Universidad de Costa Rica
Facultad de Ciencias Económicas
Escuela de Estadística

EJERCICIOS DE LA CATEDRA DE ESTADISTICA GENERAL 2

(XS 277)

[1 ciclo lectivo 2024]

PRESENTACION

El presente “Manual de Prácticas” contiene una serie de ejercicios que tienen como finalidad apoyar el aprendizaje práctico sobre los temas que se imparten en el curso XS-0277 Estadística General 2, que se imparte para las carreras de Administración en la Facultad de Ciencias Económicas de la Sede Rodrigo Facio y en las diferentes Sedes Regionales de la Universidad de Costa Rica.

El documento es una actualización de los ejercicios aportados por los profesores de la cátedra, y pretende ser una base para las lecciones prácticas y laboratorios de cómputo que se realizan durante el ciclo lectivo; así como ilustrar la aplicación de los conceptos estadísticos aprendidos durante el curso.

Adicionalmente se incluye un preámbulo con ejercicios que repasan el uso de la distribución normal para la resolución de problemas, pues dicha distribución es la base para el desarrollo de los demás temas del curso.

Se agradece de antemano cualquier observación o sugerencia, que tenga como propósito mejorar su contenido.

Su servidor,

Patricia Delvó Gutiérrez,
Coordinadora
Cátedra Estadística General

INDICE

TEMAS	Número de página
PREAMBULO	*3*
Estimación	*5*
Verificación de Hipótesis	*10*
Análisis de Variancia	*14*
Correlación y Regresión	*20*
Estadística No Paramétrica	*26*

PREAMBULO

1. Sea Z una variable aleatoria normal con $\mu = 0$ y $\sigma = 1$. Calcular:

a.	$P(Z \leq 0)$	d.	$P(Z \geq 1)$
b.	$P(Z > 1)$	e.	$P(Z > -1)$
c.	$P(-1 < Z < 1)$	f.	$P(-2 < Z < -1)$

2. Se sabe que la demanda aleatoria de gasolina durante un cierto periodo de tiempo se comporta con arreglo a la ley normal de media 150 000 litros y desviación estándar 10 000 litros, determinar la cantidad que hay que tener dispuesta a la venta en dicho periodo para poder satisfacer la demanda con una probabilidad de 0,95.

3. En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.

4. Si el número de personas que entran diariamente en unos grandes almacenes se distribuye normalmente. Determine la media y la varianza de la variable X, si hay una probabilidad 0,58 de que entren menos de 75 clientes y una probabilidad 0,38 de que entren entre 75 y 80 clientes.

5. Las calificaciones de los 500 aspirantes presentados a un examen para contratación laboral, se distribuye normalmente con media 6.5 y variancia 4.
   a. Calcule la probabilidad de que un aspirante obtenga más de 8 puntos.
   b. Determine la proporción de aspirantes con calificaciones inferiores a 5 puntos.
   c. ¿Cuántos aspirantes obtuvieron calificaciones comprendidas entre 5 y 7,5 puntos?

6. Sólo 24 de los 200 alumnos de un Centro de enseñanza media miden menos de 150 cm. Si la estatura media de dichos alumnos es de 164 cm., ¿cuál es su varianza si la variable se distribuye de forma normal?

7. Las puntuaciones de un examen se distribuyen normalmente con media 15 puntos. La puntuación A ha sido superada por un 23% de los alumnos. La puntuación B está situada a 5 puntos diferenciales por debajo de la media. Entre B y la media se encuentra el 30% de los alumnos. Calcular:
   a. La desviación típica de las notas.
   b. Las puntuaciones directas de A y B.
   c. El porcentaje de alumnos entre A y B.

8. Una máquina despachadora de refrescos está ajustada para servir un promedio de 200 ml por vaso. Si la cantidad de refresco es normalmente distribuida con una desviación estándar igual a 15 ml.

a. ¿Qué fracción de los vasos contendrá más de 224 ml?
b. ¿Cuál es la probabilidad de que un vaso contenga entre 191 y 209 ml?
c. ¿Cuántos vasos probablemente se derramarán si se utilizan vasos de 230 ml en los siguientes 1000 refrescos?
d. ¿Debajo de qué valor se obtiene el 25% más pequeño de los refrescos?

8. En el primer semestre del 2021 el salario inicial medio de los recién egresados de una Escuela de Administración de Negocios era $ 31 280 anuales. Suponga que los salarios iniciales siguen una distribución normal con desviación estándar $3 300. ¿Qué porcentaje de los egresados tiene un salario inicial medio:
a. entre $ 30 000 y $35 000?
b. superior a $40 000?
c. entre $35 000 y $40 000?

9. Se supone que las calificaciones de un examen se distribuyen normalmente, con promedio 7,5 y una desviación estándar de 1,5. Sabiendo que se concede sobresaliente al 15% de los estudiantes que hicieron el examen, y que el 10% de los estudiantes que lo hicieron no consiguen aprobar, encuentre:
a. la puntuación mínima exigida para obtener sobresaliente
b. la puntuación mínima exigida para aprobar.

10. Suponga que el costo por hora de operar un avión comercial sigue una distribución normal con una media de $ 2 100 por hora y una desviación estándar de $ 250. ¿Cuál es el costo más bajo de operación para el 3 % de los aviones?

1. Estimación

1.1 Los siguientes datos corresponden a la cantidad de transacciones que se realiza la sucursal en un banco cada día.

a) Calcule el promedio y la desviación estándar de la siguiente población:

102	92	97	97	13	75	91	34	34
34	94	23	34	32	101	52	62	31

b) Seleccione una muestra aleatoria sin reemplazo de ocho días y calcule el promedio y la desviación estándar

c) Compare los resultados de a) y b) y comente sobre las diferencias

d) Realice una estimación del promedio muestral por intervalo de la muestra de b) utilizando la desviación estándar de la muestra y un nivel de confianza del 95%.

1.2 Un auditor tomó una muestra de 50 cuentas por cobrar, obteniendo un valor promedio de cuentas por cobrar de 260 000 colones, y una desviación estándar de la población de 45 000 colones. Estime el intervalo de confianza para el promedio utilizando una confianza del 90%. Interprete.

1.3 Se desea estimar el monto promedio de las ventas de galletas para un establecimiento de ventas al detalle, durante el último año; para lo cual se toma una muestra de 25 establecimientos y se estableció un monto promedio de 758 500 colones, con una desviación estándar de 95 000 colones. Calcule e interprete el intervalo de confianza al 99% para el monto promedio de ventas al detalle.

1.4 El Ministerio de Salud desea estimar un intervalo de confianza para la diferencia de promedios de las horas de sueño de los habitantes de Matina y Talamanca con un 95% de confianza. Para ello toma una muestra aleatoria de 200 habitantes de Matina resultado que un promedio de 8,3 horas de sueño con una desviación estándar de 4,7 horas; por su parte toma una muestra de 210 habitantes de Talamanca obteniendo un promedio de 7,2 horas con una desviación estándar de 3,8 horas. Determine el intervalo citado.

1.5 Una empresa desea ofrecer los servicios de remolque de vehículos en carretera, para lo cual estudia en todo el país mediante una muestra de 235 vehículos encontrados en la carretera con problemas mecánicos, de los cuales 59 utilizaron los servicios de remolque. Determine el intervalo para la proporción de vehículos que utilizan servicios de remolque con un 98% de confianza e interprete los resultados.

1.6 En una muestra de 37 países del mundo en el año 2018, se encontró un promedio de desperdicio de alimentos de 179 kilos por habitante, con una desviación estándar de 98 kilos. Utilizando la información anterior, calcule e interprete un intervalo del 97% de confianza para el promedio poblacional de desperdicios de alimentos por habitante en los países del mundo en términos del problema. Anote el resultado final con dos decimales, usando redondeo a la unidad más próxima.

1.7 La empresa "Pizza Me gusta", se ha especializado en ventas a domicilio y está considerando una promoción para los lunes de "lleve 2 y pague 1", para incrementar el tamaño de las órdenes de sus clientes. Con el fin de determinar las bondades de la promoción realizada, selecciona una muestra de 36 órdenes con base en las cuáles determina una media de $14.37 con una desviación estándar de $4.02. Si interesa realizar una estimación del 95% de confianza, construya el intervalo respectivo e interprete el resultado.

1.8 Una empresa dedicada al diseño clásico de interiores está interesada en contar con información sobre los salarios anuales de los empleados de esta industria para poder revisar su escala salarial. Para ello toma una muestra de 18 compañías y encontró un salario promedio de $31780, con una desviación estándar de $5450. Construya un intervalo de confianza del 96% para el salario medio anual de las empresas de dicha industria.

1.9 Una empresa automotriz está interesada en construir un intervalo de confianza para el promedio y desviación estándar de las horas requeridas para armar un automóvil. Si toma una muestra de tamaño 27, obteniendo una media de 36,2 horas y con una desviación estándar de 2,6. Si se sabe que la distribución de las medias muestrales del tiempo de armado es normal, construya un intervalo de confianza del 94% para el tiempo medio verdadero real desconocido de armado de automóviles en dicha empresa.

1.10 Una universidad estatal está interesada en hacer algunas mejoras laborales para su personal administrativo y académico. Entre los aspectos que desea determinar es la proporción de personas funcionarias que están a favor de teletrabajar al menos dos días por semana. Dado que no cuenta con estudios previos se decide utilizar el supuesto de varianza máxima. Determine el tamaño de la muestra para un nivel de confianza del 95%, un error permisible del 5%, sabiendo que la población de funcionarios es de 5500.

