EMPLEADO	PERSONAL ADMINISTRATIVO
Hipocampo, S. A. de C.V.	425	204
Hotel Calinda Cancún	499	190
Zinc de México, S. A.	623	50
Gelcaps Exportadora de México	147	18
Galletera Palma, S.A. de C.V.	212	70

a) Estime el número promedio de personal empleado ocupado en 1989 en las 50 empresas, dando una cota para el error de estimación.

b) Estime el total de personal administrativo ocupado en 1989 para las 50 empresas pequeñas dando una cota en el error de estimación.

7.2.11. Una encuesta realizada en 1994 mostró que un quinto de las 2207 personas encuestadas falsearon información en su declaración del impuesto federal al ingreso. Calcule la cota del error de estimación. ¿Considera usted que la fracción muestral es cercana a la proporción real que ocultó información?

7.2.12. Una encuesta realizada por Group Attitudes Inc. fue hecha para medir las actitudes de estadounidenses hacia la universidad. La empresa que realizó la encuesta envió cuestionarios por correo a 4200 personas en todo Estados Unidos y recibió 1188 respuestas. Alrededor del 55% de los encuestados dijo haber tenido grandes problemas para poder pagar la universidad de sus hijos.

Consideraría usted esta cifra como altamente confiable y representativa de la verdadera proporción de estadounidenses con este problema?

EJERCICIOS CON BASES DE DATOS

8.1. La información de la base de datos ***FLOTAMI***, corresponde a una muestra aleatoria de 600 embarcaciones de la Flota Mayor Mexicana. Esta muestra fue obtenida de una población constituida por 3,262 barcos obtenida por la SEMARNAP.

A continuación se listan las variables de interés para cada barco de la muestra:

i. **Entidad**: la entidad federativa donde se encuentra registrada (1,2,…,31)
ii. **Oficina**: la oficina de pesca, perteneciente a la entidad federativa en donde está registrada la embarcación.
iii. **Pesquería**: el tipo de pesca al que se dedica cada embarcación: H camarón, B atún, Z tiburón, F especies de escama y D sardina
iv. **Anio**: año de construcción de la embarcación.
v. **Eslora**: longitud de la embarcación medida de proa a popa (mts)
vi. **Manga**: anchura de la embarcación (mts)
vii. **Calado**: profundidad a que llega la quilla (base) del barco en metros
viii. **Arqueo**: arqueo o cabida de la embarcación en toneladas
ix. **Bodega**: capacidad de bodega en toneladas (sólo el espacio reservado como bodega)
x. **Carga**: capacidad de carga en toneladas.
xi. **Casco**: tipo de casco

a) Para cada una de las variables, indique: el tipo de datos que se tienen, su escala de medición, el (los) parámetro(s) de interés y su(s) correspondiente(s) estimadore(s).
b) Realice un diagrama de dispersión para las variable *capcarga* y *eslora*. Pruebe si existe una relación lineal entre estas variables.
c) Lo mismo que en el inciso b), pero para las variables *eslora* y *manga*.
d) Pruebe, con un nivel de significancia del 10% si ha relación lineal entre las variables *aniocons* y *eslora*.
e) Construya la tabla de contingencia para las variables *pesqueria* y *casco*. Pruebe si hay independencia entre este par de variables. Obtenga sus conclusiones.
f) Realice las pruebas de hipótesis convenientes para comparar la capacidad de carga de los barcos que fueron construidos antes del año de 1980 y los posteriores a dicho año.
Sugerencia: en el caso del Statgraphics es necesario considerar a las variables como capcarga select (aniocons<=1980) y capcarga select (aniocons>1980)

g) Explore, mediante la gráfica normal, si la variable *bodega* sigue una distribución normal. Además obtenga las medidas descriptivas adecuadas para esta variable. En base a lo obtenido anteriormente, realice una prueba de hipótesis paramétrica o no paramétrica, relacionada con la localización y la dispersión de esta variable.

8.2. La base de datos **KALOSHA**, fue obtenida del libro Estadística Básica en Administración de los autores Berenson M. y Levine D.

Kalosha Industries es una fábrica de partes automotrices que cuenta con 9800 empleados. La base está constituida por los datos muestrales de 400 empleados que fueron seleccionados aleatoriamente. A estos trabajadores se les encuestó para obtener información sobre las siguientes variables:

i) IDNUM: número de encuestado
ii) WORKHRS: horas de trabajo
iii) OCCUP: ocupación
iv) AGE: edad
v) EDUC: años de escolaridad
vi) SEX: género
vii) EARNRS: número de aportadores
viii) RINCOME: ingresos del encuestado
ix) FINCOME: ingresos familiares
x) STATJOB: satisfacción con el trabajo
xi) RICHWORK: retiro si se enriquece
xii) JOBCHAR: características del trabajo
xiii) GETAHEAD: avanzar
xiv) TRAUMA: eventos traumáticos
xv) MEMUNION: miembro de sindicato
xvi) WRKYEARS: años trabajados
xvii) EMPYEARS: años en Kalosha Industries
xviii) NUMPROMO: Número de promociones
xix) FUTPROMO: Promoción futura
xx) SEXPROMO: Oportunidades de promoción
xxi) ADVANCES: Progreso de avances
xxii) IDECIDE: Toma de decisiones
xxiii) ORGMONEY: Decisiones presupuestarias
xxiv) PROUDORG: Orgullo de Kalosha Industries
xxv) STAYORG: Permanecer en la organización

xxvi) UNMANREL: Sindicato- gerencia
xxvii) WOWRKEREL: Relaciones de compañeros
xxviii) SCHOOLGN: Escolaridad formal
xxix) TRAINING: Capacitación formal

