



Universidad de Costa Rica  
Facultad de Ciencias Económicas  
Escuela de Estadística



# **EJERCICIOS DE LA CATEDRA DE ESTADISTICA GENERAL 2**

**(XS 277)**

[2 ciclo lectivo 2019]

## PRESENTACION

El presente “Manual de Prácticas” contiene una serie de ejercicios sobre los temas que se imparten en los cursos de Estadística General 2, en la Facultad de Ciencias Económicas y en las diferentes Sedes Regionales de la Universidad de Costa Rica.

El documento se elaboró con base en los ejercicios aportados por los profesores de la cátedra y se pretende que éste material sirva de base para las lecciones prácticas y los laboratorios de cómputo a desarrollar durante el ciclo lectivo; así como ejemplo de la aplicación de los conceptos estadísticos desarrollados.

Mucho le agradeceríamos cualquier observación o sugerencia, con el propósito de mejorar su contenido.

Su servidor,

Ramón Luis Bolaños Zamora  
Coordinador  
Cátedra Estadística General

## INDICE

<b>CONTENIDO</b>	<b>PAGINA</b>
1. Estimación.....	4
2. Verificación de hipótesis.....	8
3. Análisis de Variancia.....	11
4. Correlación y regresión.....	16
5. Estadística No Paramétrica.....	21

---

# Ejercicios de Estadística General 2

---

## 1. Estimación

- 1.1 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 \$200 del promedio real con 99% de confianza? Suponga que la desviación estándar de los depósitos mensuales totales para todos los clientes es alrededor de mil dólares.
- 1.2 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 obstante, no desea promediar los precios diarios de la población de 2 500 precios, ya que resulta muy laborioso conseguirlos, de modo que 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, a su vez, se estima que la desviación estándar del precio de esta acción es de unos \$12,50 para este período. ¿Que tan grande debe ser la muestra a tomar?
- 1.3 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. Finalmente según los datos del INEC a julio del 2008 la población de hogares de la zona rural llegó a 432 629.
- 1.4 En una empresa con un número elevado de empleados se desea estimar el promedio de horas trabajadas ( $\mu$ ) 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?
- 1.5 Los siguientes datos se refieren a la cantidad de transacciones que se realizan 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%

## Ejercicios de Estadística General 2

---

- 1.6 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.
- 1.7 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.
- 1.8 El Ministerio de Salud desea estimar un intervalo de confianza para la diferencia de promedios de las horas de sueño de los habitantes de Jicaral y Nosara con un 95% de confianza. Para ello toma una muestra de aleatoria de 200 habitantes de Jicaral 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 Nosara obteniendo un promedio de 7,2 horas con una desviación estándar de 3,8 horas. Determine el intervalo citado.
- 1.9 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.
- 1.10 El Consejo Nacional de Política Pública de la Persona Joven y el Vice ministerio de la Juventud realizaron una encuesta a los residentes en Costa Rica (costarricenses y extranjeros) con edades entre 15-35 años en abril del 2007, aplicada una a muestra de 2500 jóvenes 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 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 su vez, 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 cambiaría 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

## Ejercicios de Estadística General 2

- 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.

1.11 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.

1.12 A continuación se presentan algunos resultados de la Encuesta de Hogares de Propósitos Múltiples realizada por el Instituto Nacional de Estadística y Censos en julio del 2008, respecto la clasificación de los costarricenses ocupados según el sector institucional donde laboran.

Sector Institucional	Fuerza de trabajo ocupada	Ingreso mensual promedio (en miles de colones)	Desviación estándar del ingreso * (en 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 tabla anterior:

- a) Determine el promedio total del ingreso mensual

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

- b) Afijación proporcional.  
c) Afijación óptima (Neyman)