1.11 Un ejecutivo de un banco desea determinar la cantidad del promedio de depósitos mensuales totales por cliente en dicha entidad. Piensa que el estimado de esta cantidad promedio y el uso de un intervalo de confianza es suficiente. ¿Qué tan grande debe ser la muestra que debe tomar para estar a no más de 10 000 del promedio real con 99% de confianza? Suponga que la desviación estándar de los depósitos mensuales totales para todos los clientes es alrededor de 500 000 colones.

1.12 Suponga que durante varios años una empresa le ha dado siguiendo al precio de la acción de una aerolínea y actualmente se encuentra interesada en determinar el precio diario promedio de esta acción en un período de 10 años. No se desea promediar los precios diarios de la población de 2 500 precios, porque resulta muy laborioso conseguirlos. Por lo anterior se decide tornar una muestra aleatoria (sin reemplazo) de los precios diarios y estimar el promedio. Se desea tener 90% de confianza en los resultados y que el promedio estimado no se aleje más de $2,00 del promedio verdadero. Un experto estima que la desviación estándar del precio de esta acción es de unos $12,50 para este período. ¿Qué tan grande debe ser la muestra por tomar?

1.13 Asuma que el Ministerio de Educación desea estimar la proporción de hogares de la zona rural en Costa

Rica que tienen computador personal utilizando una muestra aleatoria. Si se desea un nivel de confianza del 93% y que la estimación citada no supere en un 5% al valor poblacional, determine el tamaño de muestra (sin remplazo) apropiado que responda los requerimientos especificados. Según los datos del INEC a julio del 2023 la población de hogares de la zona rural llegó a 488204.

1.14 Se tomó una muestra de 1500 propietarios de establecimientos comerciales con el fin de consultar acerca de los artículos de mayor consumo durante la pasada pandemia. Si se deseaba un nivel de confianza del 90% y se asumió variabilidad máxima, ¿cuál fue el error máximo permisible que se supuso para determinar el tamaño de la muestra?

1.15 Explique en sus propias palabras en qué consiste el teorema del límite central.

1.16 ¿El promedio es un estimador suficiente? Justifique su respuesta.

1.17 En una empresa con un número elevado de empleados se desea estimar el promedio de horas trabajadas (μ) por medio de una muestra. Se sabe que la desviación estándar es de 4,3 horas. Si se quiere un error estándar del promedio de 0,5 horas, ¿Qué tamaño de muestra se debe utilizar?

1.18 En una ciudad de Pennsylvania hay 250 familias. Se toma una muestra aleatoria sin reemplazo de 25 familias que revela que la contribución media a la Iglesia ronda los $450, con una desviación estándar de $75. Con base en esta información responda lo siguiente:

a) ¿Cuál es la media de la población?
b) ¿Cuál es el mejor estimador de la media de la población?
c) Construya un intervalo de confianza del 90% para la media de la población. Indique cuáles son los puntos extremos del intervalo de confianza.
d) Interprete el intervalo de confianza.
e) Indique si la media pude ser $445 o $424. Justifique su respuesta.

1.19 El ingeniero de producción de cierto proceso de fabricación está considerando realizar un cambio en el proceso de control de la fabricación de partes componentes. Para determinar si el nuevo proceso resulta en una mejoría, decide toma muestras del procedimiento existente y del nuevo proceso. Toma una muestra de 1500 artículos del procedimiento actual y encuentra 75 que son defectuosos. En la muestra de 2000 artículos del nuevo proceso encuentra 80 artículos defectuosos. Construya un intervalo de confianza del 90% para la diferencia real de la fracción de artículos defectuosos con el proceso actual de control y con el nuevo.

1.20 El Consejo Nacional de Política Pública de la Persona Joven y el Viceministerio de la Juventud realizaron en abril de 2007 una encuesta aplicada 2500 jóvenes a una de muestra de los residentes en Costa Rica (costarricenses y extranjeros) con edades entre 15-35 años que fueron seleccionados bajo el método de cuota, es decir cada entrevistador debía completar una cuota de entrevistas de jóvenes con determinadas características utilizando su mejor juicio.

a) En el informe de la encuesta citada, se indica que se utilizó un nivel de confianza del 95% una población de 1 389 198 jóvenes, de manera que se obtuvo un error de muestreo del 2% para la estimación de la proporción de interés. ¿El valor del error de muestreo indicado es correcto? Justifique

b) A suvez, la muestra citada se distribuyó como se presenta a continuación:

ZONA	POBLACIÓN	MUESTRA
GAM	748 586	1 347
Resto del país Urbano	214 084	385
Resto del país Rural	426 528	768

b1) Con base en la tabla anterior, compruebe que se utilizó la distribución proporcional.

b2) Indique cómo cambaría la distribución anterior, si se distribuye utilizando el criterio de afijación óptima, en el supuesto caso de que el estrato Resto del país Urbano fuese más variable

b3) Explique las razones de las diferencias entre las afijaciones proporcional y óptima

b4) Dentro de la metodología se indica que la muestra se asignó en primera instancia según la zona de la tabla anterior, luego dentro de cada zona se seleccionaron distritos, después viviendas y finalmente jóvenes. Sobre lo expuesto, indique que tipo de muestreo se utilizó. Justifique

c) Asuma que con base en una muestra del resto del país urbano se estimó que los jóvenes mismos devengan un ingreso promedio mensual de ¢276 188. Con base en los datos anteriores, calcule e interprete un intervalo de confianza para el promedio poblacional del ingreso citado, utilizando un nivel de confianza del 90% e ignorando el factor de corrección y asumiendo una desviación del ingreso mensual de ¢112 000.

1.21 Una empresa dedicada a la venta de carne de cerdo para la parrilla desea lanzar un nuevo producto y para ello investiga el mercado potencial entre jóvenes y adultos. Selecciona una muestra aleatoria de 80 jóvenes y 75 adultos en San José, donde 27 jóvenes les agradó el producto contra 48 adultos que también les agradó dicho producto. Calcule un intervalo de confianza del 90% para la diferencia entre las proporciones de jóvenes y adultos que les agradó el nuevo producto de carne de cerdo para la parrilla.

1.22 A continuación, se presentan algunos resultados de la Encuesta de Nacional de Hogares realizada por el Instituto Nacional de Estadística y Censos en julio del 2020, respecto la clasificación de los costarricenses ocupados según el sector institucional donde laboran.

Sector Institucional	**Fuerza de trabajo ocupada**	**Ingreso mensual promedio** *(miles de colones)*	**Desviación estándar del ingreso *** *(miles de colones)*
Sector privado	1 678 389	254,1	123,8
Sector público	276 146	451,0	234,6
Organizaciones internacionales	33 173	654,3	50,9

*estimación

Con base en la información anterior:

a. Determine el promedio total del ingreso mensual

b. Asumiendo que se desea utilizar una muestra de tamaño 1200 distribuya la muestra entre los estratos utilizando:

b.1 Afijación proporcional
b.2 Afijación óptima (Neyman)

1.23 ¿En qué circunstancias se prefiere la afijación optima en lugar de la proporcional? Justifique

1.24 En el archivo llamado “Matricula extranjeros secundaria_MEP_2022” que se puede consultar en la página del curso, se tiene la matrícula de estudiantes extranjeros en secundaria en el año 2022 según los datos administrativos del Ministerio de Educación Pública de Costa Rica. A continuación, se muestra una vista parcial de los primeros treinta registros:

CODIGO	NOMBRE	REGION	CIRES	PR	CAN	DIS	PROVINCIA	CANTON	DISTRITO	POBLADO	SECTOR	ZONA	TOTT	TOTH	TOTM
0	GRAYMAR SCHOOL	SAN JOSE CENTRAL	2	1	1	1	SAN JOSE	SAN JOSE	CARMEN	BARRIO ESCALANTE	2	1	4	2	2
0	ANGEL HIGH SCHOOL	SAN JOSE CENTRAL	3	1	1	1	SAN JOSE	SAN JOSE	CARMEN	EL CARMEN	2	1	1	0	1
3940	LICEO DE COSTA RICA	SAN JOSE CENTRAL	2	1	1	4	SAN JOSE	SAN JOSE	CATEDRAL	GONZALEZ VIQUEZ	1	1	105	105	0
3938	COLEGIO SUPERIOR DE SEÑORITAS	SAN JOSE CENTRAL	2	1	1	4	SAN JOSE	SAN JOSE	CATEDRAL	SOLEDAD	1	1	91	0	91
0	SAGRADO CORAZON	SAN JOSE CENTRAL	2	1	1	4	SAN JOSE	SAN JOSE	CATEDRAL	FRANCISCO PERALTA	2	1	4	2	2
0	SEMINARIO	SAN JOSE CENTRAL	2	1	1	4	SAN JOSE	SAN JOSE	CATEDRAL	NACIONES UNIDAS	2	1	3	3	0
4824	NOCTURNO DE HATILLO	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO 2	1	1	37	13	24
3956	LICEO EDGAR CERVANTES VILLALTA	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO CENTRO	1	1	34	18	16
3948	LICEO ROBERTO BRENES MESEN	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO 2	1	1	30	18	12
0	ADVENTISTA DE COSTA RICA	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO 1	2	1	8	5	3
0	INSTITUTO DE DESARROLLO DE INTELIGENCIA	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO 1	2	1	2	0	2
6719	C.T.P. HATILLO	SAN JOSE CENTRAL	5	1	1	10	SAN JOSE	SAN JOSE	HATILLO	HATILLO CENTRO	1	1	1	1	0
3942	LICEO DEL SUR	SAN JOSE CENTRAL	1	1	1	3	SAN JOSE	SAN JOSE	HOSPITAL	BARRIO CUBA	1	1	98	46	52
3972	LICEO MARIA AUXILIADORA	SAN JOSE OESTE	1	1	1	3	SAN JOSE	SAN JOSE	HOSPITAL	DON BOSCO	3	1	32	13	19
0	EL CARMELO	SAN JOSE CENTRAL	1	1	1	3	SAN JOSE	SAN JOSE	HOSPITAL	BARRIO CUBA	2	1	4	3	1
3943	LICEO LUIS DOBLES SEGREDA	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	SABANA ESTE	1	1	171	90	81
0	COLEGIO LA SALLE	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	SABANA SUR	2	1	37	20	17
0	LOS ANGELES	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	LAS AMERICAS	2	1	28	16	12
4156	C.T.P. EDUCACION COMERCIAL Y DE SERVICIOS	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	SABANA SUR	1	1	17	6	11
0	INSTITUTO DR. JAIM WEIZMAN	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	BARRIO HOLANDA	2	1	6	5	1
0	SAINT GEORGE HIGH SCHOOL	SAN JOSE CENTRAL	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	MATA REDONDA	2	1	5	5	0
0	COLEGIO BILINGÜE DEL VALLE	SAN JOSE OESTE	1	1	1	8	SAN JOSE	SAN JOSE	MATA REDONDA	SABANA SUR	2	1	1	1	0
3941	LICEO DE SAN JOSE	SAN JOSE OESTE	1	1	1	2	SAN JOSE	SAN JOSE	MERCED	BARRIO MEXICO	1	1	152	72	80
3968	LICEO DE PAVAS	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	PAVAS	1	1	194	93	101
3978	COLEGIO DE RINCON GRANDE	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	LOMAS DEL RIO	1	1	153	73	80
0	BRITANICO DE COSTA RICA	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	SANTA CATALINA	2	1	49	25	24
6530	C.T.P. PAVAS	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	PAVAS CENTRO	1	1	36	10	26
6530	SECCION NOCTURNA C.T.P. DE PAVAS	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	PAVAS CENTRO	1	1	31	7	24
6631	UNIDAD PEDAGOGICA DANIEL ODUBER QUIROS	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	VILLA ESPERANZA	1	1	28	11	17
0	COLEGIO HUMBOLDT	SAN JOSE OESTE	2	1	1	9	SAN JOSE	SAN JOSE	PAVAS	ROHRMOSER	2	1	24	11	13