(Se anexa a este ejercicio, una copia del cuestionario que se les aplicó a los empleados)

a) Suponga que usted ha sido contratado por el Director de recursos humanos de la empresa Kalosha con el objeto de hacer inferencias respecto a la población de los empleados Mediante la construcción de los respectivos intervalos de confianza del 95%, escriba un resumen en el cual analice sus hallazgos basándose en:

- La edad promedio
- El número promedio de años trabajado como empleado de tiempo completo.
- El número promedio de años empleados por Kalosha Industries
- La proporción de empleados que están clasificados como administrativos o profesionales.
- La proporción de empleados que dejarían de trabajar y se retirarían si se volvieran muy ricos.
- La proporción de empleados para quienes disfrutar de su trabajo es lo más importante
- La proporción de empleados que sienten que tienen una probabilidad alta o moderada de ser promovidos en los siguientes cinco años.
- La proporción de empleados cuyo trabajo les permite participar en la toma de decisiones todo el tiempo o parte de éste.
- La proporción de empleados que están muy orgullosos o algo orgullosos de trabajar para la compañía Kalosha.
- La proporción de empleados que describen las relaciones entre compañeros de trabajo o colegas, como muy buenas o buenas.

b) Supóngase que, la encuesta a los empleados de Kalosha se desea utilizar con el objeto de tomar la decisión de instrumentar un paquete de prestaciones para lo que, el Director General desea determinar:
   - Si existe evidencia de que la edad promedio de todos los empleados es menor de 40 años y

- Si existe evidencia de que la media de los ingresos familiares (totales) de los empleados es mayor a $39,000.

Explore, mediante la gráfica normal, si los datos de estas variables se distribuyen normalmente. Realice las pruebas de hipótesis adecuadas para contrastar lo que el Director desea probar.

c) ¿Existe evidencia de una diferencia de género con respecto a la cantidad promedio de tiempo (en años) que los empleados de tiempo completo han estado trabajando en Kalosha Industries?

d) ¿Se pudiera aplicar alguna prueba para muestras relacionadas con esta base de datos? Justifique su respuesta.

e) ¿Existe evidencia de que hay dependencia entre el sexo de los empleados y la ocupación escogida por ellos? (veáse preguntas 5 y 12 de la encuesta)

f) ¿Existe evidencia de que hay dependencia entre la importancia característica del trabajo y su satisfacción en éste? (veáse preguntas 11 y 9 de la encuesta)

8.3 Un analista financiero desea comparar las ganancias por acción (US dólares norteamericanos) en dos tipos de sectores - transporte y construcción-, con la finalidad de asesorar a sus clientes en su cartera de inversión. Los siguientes datos, relativos a las ganancias por acción fueron obtenidos de una muestra aleatoria de semanas de los años de 1997 y 1998. (Archivo Inversión)

TRANSP 1997	TRANSP 1998	CONSTR 1997	CONSTR 1998
8.1376	5.51742	7.4993	2.73049
6.6435	4.4706	6.17989	3.75221
7.5044	5.20668	2.2111	1.59407
3.0562	7.13943	2.89531	3.56927
10.8879	3.17597	6.01796	1.36178
10.0958	5.24941	1.97271	1.12844
5.1841	5.79179	4.42883	1.79665
5.0049	7.50518	2.69684	3.60381
7.0388	5.23778	2.35835	3.86575
8.0887	7.49034	3.76644	2.69205

5.4307	3.05732	0.14752	3.60339
3.6302	3.60018	3.69576	1.79703
7.2412	5.29199	2.00531	2.34153
7.6389	5.99691	1.56859	5.64959
4.5293	2.18445	7.77697	4.36794
4.3252	2.63824	4.29592	0.86965
7.4561	5.71824	2.64431	4.35386
7.4065	5.9645	3.91015	3.2481
3.4152	4.78038	3.50602	3.33365
5.0187	5.96256	4.62617	2.83937
9.1518	4.99567	3.15768	2.4862
3.1573	4.76436	3.72822	0.73701
6.3361	1.93907	6.2085	2.63982
3.8151	5.89071	5.82993	3.81253
5.6751	3.69666	2.99815	6.91579
4.7985	3.46943	4.6829	6.13502
6.3292	5.7579	2.72694	4.4148
6.1712	4.39056	3.90553	5.28549
5.2211	3.49841	5.10024	4.55339
8.6156	7.45382	2.85741	1.61201

Mediante un paquete de cómputo estadístico y suponiendo que las variables siguen una distribución normal:

a) Construya el intervalo de confianza del 95% para la media poblacional de las ganancias por acción para el sector de la construcción correspondiente al año de 1998.
b) Construya un intervalo de confianza del 95% para la varianza poblacional de las ganancias por acción del sector transporte para el año de 1997.

En los siguientes incisos, mediante el nivel de significancia descriptivo pruebe que:

c) La media poblacional de las ganancias por acción del sector construcción para el año 1997 es menor de $4.00 US dólares.
d) La desviación estándar poblacional para las ganancias por acción del sector de la construcción en 1997 fue mayor que $1.45 US dólares.

e) Las dispersiones de las ganancias por acción son diferentes, para ambos sectores, en el año de 1998.
f) La media poblacional de las ganancias por acción, en el sector transporte, fue significativamente mayor que la media poblacional de las ganancias por acción del sector construcción en el año de 1997.
g) La media poblacional de las ganancias por acción del sector de la construcción aumentó en el año de 1998 con respecto al año de 1999

8.4 Los siguientes datos (dat-gina.xls) corresponden a la información obtenida de una muestra aleatoria de 60 mujeres que tuvieron su bebé, por parto natural, en el período de abril a agosto de 1998 en un hospital de la ciudad de México.

