

**ESTADISTICA**  
**PARA**  
**ADMINISTRADORES**

**GUIA DE TRABAJOS**  
**PRACTICOS**

1<sup>ra</sup> edición

**AUTOR: CESAR A. FERNANDEZ MAGAN**

## Sobre el autor

### *César Augusto Fernández Magán*

- *Contador Público. Universidad de Buenos Aires*
- *Magíster en Dirección de Finanzas y Control de la Universidad Argentina de la Empresa (UADE Business School)*
- *Posgrado en docencia universitaria FCE-UBA*
- *Profesor Adjunto interino de Estadística para Administradores de la Facultad de Ciencias Económicas de la UBA*
- *Profesor Adjunto Ordinario de la Universidad Argentina de la Empresa*
- *Ex Docente contratado en Estadística Aplicada. Ingeniería Industrial (ITBA)*
- *Ex Profesor Adjunto de Estadística Aplicada I y II de la UCES*
- *Ex ayudante de segunda interino ad-honorem de Álgebra y de Análisis Matemático II de la Facultad de Cs. Económicas de la UBA*
- *Co-autor de la Guía de Estudio para el Modelo Lineal de Regresión para la cátedra de Estadística para Administradores de la Facultad de Cs. Económicas de la UBA*
- *Co-autor del libro Estadística no paramétrica. Ediciones Cooperativas*
- *Autor de varias guías prácticas sobre estadística para diversas universidades*
- *Consultor de empresas en temas de estadística y técnicas cuantitativas*

# **PRÁCTICO N° 1**

## **INFERENCIA**

**EN**

**POBLACIONES**

**NORMALES**



Temas:

- **Intervalos de confianza y pruebas de hipótesis para la varianza poblacional**
- **Intervalos de confianza y pruebas de hipótesis para la media poblacional con desvío poblacional desconocido**



**Atención:** En los ejercicios del presente capítulo, se utilizó como estimador de la varianza poblacional:

$$S^2 = \frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2$$

---

**Intervalos de confianza para la varianza de una población**

**Problema resuelto**

Una empresa productora de caños para pasar los cables de energía eléctrica desea analizar el diámetro de los mismos y como primera medida seleccionó una muestra aleatoria que arrojó los siguientes valores de sus diámetros en cm.

$$2,5 - 2,7 - 3 - 2 - 2,5 - 3,2 - 1,8 - 1,9 - 2,3 - 2,4 - 2,7 - 3$$

Suponiendo que el diámetro en la población tiene una distribución normal:

Construya un intervalo de confianza para el desvío poblacional con un 90% de confianza. Interprete el resultado obtenido.

$$\chi_{11;0,05}^2 = 4.575 \quad \chi_{11;0,95}^2 = 19.675$$
$$\left( \sqrt{\frac{11 * 0.4492^2}{19.675}} ; \sqrt{\frac{11 * 0.4492^2}{4.575}} \right)$$

**El desvío del diámetro de los cables estará entre 0,336 cm y 0,697 cm con una confianza del 90 %**

**Problemas propuestos**

1. Una muestra arrojó los siguientes valores en milisegundos de tiempo de acceso a un disco rígido para la misma instrucción. La variable en la población sigue aproximadamente una distribución normal.

3,4 – 5,6 – 4,8 – 3,7 – 6,3 – 4,7 – 3,4 – 5,8 – 4,5

- a) Construya un intervalo de confianza del 95% para estimar el desvío del tiempo de acceso al disco rígido.
  - b) Determine con un 95% de confianza la estimación del desvío máximo en el tiempo de acceso al disco rígido.
2. En una muestra aleatoria de 10 pilas tamaño AA, se obtuvo un desvío estándar de la vida útil de las mismas de 12,6 horas. Estimar al 95% el desvío de la vida útil de todas las pilas considerando que la variable sigue una distribución normal.
  3. Se tiene una máquina que llena botellas con un promedio de 500 cm<sup>3</sup> por botella y un desvío desconocido. Se desea estimarlo. Para ello se toma una muestra de 26 botellas las cuales arrojan un contenido promedio de 490 cm<sup>3</sup> cada una y un desvío de 20 cm<sup>3</sup> utilice un nivel de significación del 5%
  4. Un corredor entrena diariamente corriendo 200 mts 40 veces. Se sabe que el tiempo que tarda en cada carrera tiene distribución normal con desvío conocido equivalente a 7 décimas de segundo (0.7 segundos) ayer registró un tiempo promedio de 20.75 segundos con un desvío de 1.1 segundos. Indique al 5 % si cree que el desvío histórico se ha modificado.

### **Pruebas de hipótesis para la varianza de una población**

#### **Problema resuelto**

Roberto Piras, especialista en selección de personal de una empresa de servicios de gran tamaño sabe por análisis anteriores que el desvío estándar de la variable edad de postulantes a cargos gerenciales altos es de 12 años. Sin embargo y dado lo cambiante del mercado local sospecha que pudo haber aumentado. Para verificar esto, considerando un nivel de significación del 5%, selecciona una muestra aleatoria de los datos de las edades de 25 postulantes a estos cargos elegidos al azar que se han presentado en el lapso de los últimos 6 meses, la cual arroja un desvío de 15 años.

#### **Solución**

$$H_0: \sigma = 12$$

$$H_1: \sigma > 12$$

C.R. : Si  $\chi^2_{\text{calc}} > \chi^2_{(24;0,95)}$

$$\chi^2_{\text{calc}} = \frac{(n-1) * s^2}{\sigma^2} = \frac{24 * 15^2}{144} = 37.5$$

El valor crítico es  $\chi^2_{24;0,95} = 36.415$

**Decisión:** Como  $37.5 > 36.415$  se rechaza  $H_0$  y se verifica que el desvío de las edades ha aumentado, con un 5% de probabilidad máxima de estar equivocados en tal afirmación.

### Problemas propuestos

5. Una propietaria de una cadena de peluquerías para damas, está analizando la factibilidad de otorgar un seguro médico a sus empleados. Su investigación está dirigida a estimar el costo promedio y además comprobar que el desvío estándar no exceda los \$280. Una muestra aleatoria de ocho empleados proporciona los siguientes gastos en atención odontológica, en pesos, en el último año:

625    425    625    975    141    52    342    459

- a) Usando un nivel de significación del 5%, ¿existe evidencia suficiente de que el desvío estándar poblacional de los gastos de seguro médico sea mayor que \$280? Interprete el resultado.