a) Seleccione una muestra aleatoria de 50 centros educativos, utilizando la posición dentro del listado como identificador
b) Seleccione una muestra sistemática de 50 centros educativos.
c) Ordene la población de centros educativos según sector y matrícula total.
d) Seleccione una muestra sistemática de 50 centros educativos de c) y comente el resultado
e) Divida la población de centros educativos en dos estratos (por zona y matrícula total) y seleccione una muestra de 50 centros educativos distribuidos en forma proporcional

2. Verificación de Hipótesis

2.1 Un estudio realizado hace 5 años por una empresa de mercadeo, determinó que Paris es el lugar más costoso para vivir de 12 ciudades de la Unión Europea. En ese entonces se concluyó que dicha ciudad se clasificaba en segundo lugar en costo de vivienda para una familia de 4 personas, con una unidad de renta con un costo promedio de $4 292 al mes. Suponga que el director general en una compañía piensa que esa cantidad es muy alta y decide realizar su propia encuesta. Su asistente se comunica con propietarios de 55 unidades de rentas seleccionadas al azar y encuentra que el costo promedio muestral es $4 080, con una desviación estándar de $386. Con base en los resultados anteriores y empleando un nivel de significancia de 0,01, realice una verificación de la hipótesis para determinar si el costo promedio publicado por la empresa es demasiado alto. Si rechaza la hipótesis nula, discuta si los resultados son sustantivos.

2.2 Una empresa manufacturera ha estado promediando 18,2 pedidos por semana durante varios años. No obstante, durante el período de crisis actual los pedidos disminuyeron. Suponga que el gerente de producción de la compañía muestrea al azar 22 semanas y encuentra una media muestral de 15,6 pedidos, con una desviación estándar muestral de 2,3 pedidos. Verifique si el número promedio de pedidos disminuyó utilice un nivel de significancia del 0,10.

2.3 Se conoce que regularmente el tiempo de espera promedio para entrega de facturas en un Ministerio de Hacienda es de 60 minutos. Producto de las quejas de los usuarios se ha tomado una serie de medidas para disminuir el citado tiempo. Para verificar si las medidas han sido eficaces se seleccionó aleatoriamente 100 usuarios que hicieron uso del servicio y se encontró un tiempo promedio de espera de 55 minutos con una desviación estándar de 6,5 minutos.

a) Establezca la hipótesis a verificar (simbólicamente y en enunciado)
b) ¿Qué decidiría usted si se usa α = 0,05? Realice todos los cálculos y concluya en términos del problema.
c) ¿Cuál error se puede cometer al tomar la decisión de b) y por qué?
d) Si en este caso, μ realmente fuera igual a 45 minutos. Determine es la magnitud del error tipo II (β).

2.4 Con base en datos de diversos estudios, algunos investigadores indican que el 60% de los jóvenes de la zona urbana de Costa Rica padece de estrés. Dado el mundo actual tan acelerado y como consecuencia de la pandemia de COVID-19 se presume que dicho porcentaje ha aumentado. Según la muestra del ejercicio 1.20, dicho porcentaje llegó al 63%

a) Con un nivel de significancia del 1%, realice la verificación de hipótesis correspondiente para determinar que el porcentaje de estrés de los jóvenes de la zona urbana se incrementó.
b) Determine la magnitud del error tipo 2, si el verdadero porcentaje de estrés es de 68%.

2.5 Dentro de los aspectos estudiados en la encuesta del ejercicio 1.20, está la edad de inicio de los jóvenes de la primera relación sexual. Se tiene la premisa que la edad promedio de inicio de la primera relación sexual de los jóvenes de la zona rural es inferior al promedio respectivo de la zona urbana. Según la encuesta dicha edad promedio para la zona rural es de 16,2 años y para el área urbana de 18,1 años. Con un 5% de significancia,

verifique la hipótesis de que el promedio de la zona rural es inferior (asuma una desviación estándar para zona rural de 12 años y 15 años para la zona urbana).

2.6 Un artículo de USA Today informó que sólo hay trabajo para uno de cada cuatro nuevos graduados de la universidad a causa de la sobrepoblación de graduados universitarios y una débil economía. Se sospecha que con la crisis económica actual dicha proporción ha disminuido, por ello se selecciona aleatoriamente 20 graduados y se determina que cuatro de ellos tienen trabajo. Utilizando un nivel de significancia del 9,13% verifique la hipótesis de que la proporción de graduados que encuentran trabajo ha disminuido.

2.7 Suponga que un corredor de bienes raíces se encuentra interesado en comparar los precios de las propiedades del GAM versus fuera del GAM. Por ello, selecciona aleatoriamente una muestra de 13 propiedades del GAM generando un precio promedio de las propiedades de ¢55 millones con una desviación estándar de ¢15 millones, a su vez, seleccionó una muestra de 11 propiedades fuera del GAM y obtuvo un promedio de ¢42 millones con una desviación estándar de ¢9 millones. Con base en los resultados anteriores, realice la verificación de la hipótesis de que el precio promedio de las propiedades del GAM es superior al promedio de las propiedades fuera del GAM. Emplee un nivel de significancia del 10%

2.8 Asuma que el Ministerio de Economía necesita conocer las expectativas de mejoras de la economía del país. Para ello consultó una muestra de profesionales en el campo empresarial y otra de economistas. Se muestran a continuación los resultados obtenidos:

PROFESIONALES	PROPORCION*	TAMAÑO DE MUESTRA
Empresarios	0,35	75
Economistas	0,20	90

*Proporción de expertos que consideran que la economía del país se recuperará el próximo año

¿Generan los datos evidencia estadística suficiente que señale una diferencia entre las proporciones presentadas? Utilice un nivel de significancia del 10%

2.9 En un estudio para estimar la proporción de residentes de cierta ciudad y sus suburbios que están a favor de la construcción de una planta de energía nuclear, se encuentra que 63 de 100 residentes urbanos están a favor de la construcción mientras que sólo 59 de 125 residentes rural la favorecen. ¿Hay una diferencia significativa entre la proporción de residentes urbanos y suburbanos que favorecen la construcción de la planta nuclear? Utilice un nivel de significancia del 5%.

2.10 En una fábrica de conservas de fruta se desea verificar la Ho de que las latas resultan en promedio con un peso no inferior a 1 kg. Se sabe que el tamaño de la fruta puede introducir una variación en los pesos de las latas de manera que estos se distribuyan normalmente con una dispersión del 8%. Se efectúa una muestra de 100 latas en la que se determina los pesos, resultando con un peso promedio = 980 gr. Interesa saber si la muestra comprueba o rechaza la Ho, fijando una significancia del 5%.

2.11 Un consorcio norteamericano de tiendas de departamentos tiene sucursales en Multiplaza de Este y en Multiplaza de Oeste. El gerente de ventas desea ajustar las promociones y formato de los espacios, en función del tipo de clientes que las visitan y para ello decide comparar los promedios de los montos comprados por los clientes. Para lograr dicho fin selecciona dos muestras independientes de cada tienda, obteniendo los datos que se presentan a continuación (las cifras están en dólares):

Sucursal	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Este	77,58	91,73	112,61	71,79	61,22	51,82	155,40	105,86	110,82	55,85				
Oeste	218,19	120,22	117,38	71,96	32,92	51,87	61,47	51,96	61,79	51,74	102,4	50,57	91,79	112,32

Con base en la información anterior verifique la hipótesis que existe diferencia significativa entre los montos promedios de las compras de los clientes de las tiendas del Este y Oeste. Utilice un nivel de significancia del 5%

2.12 Los datos del ejercicio anterior se procesaron mediante el paquete de cómputo EXCEL obteniendo los siguientes resultados:

Prueba t para dos muestras suponiendo varianzas desiguales

	Este	*Oeste*
Media	89,468	85,47
Varianza	1045,73673	2275,09745
Observaciones	10	14
Diferencia hipotética de las medias	0	
Grados de libertad	22	
Estadístico t	0,24463669	
P(T<=t) una cola	0,40450252	
Valor crítico de t (una cola)	1,71714437	
P(T<=t) dos colas	0,80900504	
Valor crítico de t (dos colas)	2,07387307	

Compare los resultados con lo obtenido en el ejercicio anterior.