Individuo	Edad	Fecha	No. partos	Psicoprofiláctico	Oxitocina	Sexo bebé	Leche materna(meses)	Años de escolaridad	Ingreso familiar
1	34	12-Apr	7	no	si	f	18	2	$4,500
2	30	22-Apr	2	no	no	m	0.5	6	$10,000
3	30	27-Apr	1	si	no	f	6	14	$25,000
4	27	28-Apr	3	no	no	m		14	$53,000
5	27	2-May	2	no	si	m	1	12	$35,000
6	30	3-May	1	si	no	f	14	16	$30,000
7	25	4-May	4	no	no	m		18	$60,000
8	27	5-May	3	no	no	f	3	6	$10,500
9	25	8-May	5	no	si	f	1	9	$15,000
10	30	12-May	3	no	si	m	0.1	5	$12,600
11	23	12-May	2	si	no	m	2	2	$21,500
12	23	14-May	2	no	no	f	2	11	$10,000
13	27	15-May	2	no	no	f	9	9	$23,000
14	19	17-May	2	no	no	f	12	7	$1,500
15	22	17-May	5	no	no	m	0	7	$26,000
16	25	17-May	9	si	si	f	6	6	$15,000
17	39	18-May	2	no	no	m	12	3	$3,500
18	26	18-May	2	si	no	f	2	18	$18,000
19	27	23-May	2	si	si	m	6	14	$32,000
20	22	24-May	2	no	no	f	0.5	6	$12,500
21	33	25-May	4	si	si	f	0.1	6	$12,000
22	20	25-May	1	no	si	f		14	$10,000
23	32	25-May	1	si	si	m		10	$23,000

24	30	3-Jun	3	si	si	m	6	11	$24,000
25	27	5-Jun	2	si	no	f	8	6	$20,000
26	24	5-Jun	4	no	si	f	6	4	$6,000
27	22	6-Jun	3	no	no	m	1	6	$5,000
28	22	7-Jun	4	no	no	m	9	8	$16,000
29	29	15-Jun	4	no	no	m	3	3	$15,000
30	39	15-Jun	2	si	no	m	17	11	$8,000
31	40	15-Jun	4	si	si	m	12	9	$16,000
32	31	15-Jun	3	no	no	m	1.5	14	$39,000
33	33	18-Jun	3	no	si	m	4	16	$46,000
34	26	19-Jun	1	no	si	m		13	$26,000
35	28	20-Jun	6	no	si	f	30	6	$11,000
36	22	22-Jun	1	no	no	f		15	$12,000
37	36	23-Jun	1	no	si	m	6	16	$45,000
38	26	25-Jun	1	no	no	m	4	7	$20,000
39	32	25-Jun	2	no	si	m	3	7	$21,500
40	28	25-Jun	1	no	no	m		15	$60,000
41	18	4-Jul	2	no	no	m	1.5	9	$26,000
42	17	5-Jul	1	si	no	f		9	$10,000
43	31	7-Jul	2	no	no	f	0	9	$10,000
44	23	7-Jul	2	si	no	m	6	9	$50,000
45	26	8-Jul	3	no	no	m	3	11	$12,000
46	26	8-Jul	3	si	no	f	18	9	$30,000
47	24	9-Jul	2	no	si	f	0.5	10	$12,000
48	24	12-Jul	1	si	si	f		15	$12,000
49	30	12-Jul	1	si	no	f	2	12	$30,000
50	17	13-Jul	1	no	si	f		9	$14,000
51	29	13-Jul	4	no	no	m	6	6	$10,000
52	27	14-Jul	2	si	no	m	3	11	$18,000
53	30	16-Jul	1	si	si	m		16	$4,000
54	25	2-Aug	2	si	no	m	6	9	$30,000
55	36	5-Aug	3	no	no	m	3	17	$38,000
56	32	8-Aug	1	si	si	f		14	$27,000
57	19	11-Aug	2	no	no	m		5	$6,000
58	28	12-Aug	1	si	si	m	5	16	$30,000
59	27	17-Aug	5	si	no	m	13	3	$18,000
60	24	27-Aug	14	si	no	m		16	$49,000

Crecimiento bebés

	Al parto					Al primer mes		
	Peso	Talla	Peso			Peso	talla	Peso
1	3.225	52.0			31	3.200	49.0	4.320
2	3.050	49.0	3.750		32	3.015	50.0	4.235
3	3.175	50.0	4.112		33	2.525	49.0	3.690
4	3.025	50.0	4.445		34	2.775	48.0	3.980
5	3.090	48.0	4.350		35	3.555	51.0	4.660
6	2.690	45.0	4.060		36	3.050	51.0	3.945
7	2.485	49.0			37	3.075	49.0	4.680
8	3.525	48.0	4.365		38	3.510	52.0	5.005
9	3.730	52.0			39	3.250	52.0	4.820
10	3.850	51.0			40	3.390	52.0	3.785
11	3.050	51.0			41	2.650	48.0	
12	3.625	51.0	4.710		42	2.875	48.0	3.535
13	2.830	49.0	3.875		43	3.600	52.0	4.175
14	3.225	50.0	4.530		44	3.640	52.0	5.095
15	3.565	51.0	4.925		45	3.025	50.0	4.015
16	2.750	48.0	3.195		46	3.575	50.0	4.420
17	2.800	51.0	4.215		47	3.100	51.0	3.825
18	2.640	49.0	3.225		48	2.950	50.0	3.210
19	3.000	49.0	3.575		49	3.505	51.0	4.290
20	3.110	50.0	3.880		50	3.500	52.0	4.356
21	3.375	51.0	4.495		51	3.665	56.0	3.750
22	2.725	46.0	3.350		52	3.315	52.0	4.880
23	2.875	48.0	3.850		53	3.220	50.0	3.750
24	2.925	49.0	5.145		54	3.400	52.0	4.150
25	3.600	53.0	4.070		55	3.650	54.0	4.295
26	3.000	52.0			56	3.225	51.0	4.100
27	3.500	52.0	4.880		57	3.185	52.0	4.130
28	3.120	50.0	3.285		58	3.210	50.0	4.465
29	3.000	51.0			59	3.450	52.0	
30	3.325	50.0	4.305		60	3.400	53.0	