6. Los instrumentos científicos de medición deben proporcionar lecturas correctas y con errores de medición muy pequeños. El gerente de producción de la empresa Instrumén SA está preocupado por el índice de variación en las lecturas producidas por determinado instrumento, el que está diseñado para tener una desviación estándar de 200 metros, teniendo en cuenta que la variable se distribuye aproximadamente como una normal. Selecciona una muestra aleatoria de 7 mediciones con el mismo instrumento y obtiene una desviación estándar de 250 metros. Pista : piense que estamos hablando de mediciones de un instrumento de precisión. ¿sería lógico hacer un test bilateral?

¿A que conclusión arriba trabajando con un nivel de significación del 5%?

7. La fabricación de un repuesto para un reproductor de blu-ray se necesita realizar con tolerancias muy estrechas para que sea aceptado en el mercado. Las especificaciones del repuesto establecen que la varianza de las longitudes debe alcanzar como máximo  $0.0005\text{mm}^2$ . Suponga que con una muestra aleatoria de 25 repuestos la varianza de la muestra fue de  $0.0006\text{mm}^2$ . Pruebe con un nivel de significación del 5% si se ha violado la especificación de la varianza aceptable. Indique qué supuestos consideró.

8. Un fabricante de baterías para kartings asegura que sus baterías duran en promedio 3 años con un desvío estándar de **9 meses**. Si una muestra de 6 de estas baterías arrojó las siguientes duraciones **en años**

1.8    2.5    3    3.6    4.1    4.2

- a) Determine un intervalo de confianza del 95% para el desvío de la duración de estas baterías  
b) Indique si es válida la afirmación del fabricante al 5%

9. Una empresa embotelladora afirma que el desvío del contenido de las botellas es mayor a  $20 \text{ cm}^3$  y que de comprobarse esto detendría la producción. Para ello toma una muestra de 26 botellas observando un desvío estándar de  $20 \text{ cm}^3$ . Dados los números involucrados, antes de resolver la prueba piense cuál será la conclusión, luego realice la prueba con un 5% de riesgo.
10. Una empresa envasa latas de atún y asegura que el contenido promedio de las mismas es de por lo menos 60 gr con un desvío de 3 gr. Si esto no fuera así detendría de inmediato la producción. Se lleva a cabo un control de calidad y se abre el contenido de 15 latas que arrojan un promedio de 57 gr con un desvío de 4 gr. Verifique si la producción debe continuar o no. Se desea que el riesgo de detener la producción cuando en realidad la máquina está funcionando bien sea del 5% en lo que al desvío se refiere.
11. Una empresa fabrica tuercas las cuales deben tener exactamente un diámetro promedio de 2 mm con un desvío de 0.08mm. Se llevó a cabo un control de 25 tuercas que arrojó un promedio de 1,9 mm de diámetro con un desvío de 0,1 mm. Establezca a un riesgo del 1% si las tuercas están dentro de las especificaciones técnicas en lo que al desvío se refiere.

### **Intervalos de confianza para la media poblacional con desvío poblacional desconocido**

#### **Problema resuelto**

Una máquina llena de latas de tomate y dicho contenido tiene una distribución normal. Se desea estimar el contenido promedio de todas las latas del lote. Para ello se toma una muestra de 16 latas que arrojan un contenido promedio de 235 gr. y un desvío de 15 gr.

- a) Realice la estimación con un 95% de confianza.  
b) ¿Cuántas latas más habría que observar para reducir el error muestral de la estimación anterior en un 30%?

#### **Resolución:**

Datos del ejercicio:

$$\begin{aligned}n &= 16 \\ \bar{x} &= 235 \text{ gr.} \\ S &= 15 \text{ gr.} \\ \text{NC} = 95\% &\rightarrow \alpha = 0.05 \rightarrow 1 - \alpha/2 \rightarrow 1 - 0.05/2 = 0.975 \rightarrow \text{en } t_{(n-1; 1-\alpha/2)} = t_{15; 0.975} = 2.131\end{aligned}$$

#### **Se pide:**

- a) Estimar el contenido promedio de las latas del lote

$$\begin{aligned}\text{IC}(\mu) &: [\bar{x} \pm t_{(n-1; 1-\alpha/2)} * S/\sqrt{n}] \\ \text{IC}(\mu) &: [235 \pm t_{(15; 0.975)} * 15/\sqrt{16}] \\ \text{IC}(\mu) &: [235 \pm 2.131 * 15/\sqrt{16}]\end{aligned}$$



IC( $\mu$ ) : [235 +/- 8]

IC( $\mu$ ) : [227; 243] gramos

### Interpretación:

Con un 95% de confianza puede afirmarse que el contenido promedio de las latas se encuentra entre 227 y 243 gramos.

b) ¿Cuántas latas más se deben agregar a la muestra para que el error muestral obtenido se reduzca en un 30%?

El EM (Error muestral) de la estimación calculado previamente fue de:

$$t_{(15; 0.975)} * 15 / \sqrt{16} = 8 \text{ gramos}$$

Si deseamos reducirlo en un 30% haremos:

$$EM * (1 - 0.3) = EM * 0.7 = 8 * 0.7 = \boxed{5.6} \text{ EM deseado (EM*)}$$

Se nos pide que averigüemos el n° de latas más que debo observar →

Si  $EM = [t_{(n-1; 1-\alpha/2)} * S / \sqrt{n}]$  y quiero averiguar n, entonces debemos despejarla de la ecuación:

$$n = \left[ \frac{t_{n-1; 1-\alpha/2} * S}{EM * } \right]^2$$

La dificultad que este cálculo presenta es que la n que queremos despejar está dentro de la fórmula que vamos a usar para calcularlo. Es por eso que vamos a utilizar el **método iterativo** para calcularlo.

Este método consiste en probar distintos valores para el n a obtener y utilizarlos en nuestra  $t_{(n-1; 1-\alpha/2)}$

Luego de hacer varios cálculos si la n utilizada para la fórmula es igual a la que nos da como resultado la cuenta, entonces será ese el tamaño de muestra requerido.