2.13 Una compañía de productos de belleza está desarrollando un nuevo champú y para ello está interesada en la altura de la espuma (en mm). La altura de la espuma tiene distribución normal con una desviación estándar de 20 mm. La compañía está interesada en probar la hipótesis de que la media de la altura es de 175 mm contra la alternativa de que es mayor que ese valor. Para ello toma una muestra de tamaño 10 y rechaza la hipótesis si el promedio de la muestra es mayor que 185 mm.

a) Calcule la probabilidad de cometer un error tipo I.
b) Calcule la probabilidad de cometer un error tipo II si la media verdadera es de 195 mm.
c) Calcule la potencia de la prueba para este valor específico de la media.

2.14 Todos los días el United States Customs Service descubre unos 28 millones de dólares en artículos introducidos ilegalmente en el país, con una desviación estándar de 16 millones de dólares al día. En 64 días elegidos aleatoriamente en el año 2019. La oficina descubrió un promedio de 30,3 millones de dólares en artículos de contrabando. ¿Indica la muestra que el jefe de aduanas debe preocuparle el hecho de que se haya rebasado el nivel histórico del contrabando, si usa un nivel de significancia del 5%?

2.15 Comente la siguiente afirmación: “El error tipo 1 es más grave que el error tipo 2”

2.16 ¿En qué circunstancias se puede cometer el error tipo 2?

2.17 Si un error tipo II es considerado más serio en cierta situación, ¿seleccionaría usted un valor alfa alto o bajo? Justifique su respuesta.

2.18 ¿Porqué al formular la hipótesis nula siempre debe ponerle el signo de igualdad?

3. Análisis de Variancia

3.1 Se desea probar que el promedio de producción de petróleo es diferente para al menos un par de Continentes, de manera que a continuación se presentan los datos:

Continente A	2	3	5	2	1	2	Promedio = 2,50
Continente B	9	8	7	7			Promedio = 7,75
Continente C	3	2	2				Promedio = 2,33

Donde $\Sigma x = 53$ $\Sigma x^2 = 307$ promedio total = 4,08 SCE = 12,9167 SCT = 90,9231

Para los datos presentados realice la prueba respectiva a efectos de verificar la hipótesis de que al menos dos promedios de producción de petróleo de los continentes son diferentes, use un nivel de significancia del 5% y concluya en términos del problema. (Datos en miles de millones de barriles). En caso de rechazar la hipótesis, efectúe la prueba para la diferencia entre pares de medias (DMS)

3.2 El ejercicio 3.1 se procesó mediante el paquete estadístico MegaStat generando los siguientes resultados:

One factor ANOVA

Mean	*n*	*Std. Dev*	
2,5	6	1,38	Continente A
7,8	4	0,96	Continente B
2,3	3	0,58	Continente C
4,1	13	2,75	Total

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Treatment	78,01	2	39,003	30,20	,0001
Error	12,92	10	1,292		
Total	90,92	12			

Post hoc analysis
p-values for pairwise t-tests

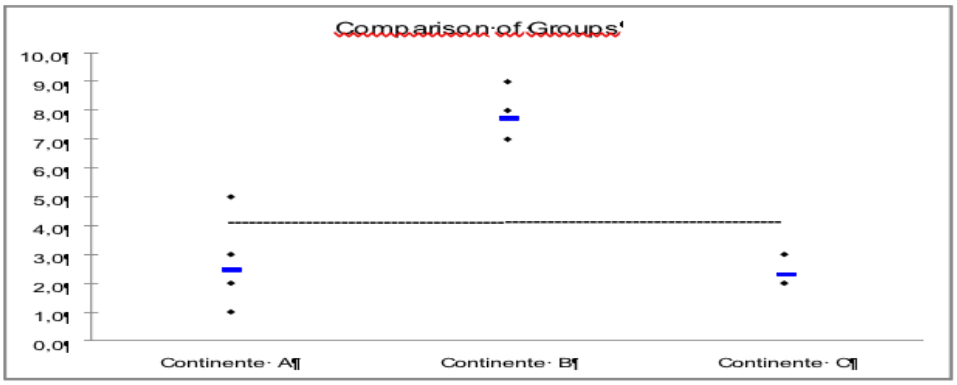
	Continente C	Continente A	Continente B
	2,3	2,5	7,8
Continente C	2,3		
Continente A	2,5	,8399	
Continente B	7,8	,0001	3,08E-05

Tukey simultaneous comparison t-values (d.f. = 10)

	Continente C	Continente A	Continente B
	2,3	2,5	7,8
Continente C	2,3		
Continente A	2,5	0,21	
Continente B	7,8	6,24	7,16

critical values for experimentwise error rate:

0,05	2,74
0,01	3,73



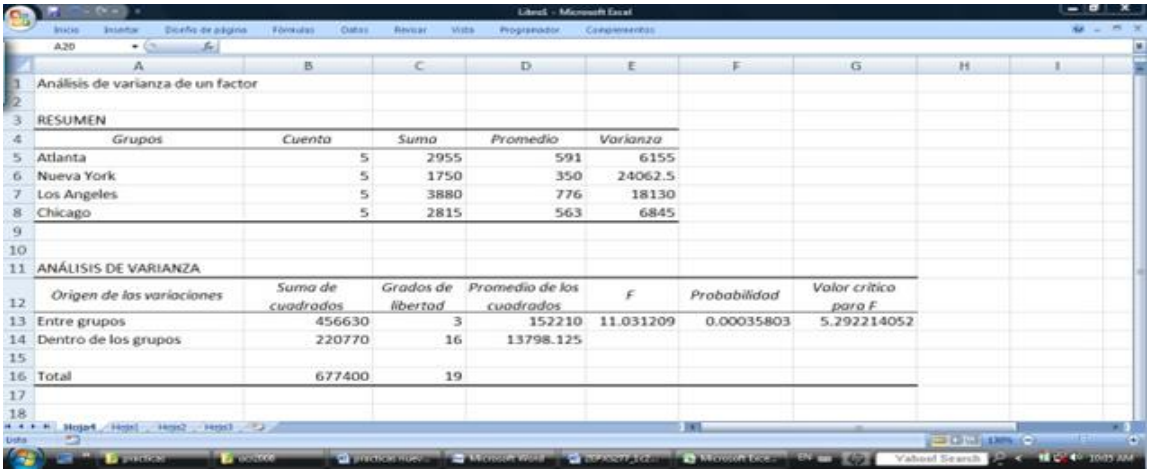
Compare los resultados con los obtenidos en el ejercicio 3.1

3.3 Los costos de transporte de las familias suelen ser más altos de lo pensado, ya que incluyen el pago del auto, combustible, reparaciones, estacionamiento y transporte público. A veinte familias seleccionadas al azar de cuatro ciudades de Estados Unidos se les solicitó el costo mensual y los resultados fueron los siguientes:

Atlanta	Nueva York	Los Ángeles	Chicago
$650	$250	$850	$540
480	525	700	450
550	300	950	675
600	175	780	550
675	500	600	600

Utilizando un nivel de significancia del 1% verifique la hipótesis de que al menos un par de promedios de los costos de transporte de las ciudades es diferente
En caso de rechazar la hipótesis, efectúe la verificación de la prueba para la diferencia entre pares de medias de Tukey

3.4 Los datos del ejercicio 3.3 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:



Análisis de varianza de un factor

RESUMEN

Grupos	Cuenta	Suma	Promedio	Varianza
Atlanta	5	2955	591	6155
Nueva York	5	1750	350	24062.5
Los Angeles	5	3880	776	18130
Chicago	5	2815	563	6845

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	456630	3	152210	11.031209	0.00035803	5.292214052
Dentro de los grupos	220770	16	13798.125			
Total	677400	19				

Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 3.3.

3.5 En un estudio realizado por el Departamento de Salud y Educación Física en una Universidad, se asignaron 3 dietas durante 3 días a 6 sujetos utilizando un diseño de bloques completos aleatorizados. Los sujetos, que desempeñan el papel de bloques, recibieron las siguientes 3 dietas en orden aleatorio:
Dieta 1: grasas mixtas y carbohidratos,
Dieta 2: alta en grasas,
Dieta 3: alta en carbohidratos.