(No se tienen registrados los datos faltantes)

Tabla de pesos y estaturas

(vigente en la Secretaría de Salud)

	Peso		Estatura	
Edad	Niños	Niñas	Niños	Niñas
Al nacer	3.4	3.2	51	49
1er mes	4.2	3.9	55	53
2do mes	5.1	4.7	59	56
3er mes	6.0	5.6	61	59
4to mes	6.6	6.2	64	62
5to mes	7.2	6.8	66	64
6to mes	7.7	7.3	68	66

a) Defina la población de interés.

b) ¿Cuál es el tipo y la escala de medición de cada una de las variables?

c) Obtenga el coeficiente de correlación para las variables años de escolaridad e ingreso familiar. Interprete su resultado.

d) Lo mismo que en el inciso anterior, pero para las variables edad y número de partos.

e) Explore el tipo de distribución subyacente en las variables años de escolaridad e ingreso familiar.

f) Obtenga el intervalo de confianza en 90% para el parámetro poblacional de la proporción de varones nacidos en ese hospital en el período de abril a agosto de 1998.

g) Obtenga el intervalo de confianza del 90% para el parámetro poblacional del número medio de partos para todas las mujeres que tuvieron bebé en este hospital en el período de abril a agosto de 1998.

h) Obtenga los intervalos de confianza el 95% para el peso medio y la talla media de los bebés al nacer.

i) La Secretaria de Salud afirma que hay un número medio de partos de 2.3 por mujer en la Ciudad de México. Pruebe si la media de partos por mujer, en esta población, sobrepasa este valor.

j) ¿Existe diferencia significativa, con respecto al sexo, entre los pesos medios y las tallas medias de los bebés al nacer?

k) Pruebe, utilizando el nivel de significancia descriptivo, que la estatura media al nacer de las niñas fue la que registra la Secretaría de Salud.

l) ¿Cree que exista una diferencia significativa en la dispersión de las estaturas de las niñas y de los niños al nacer?

m) En las últimas décadas ha habido un gran esfuerzo por tratar de informar a las madres de los beneficios de los partos psicoprofilácticos. Durante 1990 el 40% de los partos fue psicoprofiláctico en la Ciudad de México. Compare la información de esta muestra con dicho porcentaje.

ITAM

RESPUESTAS

RESPUESTAS

INTRODUCCIÓN

1.4. a) inferencial b) descriptivo c) descriptivo d) inferencial

1.5. a) descriptivo b) inferencial c) descriptivo d) descriptivo

1.9 a) inferencial población: personas en México (no está bien definida)
variable: partido con el que simpatizan
b) inferencial población: todos los posibles votantes
variable: participación en las votaciones
c) descriptivo población: alumnos que cursan Estadística II
variable: probabilidad con la que creen pasar
d) inferencial población: alumnos que toman un curso relacionado con la estadística
variable: probabilidad con la que creen pasar

1.11. i) electores que participarán en las próximas elecciones
ii) partido por el que votarán
iii) proporción de votos para cada partido
iv) el padrón electoral
v) estimación puntual del parámetro y una medida de confianza y precisión en la estimación

1.12. i) todos los egresados de esa universidad
ii) tiempo transcurrido entre la terminación de estudios y aquel en que encontró trabajo remunerado en su área
iii) El tiempo promedio que se tardan en encontrar empleo o la proporción de egresados que tardan más de un tiempo t en conseguir trabajo
v) lista de egresados de la institución con teléfono o domicilio
v) estimación puntual del parámetro y una medida de confianza y precisión en la estimación

1.14. a) Es una muestra aleatoria siempre y cuando se escoja aleatoriamente el inicio del proceso de selección

b) Sí c) 1/20 d) proporción de autos sin defectos

1.15. i) d ii) b iii) a iv) b v) d vi) a

ITAM

DISTRIBUCIONES MUESTRALES
Y EL TEOREMA CENTRAL DEL LÍMITE

2.1.1. a) falso b) verdadero c) verdadero d) verdadero

2.1.2 a) las amas de casa entrevistadas

b) Todas las amas de casa del D.F.

c) $\hat{p}$=30/50 y el parámetro es la proporción real de amas de casa del D.F. que prefieren esa marca de detergente

d) i) estadística descriptiva ii) estadística inferencial

2.1.3. $\overline{X}$, $\hat{P}$, $\sum X_i$ cuando las X_i v.a. independientes

2.1.4. disminuye la varianza de la distribución muestral

2.1.5. Error estándar

2.1.6. Sí

CONSTRUCCION DE DISTRIBUCIONES MUESTRALES

2.2.1. a)

x	P(X=x)
0	2/3
1	1/3

E(X) = 1/3 V(x) = 2/9

b)

$\hat{p}$	$f(\hat{p})$
0	4/9
1/2	4/9
1	1/9

c) $E(\hat{p}) = 1/3$ $V(\hat{p}) = 1/9$

2.2.2. a) $C_5^6=6$

2 4 6 8 10
2 4 6 8 12
2 4 6 10 12
2 4 8 10 12
2 6 8 10 12
4 6 8 10 12

b) medianas 6 6 6 8 8 8

c)

$\tilde{x}$	$P(\tilde{X}=\tilde{x})$
6	1/2
8	1/2

d) $E(\tilde{X}) = 7$ $V(\tilde{X}) = 1$

2.2.3. a)