Comenzaremos a probar desde el valor de n dado como dato para facilitar nuestro cálculo, aunque podría empezarse con otro valor cualquiera que fuera mayor a ese.

Veamos como averiguamos el tamaño de muestra deseado en este ejercicio:

n dado en el ejercicio: 16

1° intento...**si n=16**

$$n = [(t_{(15; 1-0.05/2)} * 15) / 5.6]^2$$

$$n = [(2.131 * 15) / 5.6]^2$$

$n = 32.58 \rightarrow 33$  (al redondear, si redondeo va a ser siempre hacia arriba pues quiero achicar el EM)

¿Es igual al  $n$  que utilizamos para el cálculo? No  $16 \neq 33$ , por lo tanto, vamos a tener que seguir intentando. ¿Qué  $n$  utilizamos ahora? La  $n$  obtenida en el paso anterior.

2° intento

**Si  $n=33$**

$$n = [(t_{(32; 1-0.05/2)} * 15) / 5.6]^2$$

$$n = [(2.037 * 15) / 5.6]^2$$

$n = 30 \rightarrow 30 \neq 33 \Rightarrow$  debo continuar...

3° intento

Si  $n= 30$

$$n = [(t_{(29; 1-0.05/2)} * 15) / 5.6]^2$$

$$n = [(2.045 * 15) / 5.6]^2$$

$n = 30.0049 \approx 30 \rightarrow 30 = 30 \gggg$  este es el tamaño de muestra que reduce el EM anterior en un 30%.

El ejercicio nos pide saber cuántas latas MAS debo agregar a mi muestra. Ya tenía 16, por lo tanto tengo que hacer  $30 - 16 = 14$  para obtener ese valor.

**Conclusión:** Deberé agregar 14 latas más para reducir el error muestral original en un 30%

### **Pruebas de hipótesis para la media poblacional con desvío poblacional desconocido**

#### **Problema resuelto**

Una empresa vende cajas de fósforos.

En cada caja pone 400 fósforos y cuando el cliente recibe estas cajas hace un control de calidad al azar de 16 de las mismas para ver si realmente tienen el contenido prometido por la empresa. El cliente establece en un 5% la probabilidad de rechazar todo el envío en forma equivocada.

- 1) Plantee las hipótesis adecuadas, la condición de rechazo y la regla de decisión.
- 2) Sabiendo que se tomó una muestra que arrojó un promedio de 388 fósforos por caja con un desvío de 10, decida si hay que rechazar todo el envío.

#### **Resolución:**

$$H_0 = \mu \geq 400$$

$$H_1 = \mu < 400$$

$$CR: Si \bar{X} < \bar{X}_c$$

RD: si Rechazo  $H_0 \rightarrow$  Devuelvo el envío

n = 16 cajas

$\bar{X}$  = 388 fósforos

S = 10 fósforos

**Calculo  $\bar{X}$  crítico**

$$\bar{X}_c = \mu_0 \pm t_{v;\phi} \frac{S}{\sqrt{n}} = [400 - t_{15; 0.95} * 10 / \sqrt{16}] = 400 - 1.753 * 10 / \sqrt{16} = 395.6 \text{ fósforos}$$

Como  $\bar{X}$  = 388 es menor a  $\bar{X}_c$  = 395.6  $\rightarrow$  Rechazo  $H_0$  y por tanto rechazo todo el envío.

### **Intervalos de confianza y pruebas de hipótesis para la media poblacional con desvío poblacional desconocido**

#### **Ejercicios propuestos**

1. Un restó analiza la cantidad de litros de aceite de oliva que consume por semana. Se han tomado datos del consumo de las últimas 8 semanas: 7-8-6-7-7-8-7-8.
  - a) Estime con una seguridad del 95% el consumo promedio semanal de aceite de oliva.
  - b) Si se desea reducir el error muestral de la estimación anterior en un 20%, manteniendo el mismo tamaño de muestra, ¿Qué nivel de confianza se deberá utilizar aproximadamente?
2. Un minimercado está intentando estimar sus ventas promedio semanales. A tal fin toma una muestra de 10 semanas y anota el monto de sus ventas en miles de \$  
15—18—14—16—25—20—12—15—17—21
  - a) Estime las ventas promedio semanales con un 90% de confianza
  - b) ¿Cuántas semanas más debería observar para reducir el error muestral obtenido en la estimación anterior en un 40%?
3. En un criadero de patos la cantidad de alimento que se consume por día tiene distribución normal. Se tomó una muestra de 7 días que arrojó un consumo medio de 4 kg con desvío de 3/4 kg.

- a) Estime al 95 % el consumo medio diario  
b) Si se desea mantener el mismo nivel de confianza y al mismo tiempo reducir el error muestral anterior en un 40% ¿cuántos días más habría que observar?
4. El decano de una universidad desea estimar el número promedio de alumnos que abandonan el primer año. Se sabe que este valor se distribuye normalmente. Manda tomar una muestra de 21 alumnos en primer año la cual arroja un promedio de 5 desertores con un desvío de 1. Realice la estimación con un 95% de confianza.
5. En una muestra de 25 bebés de 12 semanas de vida se obtuvo un peso medio de 5900 gr y un desvío de 94 gr.
- a) estime el peso medio poblacional con un riesgo del 5%  
b) ¿cuántos niños más habría que pesar para reducir el error muestral anterior en un 30%?
6. Una empresa que produce alfajores quiere estimar el nivel de costos medios. Para ello seleccionó una muestra aleatoria de 16 períodos de producción que dio como resultado un promedio de costos de \$30.000 y un desvío estándar de \$1.000. Construya un intervalo de confianza del 95% para el costo medio real, suponiendo que la variable en la población tiene una distribución de probabilidad aproximadamente normal.
7. La empresa La Esmeralda SA necesita delinear una política de stock para un nuevo producto a comercializar. Para ello selecciona una muestra aleatoria de 22 unidades con el fin de estimar la vida útil promedio (en días) de dicho producto. Los resultados obtenidos de la prueba son los siguientes:

7 – 11 – 8 – 12 – 18 – 15 – 22 – 15 – 17 – 14 – 18  
22 – 25 – 14 – 11 – 13 – 9 – 17 – 28 – 21 – 23 – 22