Al terminar el periodo de tres días se puso a cada sujeto en una banda caminadora y se midió el tiempo, en segundos, que transcurría hasta que se sentían exhaustos. Efectúe un análisis de varianza correspondiente. Los datos registrados son los siguientes:

Dieta	Sujetos					
	1	2	3	4	5	6
1	84	35	91	57	56	45
2	91	48	71	45	61	61
3	122	53	110	71	91	122

3.6 Una empresa vende champú para cabello normal, grasoso y teñido en el mercado local y tiene la premisa que la época del año afecta la venta del producto, por tal motivo realiza obtiene una muestra de las ventas del champú y las presenta en la siguiente tabla:

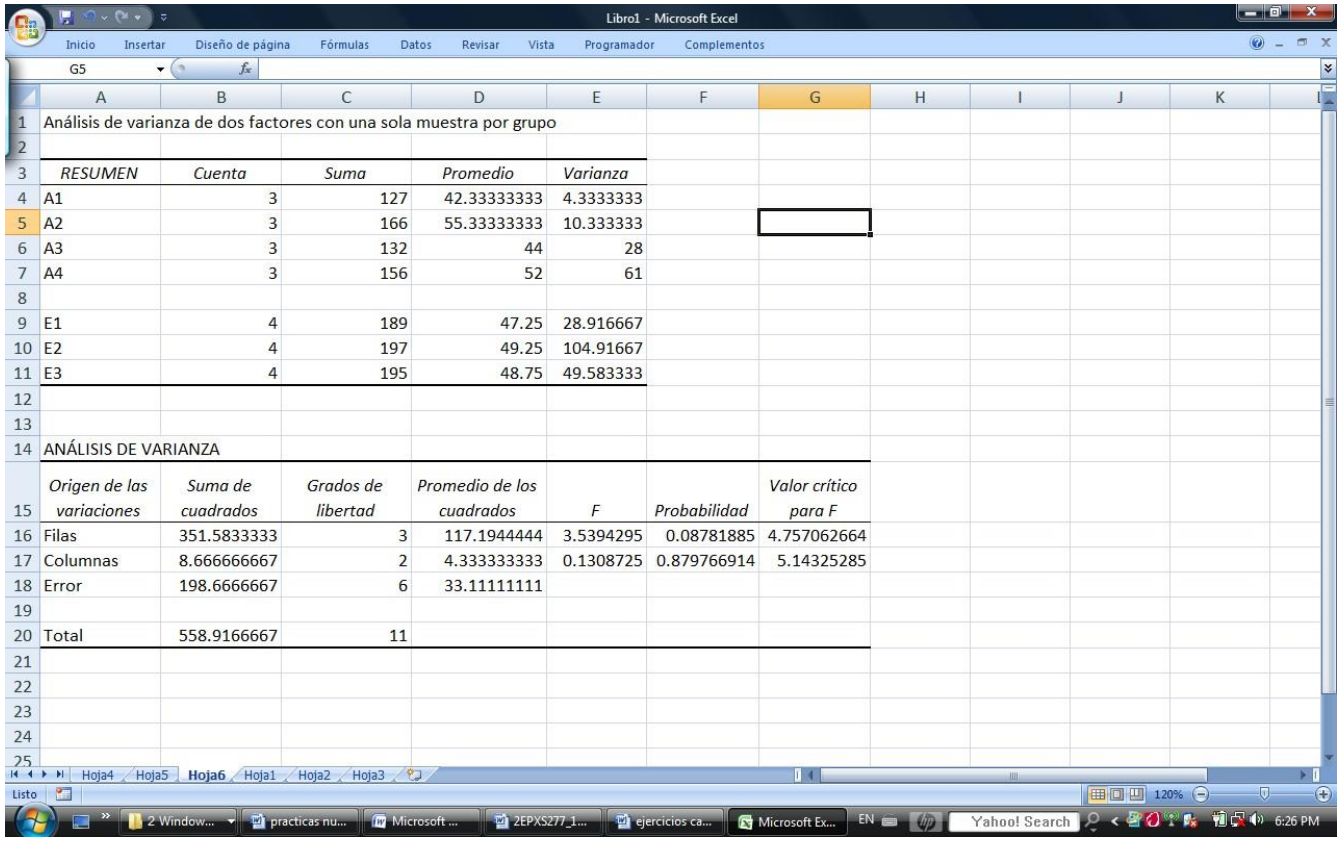
TRIMESTRE	Normal	Grasoso	Teñido
Enero-Marzo	6	7	11
Abril-Junio	10	11	13
Julio-Setiembre	12	10	7
Octubre-diciembre	8	8	7

Utilizando un nivel de significancia del 1% verifique la hipótesis de que al menos un par de las ventas promedios de los tipos de champú es diferente considerando el efecto de los trimestres

3.7 Cuatro tipos de bastidores publicitarios fueron instalados en 12 establecimientos de venta al detalle. Se asignaron aleatoriamente 3 establecimientos para cada uno de los bastidores, con el fin de estudiar el impacto de punto de venta. Dados los datos de la siguiente tabla, verifique la hipótesis de que existen diferencias en al menos un par de los valores medios de las ventas para los tres tipos de establecimientos considerando el efecto del tipo de bastidor. Utilice un 5% de significancia.

TIPO D BASTIDOR	ESTABLECIMIENTO 1	ESTABLECIMIENTO 2	ESTABLECIMIENTO 3
A1	40	44	43
A2	53	54	59
A3	48	38	46
A4	48	61	47

3.8 Los datos del ejercicio 3.7 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:



Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	Cuenta	Suma	Promedio	Varianza
A1	3	127	42.33333333	4.3333333
A2	3	166	55.33333333	10.333333
A3	3	132	44	28
A4	3	156	52	61
E1	4	189	47.25	28.916667
E2	4	197	49.25	104.91667
E3	4	195	48.75	49.583333

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Filas	351.5833333	3	117.1944444	3.5394295	0.08781885	4.757062664
Columnas	8.666666667	2	4.333333333	0.1308725	0.879766914	5.14325285
Error	198.6666667	6	33.11111111			
Total	558.9166667	11				

Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 3.7.

3.9 El ejercicio anterior se amplió la cantidad de datos en cada celda, es decir se cuenta con tres replicaciones para cada tratamiento y bloque, de la siguiente manera:

TIPO DE BASTIDOR	ESTABLECIMIENTO 1	ESTABLECIMIENTO 2	ESTABLECIMIENTO 3
A1	40	44	43
A1	40	44	42
A1	41	43	43
A2	53	54	59
A2	53	54	59
A2	52	53	58
A3	48	38	46
A3	48	37	46
A3	47	38	47
A4	48	61	47
A4	48	60	46
A4	47	61	47

Los datos anteriores se procesaron mediante el paquete de cómputo Excel generando los siguientes resultados:

Análisis de varianza de dos factores con varias muestras por grupo				
RESUMEN	E1	E2	E3	Total
A1				
Cuenta	3	3	3	9
Suma	121	131	128	380
Promedio	40,3333333	43,6666667	42,6666667	42,2222222
Varianza	0,33333333	0,33333333	0,33333333	2,44444444
A2				
Cuenta	3	3	3	9
Suma	158	161	176	495
Promedio	52,6666667	53,6666667	58,6666667	55
Varianza	0,33333333	0,33333333	0,33333333	8
A3				
Cuenta	3	3	3	9
Suma	143	113	139	395
Promedio	47,6666667	37,6666667	46,3333333	43,8888889
Varianza	0,33333333	0,33333333	0,33333333	22,3611111
A4				
Cuenta	3	3	3	9
Suma	143	182	140	465
Promedio	47,6666667	60,6666667	46,6666667	51,6666667
Varianza	0,33333333	0,33333333	0,33333333	46
Total				
Cuenta	12	12	12	
Suma	565	587	583	
Promedio	47,0833333	48,9166667	48,5833333	
Varianza	21,3560606	86,0833333	39,9015152	

ANÁLISIS DE VARIANZA						
Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Muestra	1013,19444	3	337,731481	1013,19444	2,14513E-25	3,00878657
Columnas	22,8888889	2	11,4444444	34,3333333	9,10862E-08	3,40282611
Interacción	599,555556	6	99,9259259	299,777778	2,4125E-21	2,50818882
Dentro del grupo	8	24	0,33333333			
Total	1643,63889	35				

Interprete los resultados

3.10 En una fábrica de aceite vegetal comestible, la calidad se afecta por la cantidad de impurezas dentro del aceite, que causan oxidación y esto repercute en las características de sabor y color del producto final. El proceso de "blanqueo" es responsable de eliminar las impurezas y una forma de medir su eficacia es "midiendo" el color del aceite. Para generar una aproximación a la solución de este problema inicialmente se decide estudiar el efecto de la temperatura y el porcentaje de arcilla en el color del aceite en el laboratorio. El diseño y datos de las pruebas experimentales se presentan a continuación:

Temperatura	Porcentaje de arcilla

	0,8		0,9		1,0		1.1	
90	5,8	5,9	5,4	5,5	4,9	5,1	4,5	4,4
100	5,0	4,9	4,8	4,7	4,0	4,4	4,1	4,3
110	4,7	46,0	4,4	4,4	4,1	4,0	3,7	3,6

a) Construya el modelo estadístico y formule las hipótesis pertinentes.
b) Indique el nombre del diseño utilizado
c) En condiciones reales, por lo general, se utiliza 1,1% de arcilla y 100 grados de temperatura. ¿Porqué cree usted que se eligieron esos niveles de prueba para el experimento?
d) Realice el análisis de varianza para comprobar las hipótesis y haga las respectivas conclusiones.
e) Apoyándose en las gráficas de efectos, indique cual es la relación general entre el color y los factores controlados en el rango del experimento.
f) A partir de la gráfica de interacciones, considera usted que hay un efecto no lineal.
g) Si se considera que el nivel mínimo de blancura que se acepta es 4,8 ¿qué tratamiento utilizaría?
h) ¿Considera usted que vale la pena plantear el experimento en condiciones reales?
i) ¿Qué cambios sugiere usted que se hagan si se decide realizar el experimento en condiciones reales?