$\bar{x}$	10	11	12	13	14
$P(\bar{X} = \bar{x})$	4/144	24/144	52/144	48/144	16/144

b) $E(\bar{X}) = 12.33$ $V(\bar{X}) = 0.94$

2.2.4. a) $\mu = 4$

$\sigma^2 = 31.66$

med = 1.5

b)

0 0 1	0 1 2
0 0 2	0 1 5
0 0 5	0 1 16
0 0 16	0 2 5
0 1 2	0 2 16
0 1 5	0 5 16
0 1 16	1 2 5
0 2 5	1 2 16
0 2 16	1 5 16
0 5 16	2 5 16

c)

$\overline{X}$	$P(\overline{X} = \overline{x})$
1/3	1/20
2/3	1/20
1	2/20
5/3	1/20
2	2/20
7/3	2/20
8/3	1/20
16/3	1/20
17/3	2/20
6	2/20
19/3	1/20
7	2/20
22/3	1/20
23/3	1/20

$E(\overline{X}) = 4$ Resulta que $\mu = 4 = E(\overline{X})$

$V(\overline{X}) = 6.328$

d) $P\{|\overline{X} - 4| > 1\} = 1$

e)

m	P(M = m)
0	4/20
1	6/20
2	6/20
5	4/20

E(m)= 1.9

V(m)= 2.89

E(m) no coincide con la mediana poblacional

f) La distribución muestral de la mediana presenta menos variabilidad porque se ve menos afectada por valores extremos que la media. En este caso se tiene un valor extremo de 16.

2.2.5. b)

$\bar{x}$	0	0.5	1.0	1.5	2.0	2.5	3.0
$P(\bar{X}=\bar{x})$	0.01	0.12	0.40	0.26	0.16	0.04	0.01

c) $E(\bar{X}) = 1.3$ $V(\bar{X}) = 0.305$

2.2.6. a)

$\bar{x}$	$P(\bar{X}=\bar{x})$
3	1/27
4	3/18
5	11/36
6	7/24
7	11/72
8	3/72
9	1/216
	1

b)

$\tilde{x}$	$P(\tilde{X}=\tilde{x})$
3	7/27
6	18/27
9	2/27
	1

c) $\sigma_{\bar{x}} = 1.19$ $\sqrt{V(\tilde{X})} = 1.64$

2.2.8. a)

$\bar{x}$	$P(\bar{X}=\bar{x})$
1	.222
1.5	.333
2	.289
2.5	.133
3	.022

b) $E(\overline{X}) = 1.7$ $V(\overline{X}) = 0.2665$

2.2.9. a) $\mu_X = 6$ $\sigma_X^2 = 11.66$

b)

$\overline{x}$	$P(\overline{X} = \overline{x})$
2	1/15
3	1/15
4	2/15
5	2/15
6	3/15
7	2/15
8	2/15
9	1/15
10	1/15

c) $P(\overline{X} \geq 10) = 1/15$

d)

$\overline{x}$	$P(\overline{X} = \overline{x})$
1	1/36
2	2/36
3	3/36
4	4/36
5	5/36
6	6/36
7	5/36
8	4/36
9	3/36
10	2/36
11	1/36

e) $P(\overline{X} \geq 10) = 3/36$

2.2.10. a)

$\hat{p}$	$P(\hat{p})$
0	6/15
1/2	8/15
1	1/15

b) $E(\hat{p}) = 1/3$ igual a p

c) $\sigma_p = .298$

2.2.11. a)

y	P(y)
0	2/3
1	1/3

b) $E(y) = 1/3$ $V(y) = 2/9$

c)

$\hat{p}$	$P(\hat{p})$
0	4/9
1/2	4/9
1	1/9

d) $E(\hat{p}) = 1/3$ $V(\hat{p}) = 1/9$

2.2.12. a)

S^2	$P(S^2 = s^2)$
0	3/9
4.5	4/9
18	2/9

b) $E(S^2) = 6$

$V(S^2) = 6$

c) $P(4 < S^2 < 10) = 4/9$

2.2.13. b)

w	0	1/2	2
$P(W = w)$	0.42	0.48	0.10

c) $E(W) = 0.44$

$V(W) = .3264$

d) $E(Y) = 9.8$

$V(Y) = 1.76$

2.2.14. a) $E(Z) = 3.95$ $V(Z) = .6475$

b)

s^2	0	0.5	2	4.5
$P(S^2 = s^2)$	.365	.465	.14	.03

c) $E(s^2) = .6475$ $V(s^2) = 0.8644$

DISTRIBUCION DE MUESTREO DE: $\overline{X}$, $\hat{p}$, S^2, $\overline{X}+\overline{Y}$, $\overline{X}-\overline{Y}$, $\hat{p}_1-\hat{p}_2$, S_1^2/S_2^2

2.3.2. e) .9742

2.3.3. a) $\left(\overline{X}+\overline{Y}\right) \sim N\left(35, \frac{15}{50}+\frac{20}{40}\right)$ b) .0125

2.3.4. a) $P(X \leq 4.5) = .1587$ no
b) $P(\overline{X} \leq 4.5) = .0008$ sí
c) $P(\overline{X} \leq 4.9) = .2635$ no