Suponiendo que la vida útil (en días) se distribuye como una normal:

- a) Estime con un 95% de confianza la vida útil promedio del producto.  
b) ¿Cuál es el significado del resultado obtenido en el punto anterior?
8. Una empresa inmobiliaria quiere estimar el precio de venta promedio de casas en un barrio muy importante de la ciudad. Para ello selecciona una muestra aleatoria de 25 ventas recientes de casas y calcula el precio promedio del monto de dichas ventas y el desvío estándar que son \$148.000 y \$60.000 respectivamente. Suponiendo que la variable en la población se distribuye como una normal. Obtenga un intervalo de confianza del 95% para el precio promedio de ventas recientes de casas.
9. Una empresa productora de caños de aluminio desea analizar el diámetro (en cm) de los mismos y como primera medida seleccionó una muestra aleatoria que arrojó los siguientes valores de sus diámetros en cm.

2,5 – 2,7 – 3 – 2 – 2,5 – 3,2 – 1,8 – 1,9 – 2,3 – 2,4 – 2,7 – 3

Suponiendo que el diámetro en la población tiene una distribución normal:

- a) Construya un intervalo de confianza para el diámetro promedio con un 95% de confianza. Interprete el resultado obtenido.
- b) Construya un intervalo de confianza para el desvío con un 90% de confianza. Interprete el resultado obtenido.

10. A fin de obtener elementos para una toma de decisión acerca de la compra de un terreno en una esquina con la finalidad de construir un anexo de una escuela secundaria en donde funcionará el primario, la empresa propietaria encaró, en primer lugar, una investigación sobre la intensidad del tránsito de vehículos. Durante cada una de las 12 horas elegidas al azar (de las 7 de la mañana a las 7 de la tarde, en los días de la semana) se contó el número de vehículos que pasaron por dicha esquina. Suponiendo que la variable en estudio se distribuye en la población como una normal y habiendo obtenido una estimación puntual del promedio de vehículos por hora de 186 y una estimación puntual de la desviación estándar de 84. Calcule un intervalo de confianza del 95% para la intensidad media del tránsito.

11. Se midieron las velocidades de una muestra aleatoria de 20 automóviles, a medida que pasaban por un lugar considerado como punto de control. Los datos que se obtuvieron son los siguientes, en kilómetros por hora:

50 – 83 – 60 – 68 – 70 – 78 – 86 – 74 – 55 – 71  
74 – 65 – 80 – 64 – 40 – 60 – 68 – 58 – 84 – 90

Construya un intervalo de confianza del 95% para la velocidad promedio de los automóviles que pasan por tal punto de control.

12. Se seleccionó una muestra aleatoria de 20 pollos de una granja, a los que se les suministró una determinada dieta de engorde. Al cabo de un mes se comprobó que en promedio el aumento en ese período fue de 0,80 Kg por ave con un desvío estándar de 0,3 Kg. Por experiencias anteriores se sabe que el aumento de peso por ave se distribuye como una normal. Estimar con un 95% de confianza el verdadero aumento de peso promedio por ave que produce la dieta utilizada.

13. Se registraron las siguientes mediciones del tiempo de secado, en horas, de una marca de pintura sintética:

3,4 – 2,6 – 4,6 – 3,0 – 3,5 – 2,7 – 3,0 – 6,0  
4,0 – 2,8 – 4,5 – 4,0 – 5,3 – 4,0 – 5,0 – 6,1

Considerando que las mediciones representan una muestra aleatoria de una población normal, encuentre los límites del intervalo de confianza del 95% para el tiempo promedio de secado de la pintura sintética. ¿Cuál es su significado?

14. La empresa Plus SA fabrica un componente electrónico que se usa en las computadoras personales. Para cumplir con las especificaciones que le exigen sus principales clientes debe investigar acerca del promedio de vida de dicho componente, antes de tener la primera falla.

Para ello seleccionó una muestra aleatoria de 20 componentes y el promedio de duración de los mismos antes de fallar fue de 2.800 horas, con una desviación estándar de 80 horas. Estime un intervalo de confianza del 95% para la vida útil promedio de este componente. Interprete el resultado obtenido.

15. Una empresa provincial de energía ha publicado cifras acerca de la cantidad de kilowatts-hora que consumen varios electrodomésticos. Desde el gobierno de la provincia se intenta concientizar a la población para que planifiquen el uso de la energía eléctrica con el fin de disminuir los gastos en cada hogar. Si bien las cifras que proporciona la empresa están basadas en determinadas experiencias, últimamente se inició la realización de una investigación utilizando técnicas de la inferencia estadística. En primer lugar y considerando que el consumo promedio de un lavarropas automático es de 60 kw/hora al año, mediante una muestra aleatoria de 15 hogares incluidos en la investigación planificada se obtuvo un promedio de 56 kw/hora al año con una desviación estándar de 12 kw/hora.  
¿Sugieren estos indicadores que los lavarropas automáticos consumen, en promedio menos de 60 kw/hora al año? El consumo en la población se distribuye como una normal. Utilice un nivel de significación del 5%
16. Diariamente el organismo aduanero descubre un promedio de 38 millones de pesos en artículos introducidos ilegalmente al país. En 60 días elegidos aleatoriamente en el último año, detectó un promedio diario de 40 millones de pesos con un desvío estándar de 15 millones de pesos. ¿Indica la muestra, con un nivel de significación del 5%, que al jefe de aduanas debe preocuparle que el contrabando haya superado su nivel histórico?
17. Una O.N.G. desea auditar los gastos por todo concepto en que incurre un país por cada diputado nacional. Para ello evalúa los gastos directos (sueldo, viáticos, premios, etc.) y los indirectos (asesores y empleados a cargo). Se realizó una investigación que otorgó, sobre una muestra aleatoria de 45 diputados, que el costo anual promedio asciende a \$ 724.000 con un desvío estándar de \$ 97.000. Considerando que la variable tiene una distribución aproximadamente normal, se le pide que:
- Estime con una confianza del 95% el costo anual promedio en que incurre el país por tener un diputado.
  - La semana anterior a la investigación, el vocero presidencial había anunciado que el país incurría en un costo anual por todo concepto para cada diputado de \$691.000. ¿Indica la muestra (con un nivel de significación del 5%) que la información del gobierno está desactualizada y el costo aumentó?
18. Por denuncias de una Asociación de Protección al Consumidor se realizó una investigación en una reconocida embotelladora de gaseosas bajo la presunción de que llena sus botellas con menos líquido que el indicado en el envase 946,33 ml. Se testearon 200 botellas y los resultados son los siguientes:  
 $\bar{x} = 930$  ml.  $S = 44,36$  ml.  
Con un nivel de significación del 2% ¿Cuál fue la conclusión?