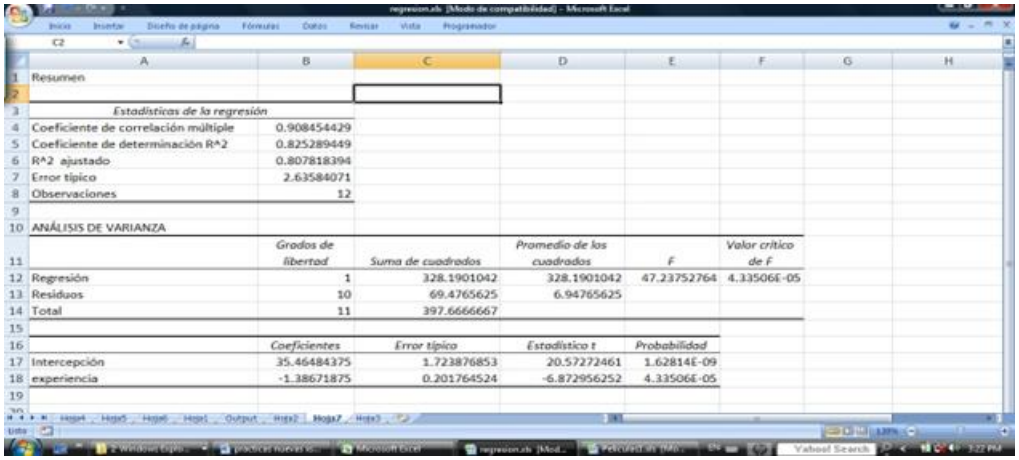
4. Correlación y Regresión

4.1 En la siguiente tabla se presentan datos relativos al número de semanas de experiencia en un empleo que implica la conexión de componentes electrónicos en miniatura y el número de componentes que fueron rechazados durante la última semana de 12 trabajadores seleccionados aleatoriamente:

Obrero muestreado	1	2	3	4	5	6	7	8	9	10	11	12
Semanas de experiencia (X)	7	9	6	14	8	12	10	4	2	11	1	8
Número de rechazos (Y)	26	20	28	16	23	18	24	26	38	22	32	25

a) Determine el coeficiente de correlación e interprételo. A su vez, verifique la hipótesis de que $\rho<0$ con α=10% e indique si es posible establecer una relación lineal entre las variables analizadas.
b) Determine la ecuación de regresión para predecir el número de componentes rechazados
c) Verifique la hipótesis que existe relación entre las variables (coeficiente de regresión menor que cero), con un nivel de significancia de un 5%
d) Estime el número de componentes rechazados en el caso de un empleado con tres semanas de experiencia en el puesto.
e) Construya un intervalo de confianza del 95% para estimar el número medio de rechazos en el caso de empleados con tres semanas de experiencia en las operaciones.
f) Calcule un intervalo de confianza al 95% para estimar el número de componentes rechazos en el caso de un empleado con tres semanas de experiencia en el puesto.

4.2 Los datos del ejercicio 4.1 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:



	A	B	C	D	E	F
1	Resumen					
2						
3	*Estadísticas de la regresión*					
4	Coeficiente de correlación múltiple	0.908454429				
5	Coeficiente de determinación R^2	0.825289449				
6	R^2 ajustado	0.807818394				
7	Error típico	2.63584071				
8	Observaciones	12				
9						
10	ANÁLISIS DE VARIANZA					
11		*Grados de libertad*	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
12	Regresión	1	328.1901042	328.1901042	47.23752764	4.33506E-05
13	Residuos	10	69.4765625	6.94765625		
14	Total	11	397.6666667			
15						
16		*Coeficientes*	*Error típico*	*Estadístico t*	*Probabilidad*	
17	Intercepción	35.46484375	1.723876853	20.57272461	1.62814E-09	
18	experiencia	-1.38671875	0.201764524	-6.872956252	4.33506E-05	

Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 4.1

4.3 Un analista toma una muestra aleatoria de 10 embarques por camión realizados por una compañía y registra la distancia en millas, como el tiempo de entrega al medio día más cercano (al momento al que el embarque estuvo listo para su carga). A continuación, se presenta la información mencionada:

Embarque muestreado	1	2	3	4	5	6	7	8	9	10
Distancia en millas (X)	825	215	1 070	550	480	920	1 350	325	670	1 215
Tiempo de entrega (días) (Y)	3,5	1,0	4,0	2,0	1,0	3,0	4,5	1,5	3,0	5,0

a) Calcule la bondad de ajuste e interprétela
b) Determine la ecuación de regresión de mínimos cuadrados y trace en el diagrama de dispersión la línea de regresión
c) Estime el tiempo de entrega a partir del momento en que el embarque esta listo para su carga en un embarque de 1000 millas
d) Calcule el error estándar de estimación
e) Verifique la hipótesis H_1: $B > 0$ para los datos de distancia de acarreo y tiempo de entrega con α=5%
f) Elabore un intervalo de confianza del 95% para el tiempo de entrega medio respecto a la distancia de acarreo de 1000 millas.
g) Elabore un intervalo de confianza del 95% para el tiempo de entrega de un solo embarque dada una distancia de acarreo de 1000 millas.

4.4 El dueño de un restaurante desea relacionar el monto de las cuentas con las propinas dadas a sus saloneros. Si selecciona una muestra de 16 cuentas y establece la siguiente relación:

$y = -199{,}78 + 107{,}78\,x$ donde x = monto de la cuenta en miles de ¢
y = monto de la propina en ¢

a) Estime la propina que se espera con una cuenta de ¢20000,00
b) Si el error estándar de estimación es de 373,45 y la desviación estándar de los montos de las cuentas es de 7,59 determine Sb

4.5 Una compañía administra una prueba a los futuros representantes de ventas antes que empiecen a trabajar. El Director General quiere determinar la relación entre las puntuaciones de la prueba y las ventas logradas por los representantes al final de un año de trabajo en el campo. Se recolectaron los siguientes datos:

Número del vendedor	Puntuación de la prueba	Número de unidades vendidas
1	**2,6**	**95**
2	**3,7**	**140**
3	**2,4**	**85**
4	**4,5**	**180**
5	**2,6**	**100**
6	**5,0**	**195**
7	**2,8**	**115**
8	**3,0**	**136**
9	**4,0**	**175**
10	**3,4**	**150**

a) Calcule la línea de regresión de mínimos cuadrados que podría servir para predecir las ventas a partir de las puntuaciones de los futuros representantes.
b) ¿Cuánto aumenta el número esperado de unidades vendidas por cada incremento de un punto en la puntuación de la prueba?
c) Utilice la línea de regresión de mínimos cuadrados para predecir el número de unidades que vendería el representante que haya obtenido una puntuación promedio en la prueba.
d) Calcule un intervalo de estimación al 95 %, para la verdadera pendiente de la recta (FRP) y el verdadero valor del intercepto de la FRP.
e) Calcule el coeficiente de determinación e interprételo.

4.6 Un coleccionista de arte estudia la relación entre la antigüedad de la pintura y el precio de subasta. En ese sentido estudió una muestra de 25 pinturas y logró establecer la siguiente ecuación:

$y = 1497{,}7 + 9{,}8\,x$ donde x= antigüedad de la pintura en años
y = precio de subasta en miles colones
Correlación (antigüedad vs precio de subasta) = 0,58

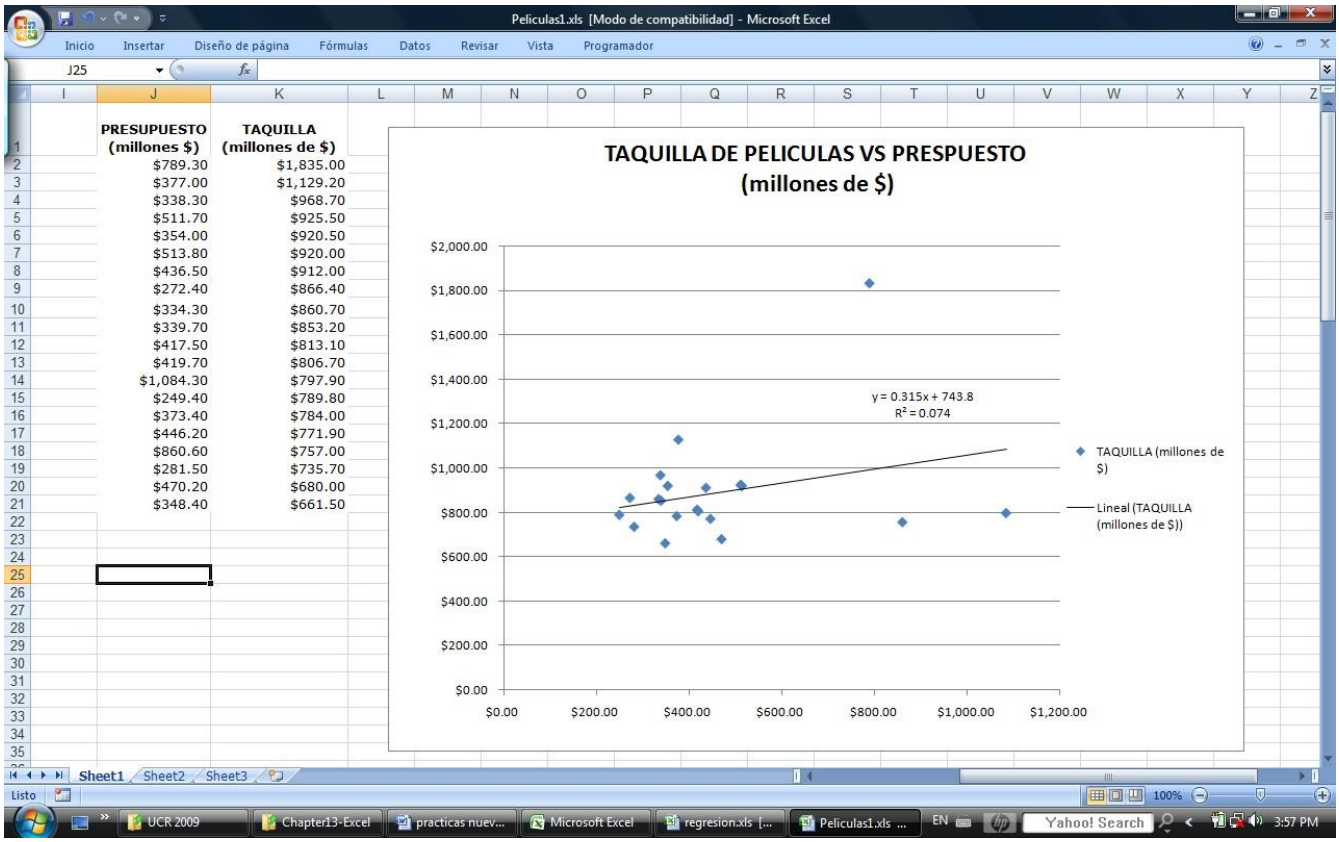
a) Estime el precio de subasta de una pintura con 100 años.
b) Se piensa que la cantidad de postores en la subasta influye en el precio, por lo tanto el coleccionista calcula los siguientes coeficientes de correlación:
Correlación (antigüedad vs postores) = - 0,60
Correlación (precio subasta vs postores) = - 0,06
Con base en lo anterior, determine el coeficiente de correlación entre (antigüedad vs precio de subasta) eliminando el efecto de los postores