2.3.5. 0.0548

2.3.6. a) 0.0548 b) 0.0228 c) 0.9986 d) 0.7698

2.3.7. 0.0232

2.3.8. a) $\hat{p} \sim N\left(p, p(1-p)/n\right)$ b) .7150

2.3.9. $\mu = 4.554$

2.3.10. .1444

2.3.11. 0.04779

2.3.12. .00051

2.3.13. a) cero

b) .0766 que la distribución es normal

c) .9234

2.3.14. a) 0.9545 b) 151

2.3.15. 0.8648

2.3.16. 0.2307

2.3.17. a) .0264 b) $P(\overline{X} \geq 53) = .0018$ Sí

2.3. 18. a) $E(\overline{X}) = .1$ $V(\overline{X}) = .001$

b) .0569

c) .0671

2.3.19. a) uno b) .0142 c) .0002

2.3.20. 0.0014 Sí

2.3.21. 0.1379

2.3.22. Normalidad 0.0228

2.3.24. a) menor a .005 (=0.00043) b) Que es demasiado grande

2.3.25. $P(F_{15,20}>0.52) = .90$

2.3.26. 0.8792 no

2.3.27. 0.6572

2.3.28. 0.1030

2.3.29. n = 1813

2.3.30. 0.4855

2.3.31. 0.07446

2.3.32. a) 20 b) .2 c) 0.3173 = 31.73%

2.3.33. 0.1017 normalidad

2.3.34. 1

2.3.36. a) $F_{m,k}$ b) χ^2_n

2.3.38. 0.9544

2.3.39. 0.1161

2.3.40. a) 0.99996 normalidad b) 0.3121

2.3.41. 0.1547

2.3.42. 0.0049

2.3.44. a) .9836 b) 24

2.3.45. a) .7977 b) .10

2.3.46. menor a 0.01 (=0.0075)

2.3.47. ≈ .05 (=.0507)

2.3.48. 0.949

2.3.49. 0.0062

2.3.50. b) .0028 c) n=15

USO DE TABLAS

RESPUESTAS

I		II	
2.4.1	.025	2.4.20	9.26
2.4.2	.995	2.4.21	27.6882
2.4.3	.95	2.4.22	-2.086
2.4.4	.99	2.4.23	b=2.57
2.4.5	.025	2.4.24	b=1.94
2.4.6	.975	2.4.25	a=.216
2.4.7	.025	2.4.26	a=.515
2.4.8	.975	2.4.27	3.22
2.4.9	.847	2.4.28	b=27.4884
2.4.10	.05	2.4.29	-1.323
2.4.11	.005	2.4.30	a=.025
2.4.12	.94	2.4.31	a=.714
2.4.13	.97	2.4.32	a=29.005 b=51.805
2.4.14	.04	2.4.33	a=43.2752
			b=104.215
2.4.15	.965	2.4.34	a=-1.383 b=1.383
2.4.16	.095	2.4.35	a=-2.718 b=2.718
2.4.17	.945	2.4.36	a=.5988
2.4.18.	.94	2.4.37	a=.227 b=3.37
2.4.19.	.945	2.4.38	a=.29 b=4.41

PROPIEDADES DE LOS ESTIMADORES

3.1.2. a) Los tres son insesgados
b) $V(T_1) = \sigma^2/n_1$ $V(T_2) = 2\sigma^2/n_1$ $V(T_3) = 3\sigma^2/4n_1$
c) T_3

3.1.3. $\hat{p}_1$ sesgado $\hat{p}_2$ insesgado

3.1.4. Si

3.1.5. a) .90 b) El 10% de las observaciones no se usaron para calcular $\hat{\mu}_2$.

3.1.6. a) E(X)= (b+a)/2 $V(X) = (b-a)^2/12$
b) E($\hat{\mu}_1$)= E($\hat{\mu}_2$) = (b+a)/2 E($\hat{\mu}_3$) = (b+a)/3
c) $\hat{\mu}_1$ y $\hat{\mu}_2$ son insesgados
d) $V(\hat{\mu}_1) = \frac{(b-a)^2}{36}$ $V(\hat{\mu}_2) = \frac{(b-a)^2}{24}$ V($\hat{\mu}_3$)= V(X)/6

3.1.7. a) $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$ son insesgados
b) ER($\hat{\mu}_2/\hat{\mu}_3$) = $n^2/8(n-2)$ ER($\hat{\mu}_1/\hat{\mu}_3$) =n/2
c) $\hat{\mu}_3$

3.1.8. a) $\hat{\mu}_2$ y $\hat{\mu}_3$ son insesgados
b) $\hat{\mu}_3$
c) $\hat{\mu}_3$

3.1.9. a) $\hat{\theta}_1$ es sesgado y $\hat{\theta}^2$ insesgados
b) ECM ($\hat{\theta}_1$) = $2.2\theta^2$ ECM($\hat{\theta}_2$) = $.25\theta^2$
c) $\hat{\theta}_2$

3.1.10. k=2/3

3.1.11. a) $\hat{\mu}$ es estimador insesgado de μ

b) $W=n_1/(n_1+n_2)$

3.1.12. .88

3.1.13. S_1^2 sesgado S_2^2 insesgado

3.1.15. $E(\hat{\lambda})=n\lambda^2$, por lo tanto no es insesgado

3.1.16 a) $E\left(\hat{\theta}_1\right)=E\left(\hat{\theta}_2\right)=\frac{2}{3}\theta$

$V\left(\hat{\theta}_1\right)=\frac{\theta^2}{54}$, $V\left(\hat{\theta}_2\right)=\frac{\theta^2}{36}$

$\therefore \hat{\theta}_1$ es el mejor estimador

b) 3.86

3.1.17. a) $\hat{\lambda}_1$ insesgado $\hat{\lambda}_2$ sesgado sesgo=$-\lambda$

b) $ECM(\hat{\lambda}_1) = \lambda/n$ $ECM(\hat{\lambda}_2) = \lambda/2 + \lambda^2$

c) $\hat{\lambda}_1$

3.1.18. b=2/3

3.1.19. a) Sí

b) sesgo= $-1/n\ \{\theta(1-\theta)/n\}$

3.1.20. a) Sí

b) Sí

c) n=43

3.1.28. El sesgo de $\hat{\lambda}_1$ es cero.

$\hat{\lambda}_2$ tiene sesgo $\left[\lambda(1-n)\right]/n$

ITAM

ESTIMACION PUNTUAL

3.2 1. $\overline{X} = 503.75$

3.2.2. $\overline{X} = 3$ $S^2 = 0.000545$

3.2.3. $\overline{X} = 443.33$ proceso estándar $\overline{X} = 451.42$ proceso nuevo