19. La producción anual de naranjas en un departamento de la provincia de Corrientes es una variable aleatoria aproximadamente normal con media 85 Tn/Ha. Pruebe con un nivel de significación del 5% si con un nuevo fertilizante que se está utilizando se obtiene un rendimiento mayor. Los datos de las 16 parcelas seleccionadas al azar en donde se probó el fertilizante tuvieron una producción promedio de 89 Tn/Ha y una desviación estándar de 6 Tn/Ha. Indique los supuestos considerados.
20. El departamento de procesamiento de datos de una gran empresa ha instalado nuevas terminales para reemplazar las unidades usadas hasta entonces. Los 25 operadores capacitados para emplear las nuevas máquinas necesitaron un promedio de 7,2 horas antes de alcanzar un nivel satisfactorio de rendimiento. El desvío muestral fue de 4.025 horas. La larga experiencia con operadores en las antiguas terminales revela que el promedio era de 8,1 horas en las máquinas antes de un rendimiento satisfactorio.

Con un nivel de significación del 5%, ¿debe el supervisor del departamento llegar a la conclusión de que es más fácil aprender a operar satisfactoriamente las nuevas máquinas? ¿Qué supuestos consideró?

21. El administrador del Hotel las Gaviotas, ubicado en la costa atlántica del país considera que la cuenta mensual del cliente promedio del hotel es de \$500. A fin de corroborar si esta temporada se ha modificado seleccionó una muestra aleatoria de 20 cuentas que dio una media muestral de \$470, una desviación estándar de \$80 y realizó una prueba una probabilidad del 10% de concluir erróneamente que el promedio ha bajado. ¿A qué conclusión llegó el administrador?
22. Un fabricante de automóviles asegura que su nuevo modelo alcanza un promedio de 10 km por litro de nafta. Se selecciona una muestra aleatoria de 10 automóviles y se realiza la prueba registrando la cantidad de km que circularon con un litro de nafta. Considerando un nivel de confianza del 95%

¿Cuál es la conclusión si los 10 automóviles dieron un promedio de 9 km/litro y una desviación estándar de 1,8 km/litro?

23. El Capital Humano, empresa que comercializa, entre otros programas para el desarrollo de los recursos humanos en las organizaciones, un programa de capacitación basado en la motivación, asegura que las tasas de rendimiento medio de los empleados aumentan después que recibieron dicha capacitación. La empresa textil La Maravilla SA, que recibió la oferta del programa de capacitación de referencia, hace un muestreo aleatorio de 25 empleados de la corporación Golden SA que recientemente lo implementó. La tasa de rendimiento medio por empleado era de 70 antes del curso y su distribución de probabilidad poblacional se considera aproximadamente normal. Los resultados de la muestra son:  $\bar{X} = 74$  ;  $S = 12$  ;  $n = 25$
- a) ¿A qué conclusión arribó la empresa textil?. Utilice un nivel de significación del 5%.
- b) Estime un intervalo de confianza del 95% para la tasa de rendimiento medio cuando se ha implementado el programa de capacitación. Explique qué relación hay entre ambos resultados.

24. Un establecimiento industrial que fabrica artículos de acero recibe una auditoría del organismo estatal de seguridad ambiental por la posible exposición de sus trabajadores al benceno. Se examinan 20 muestras de aire, tomadas durante un mes, para determinar el contenido de benceno. Los análisis produjeron los siguientes resultados:

$$\bar{X} = 2.1 \text{ partes por millón} \quad S = 0.7 \text{ partes por millón}$$

- ¿Está violando el establecimiento fabril las normas del gobierno que establece como límite máximo de contenido de benceno 1,5 partes por millón? Utilice  $\alpha = 0,05$ .
  - Calcule un intervalo de confianza del 95% para el parámetro en cuestión y establezca cuál es la relación con el test de hipótesis calculado.
  - ¿Qué supuestos consideró?
25. La empresa Lava & Limp se dedica a limpiar cristales exteriores de edificios en altura. En cada edificio trabaja un equipo de 8 personas que deben tardar como máximo 3 horas en completar su tarea. Periódicamente un supervisor controla este tiempo y si detecta que es significativamente mayor a lo estipulado, debe suspender al equipo afectado. Cada supervisor controla a un mismo equipo durante su trabajo en 30 edificios similares y luego emite su informe. El supervisor desea equivocarse como máximo un 5% de las veces que decida suspender a un equipo. Si de un control surgió un tiempo medio de 3horas y 20 minutos, con un desvío de 20 minutos, ¿qué decisión se debe tomar?



**RESPUESTAS A LOS EJERCICIOS DEL TRABAJO PRACTICO N°1**  
**Inferencia para la Varianza Poblacional**

1)  $S=1,0612$

a) IC = (0,717 ; 2,03) miliseg.

b)  $\sigma$  máx = 1,816 miliseg.