1.13 Explique en qué consiste el teorema del límite central.

1.14 El promedio es un estimador suficiente? Justifique

1.15 Bajo cuáles circunstancias se prefiere la afijación óptima en lugar de la proporcional? Justifique

## Ejercicios de Estadística General 2

1.16 A continuación se presentan las películas más taquilleras del año 1975 al 2004:

PELICULA	AÑO	TAQUILLA (millones de \$)	PELICULA	AÑO	TAQUILLA (millones de \$)
Shrek 2	2004	\$912.00	Star Wars: Episode I -- The Phantom Menace	1999	\$925.50
Harry Potter and the Prisoner of Azkaban	2004	\$789.80	The Sixth Sense	1999	\$661.50
Spider-Man 2	2004	\$784.00	Toy Story 2	1999	\$485.70
The Incredibles	2004	\$631.20	Saving Private Ryan	1998	\$479.30
The Passion of the Christ	2004	\$611.80	Titanic	1997	\$1,835.00
Meet the Fockers	2004	\$511.90	The Lost World	1997	\$614.40
LOTR: The Return of the King	2003	\$1,129.20	Men In Black	1997	\$587.20
Finding Nemo	2003	\$853.20	Independence Day	1996	\$813.10
The Matrix: Reloaded	2003	\$735.70	Twister	1996	\$495.00
Pirates of the Caribbean	2003	\$653.20	The Lion King	1994	\$771.90
Bruce Almighty	2003	\$459.00	Forrest Gump	1994	\$680.00
The Lord of the Rings: The Two Towers	2002	\$920.50	Jurassic Park	1993	\$920.00
Harry Potter and the Chamber of Secrets	2002	\$866.40	Mrs. Doubtfire	1993	\$423.20
Spider-Man	2002	\$806.70	Aladdin	1992	\$502.40
Star Wars: Episode II -- Attack of the Clones	2002	\$648.30	Home Alone	1990	\$533.80
Signs	2002	\$408.00	Ghost	1990	\$517.60
My Big Fat Greek Wedding	2002	\$356.50	Batman	1989	\$413.00
Harry Potter and the Sorcerer's Stone	2001	\$968.70	Beverly Hills Cop	1984	\$316.40
The Lord of the Rings: The Fellowship of the Ring	2001	\$860.70	Ghostbusters	1984	\$291.60
Monsters, Inc.	2001	\$524.20	Return of the Jedi	1983	\$573.00
Shrek	2001	\$469.70	E.T.	1982	\$757.00
Rush Hour 2	2001	\$329.10	Raiders of the Lost Ark	1981	\$384.00
Mission: Impossible 2	2000	\$545.40	The Empire Strikes Back	1980	\$533.90
Cast Away	2000	\$424.30	Star Wars	1977	\$797.90
Dr. Seuss' How the Grinch Stole Christmas	2000	\$340.00	Jaws	1975	\$471.00

- a) Seleccione una muestra aleatoria de 10 películas
- b) Seleccione una muestra sistemática de 10 películas
- c) Ordene la población de películas por taquilla
- d) Seleccione una muestra sistemática de 10 películas de c) y comente el resultado
- e) Divida la población de películas en dos estratos (según monto de la taquilla) y seleccione una muestra de 10 películas distribuidas en forma proporcional

### 2. Verificación de Hipótesis

- 2.1 Un estudio llevado a cabo por una empresa mostró que Paris es el lugar más costoso para vivir de las 12 ciudades de la Unión Europea. Paris está clasificado segundo en costo de vivienda, con una unidad de renta con un costo promedio de \$4 292 al mes. Suponga que el director general en una compañía piensa que esta cantidad es demasiado 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 \$4 008, 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 la hipótesis nula se rechaza, 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 utilizando 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, encontrándose un tiempo total de espera de 5500 minutos con una desviación estándar de 65 minutos.
- Establezca la hipótesis a verificar (simbólicamente y en enunciado)
  - Qué decidiría usted si se usa  $\alpha = 0,05$ ? Realice todos los cálculos y concluya en términos del problema.
  - ¿Cuál error se puede cometer al tomar la decisión de b) y por qué?
  - Si en este caso  $\mu$  fuera realmente igual a 45 minutos. ¿Cuál es la magnitud del error tipo II ( $\beta$ ) ?
- 2.4 Asuma que con base en datos de estudios el 60% de los jóvenes de la zona rural de Costa Rica padecía de estrés. Dado el mundo actual tan acelerado se presume que dicho porcentaje ha aumentado. Según la muestra del Resto del país Rural de la encuesta citada en el ejercicio 1.9, dicho porcentaje llegó al 62%
- 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 rural se incrementó
  - Determine la magnitud del error tipo 2, si el verdadero porcentaje de estrés fue de 68%
- 2.5 Dentro de los aspectos estudiados en la encuesta del ejercicio 1.9 se encuentra la edad de inicio de los jóvenes de la primera relación sexual, de manera que 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)