3.2.4. $S^2 = .001$

3.2.5. a) $\overline{X} = 158.3125$ $S^2 = 268.23$

3.2.6. a) $\overline{X} = 587.22$ b) $S^2 = 1{,}016{,}942.69$ c) $\hat{p} = \frac{5}{9} = .556$

3.2.7. $S^2 = .149318$

3.2.8. $\overline{X} = 90$ $S^2 = 157{,}3469$

3.2.9. $\overline{X} = 2994.75$ $S^2 = 1{,}226{,}572$

3.2.10. 0.866372

INTERVALOS DE CONFIANZA.

4.1.4. (.4525, .6475) (.4216, .6783)

4.1.5. (.0002, .0023)

4.1.6. a) (.2864, .4687)
b) (.2743, .4935)
c) Sí

4.1.7. (0.00674, 0.0253) (0.0015, 0.0305)

4.1.8. (.9781, 1.0331) normalidad

4.1.9. Supuesto: las lecturas del instrumento se distribuyen como normal;
$\sigma^2 \in (1.4075, 31.2637)$ y $\sigma \in (1.1864, 5.5914)$. No es adecuada la garantía.

4.1.10. a) (147.65, 168.96) normalidad
b) (146.36, 642.52) normalidad

4.1.11. a) (14.204, 14.596) normalidad
b) .2

4.1.13. a) (19.44, 55.87)
b) La distribución de los puntos del bolichista es normal
c) No

4.1.14. n = 2401

4.1.15. a) (514.21, 535.79)

b) Como los valores del intervalo son menores a 600, se puede pensar que sí hubo un daño por las heladas.

4.1.16. (.0716, .1284) Sí

4.1.17. a) (-.344, 8.344)

4.1.18. a) (.5216, .6784)

b) (1564.8, 2035.2)

4.1.19. (31.20, 90.32)

4.1.20. (19.53, 22.47)

4.1.21. (14.13, 14.85) Sí

4.1.22. a) i) $\overline{X}$ = 12.06 CV = 25.31% ii) $\overline{X}$ =15.44 CV = 15.28%

b) Para la media G(10.15, 13.97) Mex(13.96, 16.92)

Para la varianza G(5.17, 21.57) Mex(3.09, 12.9)

c) Las utilidades de las industrias se distribuyen como normal.

4.1.23. a) (.8440, .8725)

b) n = 385

4.1.24. a) (6772.79, 10547.21)

b) normalidad

c) No

4.1.25. (.5040, .696)

4.1.26. Intervalo de confianza para la varianza (11.6038, 135.5491)

Intervalo de confianza para la desviación estándar (3.4064, 11.6426) No

Los rendimientos de los instrumentos de inversión tienen una distribución normal.

DETERMINACION DEL TAMAÑO DE MUESTRA

4.2.1. n = 61

4.2.2. n = 5 suponer que el tiempo que tarda pintar el departamento se distribuye como normal

4.2.3. n = 218

4.2.4. n = 62

4.2.5. a) .8664 b) 44

4.2.6. n = 47

4.2.7. n = 62

4.2.8. n = 209

PRUEBAS DE HIPÓTESIS PARAMÉTRICAS

5.2.1. a) = .03214

b) (.6)=.416 (.4)=.02103

5.2.2. a) F_{obs}= 0.44 con = .01 , no se rechaza H_o.

b) Las utilidades de las compañías se distribuyen como normal.

5.2.3. a) H_a: >48000 Si se rechaza H_o el fabricante usa la nueva

b) H_a: <48000 Si se rechaza H_o el fabricante usa la anterior.

5.2.4. Supuestos: La edad de las mujeres de ambos medios sigue una distribución Normal.

Se rechaza H_o . La varianza es mayor en el medio urbano. Valor p < .005

5.2.5. a) (11) = .0012

b) n = 13

5.2.6. a) Z_{obs} =1.33, con α=0.05 no se rechaza H_0

b) valor-p = $P(Z \geq 1.33)$ = 0.0918

5.2.7. a) t_c =2.11 > $t_{.05,7}$ =1.89 Se rechaza H_o

b) P(t_c >2.11) esta entre .025 y .05 (=0.031953)

c) Las diferencias d_i se distribuyen como Normal.

5.2.8. a) t_c =6.455 Se rechaza H_o

c) $\approx$ 0.01

b) χ^2_c = 45.36 > $\chi^2_{14,.05}$ =23.69 Se rechaza H_o

5.2.9. $t = -2.52$ Se rechaza H_o

5.2.10. Error tipo I excesivo gasto
Error tipo II disminución en la ventas

5.2.11. a).0850 b).3410 c).0044 d).0039

5.2.12. $t=-.907$ No se rechaza H_o

5.2.13. $t_c = 4.0599 > t_{15,.01} = 2.602$ Se rechaza H_o . El programa parece ser efectivo.

5.2.14. a) $H_o : p=.15$ $H_a : p>.15$
b) =.3523
c) (.20)=.4114 1- (.20)=.5886
(.25)=.2252 1- (.25)=.7748

5.2.15. a) $Z_c = 2.09$ no
b) $P(z>=2.09) = 0.0183$

5.2.16. $t_c = 8.5714$ Se rechaza H_o a favor de que, el coeficiente intelectual aumenta con la edad.