2)  $S=12,6$  hs

IC= [8hs 40min ; 23hs] al 95%

3)  $S=20$

IC=[15,68 ; 27,61] al 95%

4)  $S=0,70$

a) IC= [0,901 ; 1,412] al 95%

5)  $\sigma_0=280$ ;  $S=293,47$ ;  $X^2_{calc}=7,70$ ;  $X^2_{7;0,95}=14,0671$ . **No rechazo Ho.**

6)  $\sigma_0=200$ ;  $S=250$ ;  $X^2_{calc}=9,375$ ;  $X^2_{6;0,95}=12,592$ . **No rechazo Ho.**

7)  $\sigma_0^2=0,0005$ ;  $S^2=0,0006$ ;  $X^2_{calc}=28,8$ ;  $X^2_{24;0,95}=36,41$ . **No rechazo Ho.**

8)  $\sigma_0=0,75$ ;  $S=0,9445$ ; IC=[215 días ; 2 años y 116 días] al 95%;  $X^2_{calc}= 7.93$ ;  $X^2_{5;0,95}=11,07$ .  
**No rechazo Ho.**

9)  $\sigma_0=20$ ;  $S=20$ ;  $X^2_{calc}=25$ ;  $X^2_{25;0,95}=37,652$ . **No rechazo Ho.**

10)  $\sigma_0=3$ ;  $S=4$ ;  $X^2_{calc}=24.89$ ;  $X^2_{14;0,95}=23,685$ . **Rechazo Ho.**

11)  $\sigma_0=0,08$ ;  $S=0,1$ ;  $X^2_{calc}=37,5$ ;  $X^2_{24;0,99}=42,98$ . **No rechazo Ho.**

**Inferencia para la Media Poblacional con Desvío Poblacional Desconocido**

1)  $\bar{x}_{raya}=7,25$ ;  $S=0,7071$

a) IC=[6,659 ; 7,841] al 95%

b) NC= aprox 90%

2)  $\bar{x}_{raya} =17,3$ ;  $S=3,83$ ;

a) IC=[15,08 ; 19,52] al 95%

b)  $EM^*= 1,332$ ;  $n^*=14$  semanas más.

3)  $\bar{x}_{raya}=4$ ;  $S=0.75$

a) IC=[3,31 ; 4,69] al 95%

b)  $EM^*=0,41616$  ;  $n^*=8$  días más

4)  $\bar{x} = 5$  ;  $S = 1$

IC=[4,544 ; 5,45] al 95%

5)  $\bar{x} = 5900$  ;  $S = 94$

a) IC=[5861,2 ; 5938,8] al 95%

b) Habría que pesar **24** niños más.

6)  $\bar{x} = 30000$  ;  $S = 1000$

IC=[29467,25 ; 30532,75] al 95%

7)  $\bar{x} = 16,4545$  ;  $S = 5,738$

IC=[13,91 ; 18,99] al 95%

8)  $\bar{x} = 148000$  ;  $S = 60000$

IC=[123.233,22 ; 172.766,78] al 95%

9)  $\bar{x} = 2,5$  ;  $S = 0,4492$

a) IC=[2,21 ; 2,79] al 95%

b) IC=[**2,27 ; 2,73**] al 90%

10)  $\bar{x} = 186$  ;  $S = 84$

IC=[132,63 ; 239,37] al 95%

11)  $\bar{x} = 68,9$  ;  $S = 12,8386$

IC=[62,89 ; 74,91] al 95%

12)  $\bar{x} = 0,80$  ;  $S = 0,30$

IC=[0,66 ; 0,94] al 95%

13)  $\bar{x} = 4,03$  ;  $S = 1,137$

IC=[3,42 hs ; 4,64 hs] al 95%

14)  $\bar{x} = 2800$  ;  $S = 80$

IC=[2762,56 hs ; 2837,44 hs] al 95%

15)  $\mu = 60$ ;  $\bar{x} = 56$ ;  $S = 12$ ;  $X_{\text{critico}} = 54,57$  kw/hora. **No rechazo Ho.**

16)  $\mu = 38$  millones;  $\bar{x} = 40$  millones;  $S = 15$  millones;  $x_{\text{critico}} = 41,24$  millones. **No rechazo Ho.**

17)  $\mu = 691000$ ;  $\bar{x} = 724000$ ;  $S = 97000$

a) IC=[694.857,94 ; 753.142,06] al 95%

b)  $x_{\text{critico}} = 715292,64$ . **Rechazo Ho.**

18)  $\mu = 946,33$ ;  $\bar{x} = 930$ ;  $S = 44,36$ ;  $x_{\text{critico}} = 939,85$ . **Rechazo Ho.**

19)  $\mu=85$ ;  $x_{\text{raya}}=89$ ;  $S=6$ ;  $x_{\text{raya crítico}}=87,63$ . **Rechazo  $H_0$ .**

20)  $\mu=8,1$  hs;  $x_{\text{raya}}=7,2$  hs;  $S=4,025$ hs;  $x_{\text{raya crítico}}=6,72$ . **No rechazo  $H_0$ .**

21)  $\mu=500$ ;  $x_{\text{raya}}=470$ ;  $S=80$ .  $x_{\text{raya crítico}}=476,25$  **Rechazo  $H_0$ .**

22)  $\mu=10$ ;  $x_{\text{raya}}=9$ ;  $S=1,8$ ;  $x_{\text{raya crítico}}=8,957$ . **No rechazo  $H_0$ .**

23)  $\mu=70$ ;  $x_{\text{raya}}=74$ ;  $S=12$

a)  $x_{\text{raya crítico}}=74,11$ . **No rechazo  $H_0$ .**

b) IC=[69,05 ; 78,95] al 95%

24)  $\mu=1,5$ ;  $x_{\text{raya}}=2,1$ ;  $S=0,7$

a)  $x_{\text{raya crítico}}=1,77$ . **Rechazo  $H_0$**

b) IC=[1,77 ; 2,43] al 95%

25)  $\mu=3$ ;  $x_{\text{raya}}=3,333$ ;  $S=0,333$ .  $x_{\text{raya crítico}}=3,1024$ . **Rechazo  $H_0$ .**

**PRÁCTICO N° 2**

**COMPARACION**

**DE**

**POBLACIONES**

**NORMALES**



Temas:

- **Comparación de dos poblaciones normales a través de sus varianzas**
- **Comparación de dos poblaciones normales a través de sus promedios**
- **Comparación de dos poblaciones a través de sus proporciones**
- **Análisis de Varianza a uno y dos factores**



**Atención:** *En los ejercicios del presente capítulo, se utilizó como estimador de la varianza poblacional :*

$$S^2 = \frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2$$

---

**Comparación de dos poblaciones normales a través de sus varianzas**

1. A los fines de encarar una investigación detallada de los salarios de sus empleados, la empresa Los Molinos SA quiere saber en primer término si la variabilidad del salario por hora es la misma para dos de sus sucursales más importantes. Se toman muestras aleatorias de los salarios por hora en cada sucursal, las que dieron los siguientes indicadores:

$$\begin{array}{ll} s_1 = \$3,79 & n_1 = 21 \\ s_2 = \$2,48 & n_2 = 25 \end{array}$$

- a) Realice la prueba con un nivel de significación del 5%.