## Ejercicios de Estadística General 2

---

- 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. A continuación se presentan los resultados obtenidos:

PROFESIONALES	PROPORCION*	TAMAÑO DE MUESTRA
Empresarios	0,35	60
Economistas	0,20	65

\*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 Una corporación norteamericana tiene una tienda en Multiplaza de Este y otra en Multiplaza de Oeste y desea comparar los promedios de los montos comprados por los clientes. Por ello selecciona una muestra de 10 clientes de la primera tienda (Este) obteniendo los siguientes montos de compras (en dólares):

17,58 19,73 12,61 17,79 16,22 15,82 15,40 15,86 11,82 15,85

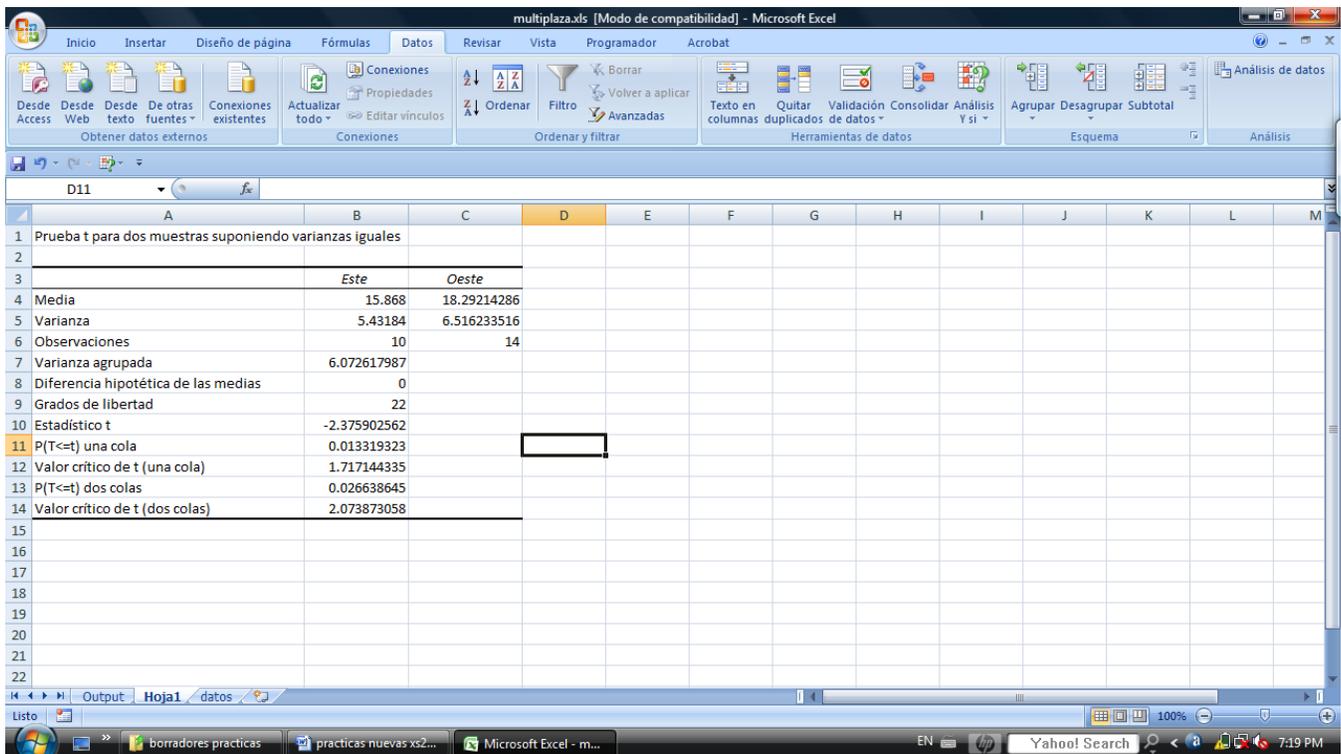
A su vez se tomó una muestra de 14 clientes de la segunda tienda (Oeste) obteniendo lo siguiente:

18,19 20,22 17,38 17,96 23,92 15,87 16,47 15,96 16,79 16,74 21,4 20,57 19,79 14,83

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%

## Ejercicios de Estadística General 2

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



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Prueba t para dos muestras suponiendo varianzas iguales												
2													
3		Este	Oeste										
4	Media	15.868	18.29214286										
5	Varianza	5.43184	6.516233516										
6	Observaciones	10	14										
7	Varianza agrupada	6.072617987											
8	Diferencia hipotética de las medias	0											
9	Grados de libertad	22											
10	Estadístico t	-2.375902562											
11	P(T<=t) una cola	0.013319323											
12	Valor crítico de t (una cola)	1.717144335											
13	P(T<=t) dos colas	0.026638645											
14	Valor crítico de t (dos colas)	2.073873058											
15													
16													
17													
18													
19													
20													
21													
22													

Compare los resultados con lo obtenido en el ejercicio anterior.

2.11 El error tipo 1 es más grave que el error tipo 2? Justifique

2.12 Bajo cuáles circunstancias se podría cometer el error tipo 2.

## Ejercicios de Estadística General 2

### 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	Continente B	Continente C
2	9	3
3	8	2
5	7	2
2	7	
1		
2		
Promedio = 2,5	Promedio = 7,75	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 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 Angeles	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

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

## Ejercicios de Estadística General 2

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

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

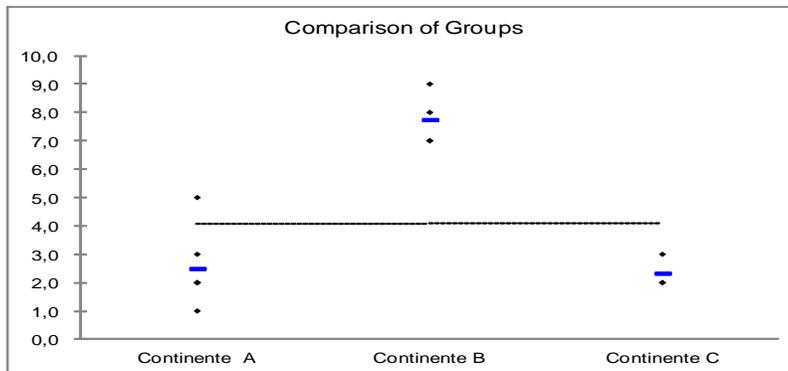
	Continente C	Continente A	Continente B
Continente C	2,3		
Continente A	2,5	,8399	
Continente B	7,8	,0001	3,08E-05

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

	Continente C	Continente A	Continente B
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

- 3.4 Una empresa vende champú para cabello normal, graso 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

## Ejercicios de Estadística General 2

3.5 Los datos del ejercicio 3.2 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:

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			
<b>Total</b>	<b>677400</b>	<b>19</b>				

Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 3.2

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

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

## Ejercicios de Estadística General 2

The screenshot shows an Excel spreadsheet with the following data:

RESUMEN					
	Cuenta	Suma	Promedio	Varianza	
A1	3	127	42.33333333	4.33333333	
A2	3	166	55.33333333	10.33333333	
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 3.6