5.2.17. a) $t_{22,.05} = -1.717$ y $t_c = -1.5816$ No se rechaza H_o
b) $\pi(1000) = P(T > -1.25)$, $0.75 < \pi(1000) < 0.90$

5.2.18. a) Fobs = 2.49 Se rechaza la hipótesis de igualdad de varianzas

b) P(Fobs > 2.49) < 0.005. La probabilidad de incorrectamente rechazar H_o cuando ésta es cierta es menor a .005 (=0.002435)

5.2.19. a) La varianza de la línea A no es significativamente diferente de 10

La varianza de la línea B es significativamente diferente de 10

b) Fobs =.36 < $F_{.975,19,15}$ = 0.3817 $\therefore$ se rechaza que las varianzas sean iguales a favor de que la línea B presenta más variabilidad en el número de defectos.

5.2.20. a) NO. Valor p > 0.1

b) Supuesto: las diferencias en el número de defectuosos se distribuye como normal

c) $1-\alpha$, β

5.2.21. a) No se rechaza H_o: $\sigma_A^2 = \sigma_B^2$

b) No se rechaza H_o: $\mu_A = \mu_B$

5.2.22. a) H_o p = .5 H_a: p > .5

b) = P(X > 7|n = 10 p = .5) = .1719

c) (.6) = .6177 (.7) = .3504 (.8) = .1209

(.4) = .9452 (.3) = .9894 (.2) = .9991

5.2.23. a) H_o: ≥ 2 H_a: < 2 error tipo I es rechazar que ≥ 2 cuando en realidad ≥ 2.

b) χ^2_{obs} = 25.35 $\therefore$ se rechaza H_o

c) P(χ^2_{60}< 25.35) < .005

5.2.24. Z_c=4.08. Se rechaza H_o

5.2.25. n = 68

5.2.26. a) No se rechaza Ho es decir el porcentaje no ha aumentado

b) 1- (.2) = .2054 1- (.3) = .5075

ESTADÍSTICA NO PARAMÉTRICA

6.1. Con α = .05, se rechaza H_O ya que $P(X \geq 9 | n = 11,\ p = .5) = .0327$

6.2. a) No, T_X = 9 con α = .05 b) No, T_X = 15 con α = .05

6.3. Sí hay preferencia entre una proposición y la otra (Se rechaza Ho)

6.4. No se rechaza Ho (Valor p>.1)

6.5. a) T_X = 61 $T_{.05}(12,12) = 101$ No se rechaza Ho

b) $\left.\begin{array}{l} T_X = 50 \quad T_{.025}(12,12) = 106 \\ T_{.975}(12,12) = 144 - 106 = 38 \end{array}\right\}$ No se rechaza H_0. Las dispersiones son iguales

6.6. Si existe relación entre el tipo de sangre y la severidad del resfriado. $\chi^2 = 28.20$

6.7. $\chi_c^2 = 0.6498$. No se rechaza H_0.

6.8. a) No, T_X = 39 No se rechaza H_0.

b) No, T_X = 80.5 No se rechaza Ho con α = .05

6.9. a) No se rechaza H_0 ya que el valor p es de .2744

b) t_{obs} = 2.12 con α = .05 Se rechaza Ho

c) Si, r_S = .87

6.10. No T_X = 32 No se rechaza Ho con α = .05

6.11. a) Si, $r_S = .83$ b) Valor p = .01

6.12. a) Valor p = 2(0.0384) = 0.0768.
b) $r_S = 0.7121$

6.13. Si, $T_X = 106$, con $\alpha = .05$ Se rechaza Ho

6.14. m = 6 $2P(M \geq 6 | n = 9 \;\; p = .5) = 2(0.2539) = 0.5068$. No se rechaza Ho

6.15. b) $H_0 : M_{pz} = M_s$ $T_x = 23$. Se rechaza H_0 c) $\chi_c^2 = 6.83$ $.005 < p < .01$.

6.17. a) No $T_X = 9.5$. Si se rechaza Ho b) Si t =3.54 Se rechaza Ho

6.18. a) No se rechaza la hipótesis nula de independencia.

6.19. c) Valor *p* = .0733.

6.20. c) Valor p = .0227

6.22. c) Valor $p = P(X \geq 9) = .1334$ d) No se rechaza H_0 con $\alpha = .06$

7.2.2. a) $\bar{x}$ = 5.0125, .8711
b) n = 21

7.2.3. a) $\bar{x}$ = 2, .938
b) $\hat{\tau}$ = 2000, 938.08
c) $\hat{p}$ = .4 .3266
e) n = 19
d) n = 16

7.2.4. Las secretarias tienen razón. El Intervalo de Confianza calculado con la muestra está muy por debajo del valor μ = 20100

7.2.5. Sí. Con n = 1091 el límite máximo para el error de estimación es 2.9%

7.2.6. a) $\overline{X}$ = 12.5, 2.23
b) $\hat{\tau}$ = 125000, 22266.39

7.2.7. a) (82.787, 85.413)
b) (83.17, 85.02)

7.2.8. [1.9337, 2.2663]

7.2.9. Con ese tamaño se muestra el error de estimación es menor al 6%

7.2.10. a) 161.8
b) $\hat{V}(\overline{X})$ = 138.63 2 $\hat{\sigma}_{\overline{X}}$ = 23.55
c) 16180 $\hat{V}(\hat{\tau})$ = 1386300

d) $2\ \hat{\sigma}_{\hat{\tau}} = 2354.82$

7.2.11. a) 381.20, 168.50
b) 5320, 3602

7.2.12. Si el muestreo fue realmente aleatorio simple y no se presentó el problema de la respuesta inexacta el error de estimación máximo asociado a este tamaño de muestra n = 2207 es de .0213

7.2.13. Sólo el 27% contestaron. No