**Solución:**

Las hipótesis nula y de alternativa son:

$$\begin{array}{l} H_0: \sigma_1^2 = \sigma_2^2 \\ H_a: \sigma_1^2 \neq \sigma_2^2 \\ \alpha = 0.05 \end{array}$$

El estadístico de prueba F se distribuye según una F de Snedecor con  $(n_1-1)$  grados de libertad en el numerador y  $(n_2-1)$  grados de libertad en el denominador:

$$F_{calc} = \frac{s_1^2}{s_2^2} = \frac{14.3641}{6.1504} = 2.34$$

La C.R. es: si  $F_{calc} < F_{v1;v2;\alpha/2}$  o si  $F_{calc} > F_{v1;v2;1-\alpha/2}$  se rechaza la hipótesis nula  $H_0$  de que no existen diferencias significativas entre las varianzas de los salarios por hora.

$$F_{20;24;0,025} = 0,415 \quad ; \quad F_{20;24;0,975} = 2,327$$

**Decisión:**  $2,34 > 2,327$  es decir  $F_{calc} > F_c$ , luego se rechaza  $H_0$ . Hay evidencia muestral suficiente para rechazar la  $H_0$  de que no hay diferencias entre las varianzas de los salarios por hora de las dos sucursales.

- La Escuela Bancaria va a renovar su contrato de servicio de micro escolar para el año próximo y debe seleccionar entre las empresas Transporte Rápido SA y Transportes Eficaces SRL. Usarán la variancia de los tiempos de recepción como medida principal de la calidad del servicio, ya que los tiempos promedio son parecidos. Los valores bajos de variancia indican que el servicio es más consistente y de mejor calidad. Si son iguales seleccionarán la empresa que ofrezca mejores condiciones económicas. Se obtuvo una muestra aleatoria de 24 tiempos para Transportes Rápidos SA (1) y una de 15 para Transportes Eficaces SRL (2). La información que proporcionó es la siguiente: (Trabaje con un nivel de significación del 5%)  
 $S_1 = 12$  minutos       $S_2 = 15$  minutos

- Una empresa farmacéutica está considerando la compra de un nuevo instrumento diseñado para identificar los componentes de ciertos medicamentos, del que existen dos marcas en el mercado. Como parte de una comparación en la precisión, (consistencia con que el instrumento identifica los componentes del mismo material), se seleccionaron 10 muestras en tubos de ensayo de un lote bien mezclado de un medicamento; 5 se analizaron con el instrumento Exun y 5 con el instrumento Bons. Los datos de la tabla son los porcentajes del componente primario del medicamento identificados por los instrumentos. Suponiendo normalidad en las distribuciones de probabilidad en las poblaciones:

INSTRUMENTO EXUN	INSTRUMENTO BONS
40	46
45	48
39	43
50	41
45	44

¿Qué diría acerca de la precisión de los dos instrumentos?.

**Comparación de dos poblaciones normales e independientes a través de sus promedios  
(cuando se desconocen los desvíos poblacionales)**

**Problema resuelto**

Una empresa reconocida en el mercado fabricante de zapatos, que distribuye en el país y exporta, quiere determinar si aplicando un plan de incentivos para sus empleados la productividad aumenta. Implementa el plan de incentivos en 8 plantas elegidas al azar para comparar con las 10 plantas restantes que sólo perciben el salario normal. Los datos obtenidos son los siguientes y se refieren a la producción diaria: (Se sabe por investigaciones previas que los desvíos poblacionales pueden suponerse iguales, lo cual deriva en un caso II)

$$\begin{aligned}n_1 = 8 & \quad \bar{X}_1 = 150 \text{ unidades} \quad s_1 = 20 \text{ unidades} \\n_2 = 10 & \quad \bar{X}_2 = 130 \text{ unidades} \quad s_2 = 18 \text{ unidades}\end{aligned}$$

¿Cuál será la conclusión de la investigación si trabaja con un nivel de significación del 1%?

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 > 0 \quad \text{C.R.: } t_{\text{calc}} > t_{8+10-2; 0,99} = 2.921$$

$$s_a^2 = \frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2} = \frac{7 * 400 + 9 * 324}{16} = 357.25$$

$$t_{\text{calc}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_a^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{150 - 130}{\sqrt{357,25 * \left(\frac{1}{8} + \frac{1}{10}\right)}} = 2.23$$

**Conclusión : Como  $t_{\text{calc}} = 2.23 < 2.921$  no se rechaza  $H_0$ . No existe evidencia suficiente para rechazar la  $H_0$ . Es decir, no se puede inferir que la productividad aumentó.**



### Problemas propuestos

1. La empresa SELENI SRL quiere analizar los tiempos de entrega de dos proveedores de materia prima para su producción (Domínguez y Lexis). El proveedor Domínguez satisface con su trabajo a la empresa, pero esta lo conservará si el promedio de tiempo de entrega de la materia prima que provee es igual o menor que el promedio de tiempo que emplea Lexis, también proveedor de la misma materia prima. Si no es así, la empresa solo le comprará a Lexis.

Suponiendo que se obtuvieron los siguientes indicadores de dos muestras aleatorias de entregas de ambos proveedores, indique cuál sería la conclusión. Utilice un nivel de significación del 5%.

Lexis	n=50	$\bar{x} = 14$	días	s=3	días
Dominguez	n=31	$\bar{x} = 12.5$	días	s=2	días

2. La fábrica de automóviles “A” realiza un estudio para rebatir la aseveración de su principal competidora “B” de que los automóviles que ésta produce consumen menos combustible. El gerente de producción considera que la variancia del consumo de combustible de los autos de la empresa competidora es mayor que la de los modelos de “A”. Se llevó a cabo una prueba con dos muestras de automóviles de ambas fábricas que proporcionaron los siguientes resultados en km por litro de nafta y para realizarla se utilizó un nivel de significación del 5%.