4.7 A continuación, se presentan las 20 películas con mayores taquillas del año 1982 al 2004:

PELICULA	AÑO	TAQUILLA	PRESUPUESTO
Titanic	1997	$1,835.00	$789.30
LOTR: The Return of the King	2003	$1,129.20	$377.00
Harry Potter and the Sorcerer's Stone	2001	$968.70	$338.30
Star Wars: Episode I -- The Phantom Menace	1999	$925.50	$511.70
The Lord of the Rings: The Two Towers	2002	$920.50	$354.00
Jurassic Park	1993	$920.00	$513.80
Shrek 2	2004	$912.00	$436.50
Harry Potter and the Chamber of Secrets	2002	$866.40	$272.40
The Lord of the Rings: The Fellowship of the Ring	2001	$860.70	$334.30
Finding Nemo	2003	$853.20	$339.70
Independence Day	1996	$813.10	$417.50
Spider-Man	2002	$806.70	$419.70
Star Wars	1977	$797.90	$1,084.30
Harry Potter and the Prisoner of Azkaban	2004	$789.80	$249.40
Spider-Man 2	2004	$784.00	$373.40
The Lion King	1994	$771.90	$446.20
E.T.	1982	$757.00	$860.60
The Matrix: Reloaded	2003	$735.70	$281.50
Forrest Gump	1994	$680.00	$470.20
The Sixth Sense	1999	$661.50	$348.40

a) Elabore un diagrama de dispersión, bajo el entendido que el presupuesto de la película determina la taquilla.
b) Calcule el coeficiente de correlación y la bondad de ajuste para confirmar lo encontrado en a).
c) Con base en a) y b) se puede concluir que es apropiado establecer una relación lineal entre el monto presupuestado a las películas y la taquilla

4.8 Los datos del ejercicio 4.7 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:



PRESUPUESTO (millones $)	TAQUILLA (millones de $)
$789.30	$1,835.00
$377.00	$1,129.20
$338.30	$968.70
$511.70	$925.50
$354.00	$920.50
$513.80	$920.00
$436.50	$912.00
$272.40	$866.40
$334.30	$860.70
$339.70	$853.20
$417.50	$813.10
$419.70	$806.70
$1,084.30	$797.90
$249.40	$789.80
$373.40	$784.00
$446.20	$771.90
$860.60	$757.00
$281.50	$735.70
$470.20	$680.00
$348.40	$661.50

Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 4.7.

4.9 Con el fin de determinar la relación entre la calificación de su desempeño laboral (y) y las calificaciones en cuatro exámenes, el departamento de personal de cierta empresa industrial realizó un estudio en el que participaron 12 sujetos. Los datos son los siguientes:

y	x_1	x_2	x_3	x_4
11.2	56.5	71.0	38.5	43.0
14.5	59.5	72.5	38.2	44.8
17.2	69.2	76.0	42.5	49.0
17.8	74.5	79.5	43.4	56.3
19.3	81.2	84.0	47.5	60.2
24.5	88.0	86.2	47.4	62.0
21.2	78.2	80.5	44.5	58.1
16.9	69.0	72.0	41.8	48.1
14.8	58.1	68.0	42.1	46.0
20.0	80.5	85.0	48.1	60.3
13.2	58.3	71.0	37.5	47.1
22.5	84.0	87.2	51.0	65.2

Estime los coeficientes de regresión del modelo: $\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4$.

4.10 Un banco desea estimar el ingreso de altos ejecutivos con base en el valor de la casa de habitación, los años de educación, la edad y el monto de la hipoteca de la casa.

Ingreso*	Educa**	Edad**	Hipoteca***	Valor***	Ingreso*	Educa**	Edad**	Hipoteca***	Valor***
1 907 533	14	53	107 920	130 640	1 846 000	14	37	72 136	124 960
1 874 400	15	49	68 728	210 160	1 869 667	14	50	86 904	153 360
1 931 200	14	44	91 448	225 496	1 921 733	14	50	82 360	158 472
1 907 533	14	39	91 448	102 808	1 907 533	15	52	98 832	186 872
1 893 333	14	53	101 672	214 704	1 898 067	15	47	100 536	155 632
1 803 400	14	46	56 232	172 672	1 973 800	15	49	106 784	245 944
1 912 267	15	42	64 752	161 880	1 898 067	15	53	86 904	189 144
1 926 467	14	49	114 736	312 968	1 921 733	16	58	85 200	84 064
1 931 200	13	37	104 512	210 160	1 912 267	13	42	98 264	221 520
1 756 067	14	43	51 120	76 680	1 935 933	14	46	92 584	80 656
1 888 600	14	48	102 808	188 576	1 898 067	15	50	85 200	194 824
1 912 267	15	54	81 224	123 256	1 822 333	14	45	78 952	211 864
1 798 667	14	44	74 976	278 320					

* colones **en años ***miles de colones

Los datos anteriores se procesaron en el paquete de cómputo EXCEL y se generó lo siguiente:

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,799336688
Coeficiente de determinación R^2	0,638939141
R^2 ajustado	0,566726969
Error típico	32757,72036
Observaciones	25

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	4	37978378657	9494594664	8,85	0,000278236
Residuos	20	21461364866	1073068243		
Total	24	59439743524			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*
Intercepción	1334149,05	165456,3662	8,0634495	0,00000010	989013,1134	1679284,9773
Educa**	30400,71	13368,05127	2,2741321	0,03412519	2515,4478	58285,9804
Edad**	-1979,64	1726,758092	-1,1464465	0,26514215	-5581,5900	1622,3185
Hipoteca***	2,64	0,474397587	5,5600656	0,00001924	1,64811	3,6273
Valor***	-0,10	0,121303802	-0,7834875	0,44252096	-0,3481	0,1580

a) Establezca la ecuación de regresión
b) Estime el ingreso de un ejecutivo de 50 años de edad, 14 años de estudios, con una casa de una hipoteca de ¢90 millones y un valor de ¢190 millones

c) Si analiza cada coeficiente de regresión y respectiva prueba del estadístico t, qué se puede concluir sobre el aporte de cada variable al modelo de regresión múltiple.
d) Calcule el coeficiente de correlación ingreso vs edad
e) Calcule el coeficiente de correlación ingreso vs valor
f) Calcule el coeficiente de correlación valor vs edad
g) Calcule el coeficiente de correlación ingreso vs valor eliminado el efecto de la edad

5. Estadística No Paramétrica

5.1 En el estudio se investigó el nivel educativo para una muestra de conductores, a fin de analizar su nivel educativo y la aceptación de los biocombustibles, presentando los resultados del siguiente cuadro:

DISTRIBUCION DE LOS CONDUCTORES SEGÚN NIVEL DE EDUCATIVO Y ACEPTACION DE BIOCOMBUSTIBLES

Nivel educativo	Aceptación de biocombustibles	
	No	Sí
Sin estudios	122	141
Primaria	135	188
Secundaria	83	93

Con base en la información presentada en el cuadro, verifique que la aceptación de biocombustibles depende de su nivel educativo. Utilice un nivel de confianza del 99% y concluya en términos del problema.

5.2 Un estudio acerca de la dependencia entre la edad y la cantidad de presión que siente el personal de ventas en su trabajo reveló la siguiente información que se muestra a continuación:

DISTRIBUCION DE LOS EMPLEADOS SEGÚN SU EDAD Y NIVEL DE PRESION DEL TRABAJO

EDAD	GRADO DE PRESION EN EL TRABAJO		
	Bajo	Medio	Alto
Menores a 25 años	20	18	22
De 25 a menos de 40 años	50	46	44
De 40 a menos de 60 años	58	63	59
60 años y más	34	43	43

Con base en la información presentada en el cuadro, verifique que el grado de presión en el trabajo depende de la edad. Utilice un nivel de confianza del 95% y concluya en términos del problema.

5.3 Se desea determinar si las opiniones de los votantes residentes en la provincia de San José respecto a una nueva reforma fiscal son independientes de sus niveles de ingreso, con un nivel de significancia del 5%. Los sujetos de una muestra aleatoria de 1000 votantes registrados en la provincia de San José se clasifican de acuerdo con su posición en las categorías de ingreso bajo, medio o alto, en relación con si están a favor o no de la nueva reforma fiscal. Las frecuencias observadas se muestran en la tabla de contingencia siguiente:

Reforma fiscal	Nivel de ingreso			Total
	Bajo	Medio	Alto	
A favor	182	213	203	598
En contra	154	138	110	402
Total	336	351	313	1000

5.4 Una empresa brinda un servicio de entrega de paquetes en San José utilizando bicicletas como medio de

transporte. Existen quejas del peso de los paquetes por parte de los repartidores y por tanto desea realizar algunos cálculos de probabilidades, pero antes se debe verificar si el peso de los citados paquetes tiene una distribución normal. A continuación, se presenta la distribución de los pesos:

PESO DE LOS PAQUETES EN KILOS	CANTIDAD DE PAQUETES
De 0,55 a 1,55	5
De 1,55 a 2,55	8
De 2,55 a 3,55	15
De 3,55 a 4,55	15
De 4,55 a 5,55	16
De 5,55 a 6,55	16
De 6,55 a 7,55	10
De 7,55 a 8,55	8

Utilice chi cuadrada para verificar que la distribución de los pesos de los paquetes se ajusta a una distribución normal con un nivel de significancia del 5% (Promedio=4,74 desviación estándar=1,93)

5.5 Los siguientes datos representan el número de horas que operan dos diferentes tipos de calculadoras científicas de bolsillo antes de que necesiten recargarse.