3.8 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

## Ejercicios de Estadística General 2

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		
<b>A1</b>						
Cuenta	3	3	3	9		
Suma	121	131	128	380		
Promedio	40,33333333	43,66666667	42,66666667	42,22222222		
Varianza	0,333333333	0,333333333	0,333333333	2,444444444		
<b>A2</b>						
Cuenta	3	3	3	9		
Suma	158	161	176	495		
Promedio	52,66666667	53,66666667	58,66666667	55		
Varianza	0,333333333	0,333333333	0,333333333	8		
<b>A3</b>						
Cuenta	3	3	3	9		
Suma	143	113	139	395		
Promedio	47,66666667	37,66666667	46,33333333	43,88888889		
Varianza	0,333333333	0,333333333	0,333333333	22,36111111		
<b>A4</b>						
Cuenta	3	3	3	9		
Suma	143	182	140	465		
Promedio	47,66666667	60,66666667	46,66666667	51,66666667		
Varianza	0,333333333	0,333333333	0,333333333	46		
<b>Total</b>						
Cuenta	12	12	12			
Suma	565	587	583			
Promedio	47,08333333	48,91666667	48,58333333			
Varianza	21,3560606	86,08333333	39,9015152			
<b>ANÁLISIS DE VARIANZA</b>						
<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
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	0,333333333			
Total	1643,63889	35				

Interprete los resultados

## Ejercicios de Estadística General 2

### 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:

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

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

4.2 Un analista toma una muestra aleatoria de 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:

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

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

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

## Ejercicios de Estadística General 2

The screenshot shows an Excel spreadsheet with the following data:

ANÁLISIS DE VARIANZA					
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	328.1901042	328.1901042	47.23752764	4.33506E-05
Residuos	10	69.4765625	6.94765625		
Total	11	397.6666667			

	Coefficientes	Error típico	Estadístico t	Probabilidad
Intercepción	35.46484375	1.723876853	20.57272461	1.62814E-09
experiencia	-1.38671875	0.201764524	-6.872956252	4.33506E-05

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

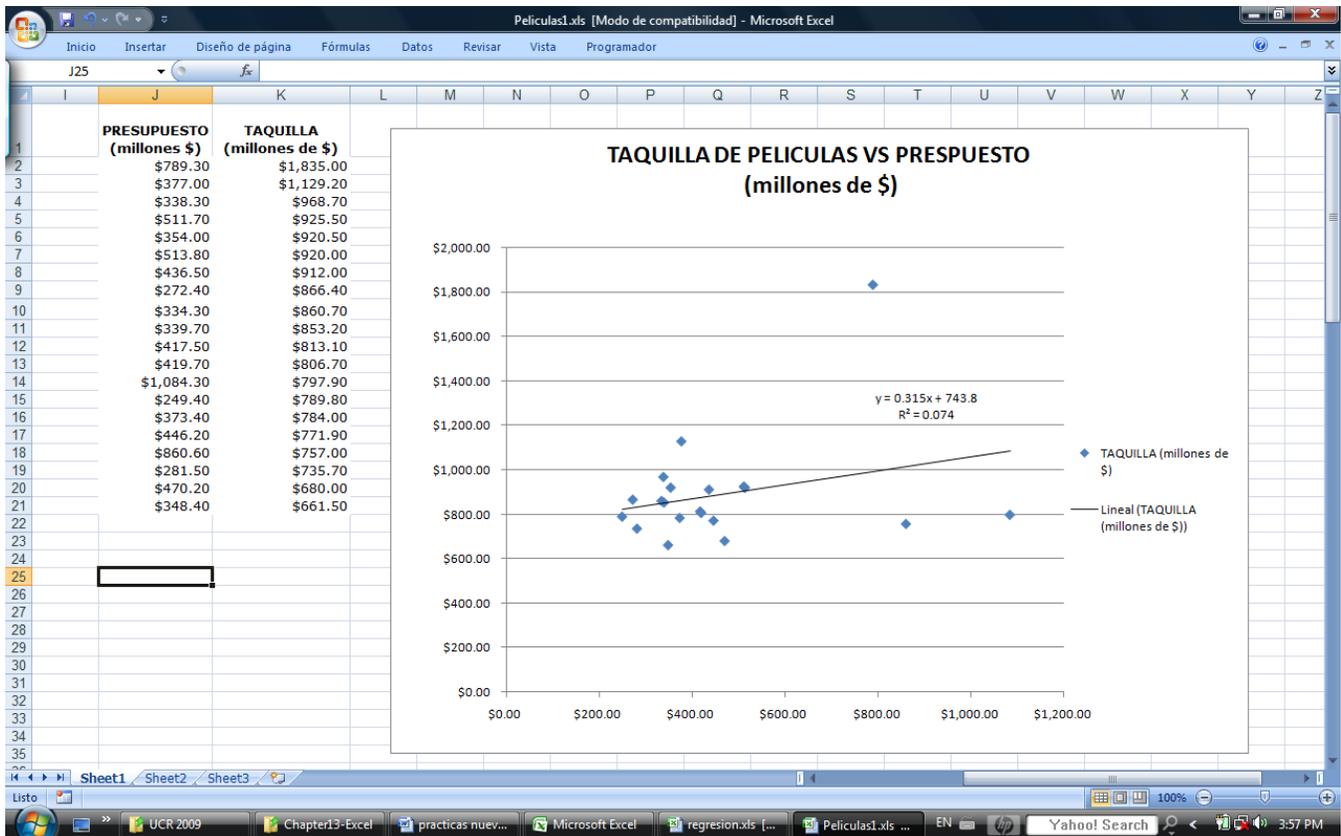
4.4 A continuación se presentan las 20 películas con mayores taquillas del año 1982 al 2004:

PELICULA	AÑO	TAQUILLA (millones de \$)	PRESUPUESTO (millones \$)
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

## Ejercicios de Estadística General 2

- Elabore un diagrama de dispersión, bajo el entendido que el presupuesto de la película determina la taquilla
- Calcule el coeficiente de correlación y la bondad de ajuste para confirmar lo encontrado en a)
- 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

4.5 Los datos del ejercicio 4.4 se procesaron mediante el paquete de cómputo Excel y se generaron los siguientes resultados:



Compare los resultados anteriores con los encontrados mediante el desarrollo del ejercicio 4.4

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:

$$y = 1497,7 + 9,8x \quad \text{donde } x = \text{antigüedad de la pintura en años}$$

$$y = \text{precio de subasta en miles } \text{€}$$

Correlación (antigüedad vs precio de subasta) = 0,58

- Estime el precio de subasta de una pintura con 100 años

## Ejercicios de Estadística General 2

---

b) Se piensa que la cantidad de postores en la subasta influye en el precio, por tanto el coleccionista calcula los siguientes coeficientes de correlación:

Correlación (antigüedad vs postores) = - 0,6

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

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

$$y = \text{monto de la propina en } \text{¢}$$

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  $S_b$

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

\*en colones

\*\*en años

## Ejercicios de Estadística General 2

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

The screenshot shows an Excel spreadsheet with the following data:

Estadísticas de la regresión					
Coefficiente de correlación múltiple	0.799336884				
Coefficiente de determinación R <sup>2</sup>	0.638939454				
R <sup>2</sup> ajustado	0.566727345				
Error típico	32757.7558				
Observaciones	25				

ANÁLISIS DE VARIANZA					
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	4	37978512397	9494628099	8.848092946	0.00027823
Residuos	20	21461411307	1073070565		
Total	24	59439923704			

	Coefficientes	Error típico	Estadístico t	Probabilidad
Intercepción	1334149.66	165456.5452	8.063444443	1.03152E-07
Hipoteca	0.002637688	0.000474398	5.560072029	1.92406E-05
Educa	30400.62882	13368.06573	2.27412323	0.034125817
Edad	-1979.627455	1726.75996	-1.146440444	0.265144598
Valor	-9.50419E-05	0.000121304	-0.783502587	0.442512287

- 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

## Ejercicios de Estadística General 2

### 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	ACEPTACION 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
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 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 tienen 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 la 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)

## Ejercicios de Estadística General 2

---

5.4 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 una confianza del 90% y concluya en términos del problema)

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

Educación: 2, 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.

5.6 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)

5.7 Se presume que el nivel de escolaridad de los habitantes del país tiene asociación con el grado de abstencionismo, por 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

## Ejercicios de Estadística General 2

---

5.8 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.

5.9 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.

5.10 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:

CCFCCCCFFCCCCFCFFCCCCFCFFCCCCFCFFCCCCFCFFCCCCFCFFCCCCFCFF

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.