Calculadora *A*	5,5	5,6	6,3	4,6	5,3	5,0	6,2	5,8	5,1
Calculadora *B*	3,8	4,8	4,3	4,2	4,0	4,9	4,5	5,2	4,5

Utilice la prueba respectiva, con $\alpha = 0.01$, para determinar si la calculadora *A* opera más tiempo que la calculadora *B* con una carga completa de la batería.

5.6 Se consultó sobre el nivel de conservación de recursos naturales para una pequeña muestra de países desarrollados y en vías de desarrollo, a saber:

Desarrollados: 2, 1, 1, 2, 3

En vías de desarrollo: 2, 2, 3, 4, 3, 1

Nivel de conservación de recursos naturales: 1=Bajo 2= Medio 3= Alto 4=Muy alto

Verifique que el nivel de conservación de recursos naturales es superior para los países en vías de desarrollo *(Use un nivel de confianza del 90% y concluya en términos del problema)*

5.7 Una universidad privada ofrece dos programas de maestría en administración de empresas. En el primer programa los estudiantes deben reunirse dos noches por semana en el campus principal en el centro de la ciudad. En el segundo programa, los estudiantes solo se comunican por internet con el profesor. El director académico de la Universidad está interesado en comparar las horas de estudio que estudiaron la semana pasada ambos grupos. Toma una muestra de 10 estudiantes que estudian en el campus y otra muestra de 12 estudiantes que estudian solo por internet.

Campus	28	16	42	29	31	22	50	42	23	25		
Internet	26	42	65	38	29	32	59	42	27	41	46	18

Por experiencia se sabe que distribuciones de tiempo de estudio en horas no se distribuyen normalmente. Usando un nivel de significancia del 5%, ¿es posible concluir que los estudiantes por internet estudian más horas?

5.8 Los siguientes datos representan los pesos, en kilogramos, del equipaje personal que llevan, en diferentes vuelos, un jugador de un equipo de beisbol y un jugador de un equipo de basquetbol.

Jugador de béisbol			**Jugador de basquetbol**	
16.3	20.0	18.6	15.4	16.3
18.1	15.0	15.4	17.7	18.1
15.9	18.6	15.6	18.6	16.8
14.1	14.5	18.3	12.7	14.1
17.7	19.1	17.4	15.0	13.6
16.3	13.6	14.8	15.9	16.3
13.2	17.2	16.5		

Utilice la prueba de la suma de rangos con α = 0.05 para probar la hipótesis nula de que los dos atletas llevan la misma cantidad de equipaje en promedio, en comparación con la hipótesis alternativa de que el peso promedio del equipaje de los dos atletas es diferente.

5.9 Se consultó sobre el nivel aceptación de los biocombustibles de tres sectores laborales obteniendo lo siguiente:

Educación: 2, 3, 3, 3, 3, 3, 2, 1, 2
Banca: 2, 2, 3, 3, 3, 2, 2
Agropecuario: 2, 1, 1, 2, 1, 1
Nivel de aceptación de biocombustibles: 1: Bajo 2: Medio 3: Alto

Con base en los datos anteriores, verifique que el nivel de aceptación de los biocombustibles es diferente para por lo menos dos sectores laborales con un nivel de significancia del 5% y concluya en términos del problema.

5.10 Asuma que se consultó sobre el apoyo del TLC a China para una pequeña muestra de costarricenses casados, solteros y divorciados, a saber:

Solteros: 2, 1, 1, 2, 3
Casados: 2, 2, 3, 4, 3, 1
Divorciados: 2, 3, 4, 4, 4, 3, 2
Nivel de apoyo al TLC: 1=Bajo 2= Medio 3= Alto 4=Muy alto

Verifique que el nivel de apoyo al TLC es diferente para al menos dos estados civiles (solteros, casados o divorciados) (Use una confianza del 90%, la prueba respectiva y concluya en términos del problema)

5.11 Se presume que el nivel de escolaridad de los habitantes del país tiene asociación con el grado de abstencionismo, por lo tanto, se recopiló la siguiente información de una muestra de siete personas:

Nivel de escolaridad: 1, 4, 3, 3, 4, 4, 2
Grado de abstencionismo: 1, 3, 1, 2, 3, 2, 1
Donde:
Nivel de escolaridad 1: Ninguno 2: Primaria 3: Secundaria 4: Universitaria
Grado de abstencionismo: 1: Bajo 2: Medio 3: Alto

Con base en los datos anteriores, calcule el coeficiente de correlación de Spearman y verifique la hipótesis para comprobar si existe asociación entre las variables estudiadas (con un nivel de significancia del 10%) y concluya en términos del problema

5.12 Se tiene la siguiente muestra de datos siete países latinoamericanos, a efectos de determinar si el grado de felicidad de los habitantes se encuentra asociado con el nivel de riqueza.

Grado de felicidad: 1, 3, 3, 2, 3, 4, 3
Nivel de riqueza: 1, 3, 1, 2, 2, 2, 1
Grado de felicidad: 1: Infeliz 2: Poco feliz 3: Feliz 4: Muy feliz
Nivel de riqueza: 1: Bajo 2: Medio 3: Alto

Calcule el coeficiente de correlación apropiado a la naturaleza de los datos anteriores y verifique la hipótesis para comprobar si existe asociación entre las variables estudiadas (con un nivel de significancia del 10%) y concluya en términos del problema.

5.13 Un grupo de consumidores prueba la calidad general de nueve marcas de hornos de microondas. Los rangos asignados por el grupo y los precios de venta al menudeo sugeridos son los siguientes:

Clasificación del grupo	**Precio**	**Fabricante sugerido ($)**
A	6	480
B	9	395
C	2	575
D	8	550
E	5	510
F	1	545
G	7	400
H	4	465
I	3	420

¿Existe una relación significativa entre la calidad y el precio de un horno de microondas? Utilice un nivel de significancia de 0.05?

5.14 Los siguientes datos representan el número de horas que funciona una desbrozadora antes de requerir una recarga:

1,5 2,2 0,9 1,3 2,0 1,6 1,8 1,5 2,0 1,2 1,7

A un nivel de significancia de 0.05 utilice la prueba de signo para probar la hipótesis de que esta desbrozadora específica funciona con una mediana de 1,8 horas antes de requerir una recarga.

5.15 Una empresa de taxis intenta decidir si utilizar neumáticos radiales en vez de neumáticos regulares con cinturón le serviría para ahorrar combustible. Se equipan 16 automóviles con neumáticos radiales y se conducen por un recorrido de prueba establecido. Después se equipan los mismos automóviles con los neumáticos regulares con cinturón y se hace que los mismos conductores vuelvan a realizar el recorrido de prueba. El consumo de gasolina, en kilómetros por litro, se presenta en la tabla que se presenta a continuación. ¿Con un nivel de significancia de 0,05 se puede concluir que los automóviles equipados con neumáticos radiales ahorran más combustible que los equipados con neumáticos regulares con cinturón?

Automóvil	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Neumáticos radiales	4,2	4,7	6,6	7,0	6,7	4,5	5,7	6,0	7,4	4,9	6,1	5,2	5,7	6,9	6,8	4,9
Neumáticos con cinturón	4,1	4,9	6,2	6,9	6,8	4,4	5,7	5,8	6,9	4,9	6,0	4,9	5,3	6,5	7,1	4,8

5.16 En un EBAIS se les indica a sus pacientes que la mediana del tiempo de espera de consulta es 30 minutos. Se desea verificar que dicho tiempo espera es superior a la mediana y para ello se toma una muestra de 204 pacientes, de los cuales 120 esperan más de 30 minutos, 4 esperan 30 minutos y 80 esperan menos de 30 minutos. Con un nivel de significancia del 5% verifique la hipótesis de que los pacientes del EBAIS los atienden en más de 30 minutos. Realice una Prueba de Signos.

5.17 Un inspector de tránsito se ubica en el Parque Central de San José para realizarle boletas a los conductores que circulan con automóviles con restricción. Dentro de sus anotaciones registra los conductores que viven en la capital (C) y fuera de ella (F), obteniendo los siguientes resultados:

CCFCCCCFFCCCCFCFFCCCCFCFFCCCCCFCCCFCFCCCCFFCFFFFCCCCFCCF

Utilizando un nivel de significancia del 5% ¿se puede afirmar que durante el periodo que estuvo el inspector de tránsito en el Parque Central realizado boletas, los conductores de los automóviles que viven en la capital y fuera de ella pasaron por ese lugar en forma aleatoria? Aplique la Prueba de Rachas de aleatoriedad.

5.18 Con el fin de estimar la proporción de votantes que favorecen a cierto candidato para alcalde, se selecciona una muestra aleatoria de 15 adultos que viven en una pequeña ciudad. También se le pregunta a cada individuo si se graduó de la universidad. Al denotar con *S* y *N* las respuestas “si” y “no”, respectivamente, a la pregunta sobre la escolaridad, se obtuvo la siguiente secuencia:

N N N N S S N S S N S N N N N

Utilice la prueba de rachas a un nivel de significancia de 0,1 para determinar si la secuencia apoya la afirmación de que la muestra se seleccionó al azar.