A : 10 – 11 – 15 – 9 – 8 – 12 – 16 – 13  
B: 11 – 18 – 14 – 17 – 9 – 9 – 18 – 12

- a) ¿La empresa “B” consume significativamente mayor cantidad de combustible que la empresa “A”?
3. Una empresa de golosinas decide cambiar el saborizante que usa en la fabricación de sus caramelos. Para probar si existen diferencias significativas en cuanto a los niveles medios de azúcar concentrados, luego de aplicar el saborizante, toma dos muestras y obtiene los siguientes indicadores:

Con el saborizante anterior:	n = 80	$\bar{X} = 5\text{mg}$	S = 1mg
Con el nuevo saborizante:	n = 120	$\bar{X} = 5,3\text{mg}$	S = 1,5mg

¿Existen diferencias significativas considerando un nivel de significación del 10%?

4. Un empresario trabaja dos marcas similares de remeras deportivas y quiere confirmar con evidencia concluyente que la marca “sport” ha disminuido en promedio su demanda en comparación a la demanda de la marca “fitness”

De no encontrar evidencia concluyente, concluirá que se trata de un problema de preferencias del consumidor.

Luego de 24 semanas, llegó a la conclusión de que la marca “sport” se había vendido a un promedio de 40 unidades por semana, con un desvío estándar de 15 unidades, mientras que de la otra marca se vendió a un promedio de 60 unidades por semana, con un desvío estándar de 24 unidades.

Probar, con un nivel de significación del 1%, si el problema tiene que ver con las preferencias respecto de las marcas o no.

5. Con el fin de analizar el ausentismo en su empresa, con sede en el microcentro porteño, el gerente de RR.HH. intenta analizar varias variables que puedan estar incidiendo en el tema. En primer lugar quisiera determinar si el número de días no trabajados en promedio por año es el mismo para los trabajadores que residen en Capital y los trabajadores que residen en el GBA. Se seleccionan dos muestras aleatorias de  $n_1 = 50$  trabajadores residentes en Capital y  $n_2 = 45$  trabajadores residentes en GBA. Los indicadores calculados con los datos obtenidos de las muestras son los siguientes:

$$\begin{array}{ll} \bar{X}_1 = 8,4 \text{ días} & S_1 = 2,2 \text{ días} \\ \bar{X}_2 = 9,1 \text{ días} & S_2 = 3 \text{ días} \end{array}$$

- a) Realice la prueba al nivel de significación del 1%. Interprete el resultado obtenido.  
b) (PARA DEBATIR) Si Usted fuera el gerente de recursos humanos de la empresa, ¿cuáles análisis de información encararía?
6. Nadia R. analista de una empresa industrial, quiere analizar los tiempos medios de una fase de la producción. Los resultados de su investigación son los siguientes. Para ello toma dos muestras independientes de trabajadores, obteniendo lo siguiente:

$$\bar{X}_1 = 320 \text{ minutos} \quad S_1 = 25 \text{ minutos} \quad n_1 = 32 \quad \bar{X}_2 = 350 \text{ minutos} \quad S_2 = 37 \text{ minutos} \quad n_2 = 37$$

- a) ¿Tiene que suponer poblaciones normales?  
b) Realice la prueba con un nivel de significación del 5 %.  
c) ¿Cuál es la conclusión de Nadia?
7. Se realizó un estudio para probar si hay diferencias en los sueldos de los empleados de dos grandes empresas siderúrgicas. Una muestra aleatoria de 20 operarios de la empresa Acero Cali SA arrojó un salario mensual promedio de \$1500 con una desviación estándar de \$200. Una muestra aleatoria de 25 operarios de la empresa Acería Sud SRL produjo un salario promedio de \$1700 con una desviación estándar de \$220. ¿Existen diferencias significativas entre dichos promedios? Utilice un nivel de significación del 5%. Asuma que las poblaciones se distribuyen como una normal.
8. Los siguientes datos representan la TEA (Tasa Efectiva Anual), en porcentaje, sobre colocaciones de dinero en el Mercado Abierto electrónico de una muestra de 10 bancos comerciales de la ciudad capital de la provincia más importante del país y 10 cuentas de empresas financieras de la misma ciudad:

Entidades	TEA									
	Bancos	2,25	2,32	2,02	1,92	2,02	1,82	2,10	2,38	2,30
Financieras	2,43	2,53	2,38	2,50	3,00	2,50	2,55	3,20	2,50	2,02

- a) ¿Existe evidencia de una diferencia significativa en las TEA sobre las colocaciones en los dos tipos de entidades? Utilice un nivel de significación del 5%.
- b) ¿Qué supuestos consideró para efectuar la prueba?

**Comparación de dos poblaciones normales a través de sus promedios en muestras pareadas**

9. Con el fin de lograr una corrección de los exámenes parciales de 5 de los cursos de Estadística para Administradores que se dictan en la Facultad y basada en la mayor similitud respecto de los criterios utilizados, se decidió analizar usando muestreo, la corrección efectuada por los ayudantes de prácticos de tales cursos, Suárez y Benítez. Para ello se seleccionaron 7 exámenes del primer parcial del total correspondiente al último cuatrimestre y se les solicitó efectuar su corrección a cada ayudante. Las calificaciones asignadas fueron las siguientes:

AYUDANTE SUAREZ	AYUDANTE BENITEZ
8	7
7	8
7	5
5	4
9	6
6	4
5	7

Sería deseable que no existieran diferencias significativas entre las calificaciones promedio por corrector. Utilice un nivel de significación del 5%. Interprete el resultado.

10. Una firma comercializadora de electrodomésticos contrató a una empresa publicitaria y le encargó el desarrollo de una campaña de publicidad para sus productos. Teniendo en cuenta la incidencia que puede tener la publicidad sobre las ventas y con el fin de evaluar el resultado de la campaña, el gerente de comercialización de la firma toma datos de las ventas (en miles de pesos) antes y después de realizar la campaña en 10 sucursales seleccionadas de su empresa. En la tabla siguiente se presentan los datos.

SUCURSAL	VENTAS PROMEDIO ANTES DE LA CAMPAÑA	VENTAS PROMEDIO DESPUES DE LA CAMPAÑA
A	60	62
B	48	50
C	51	53
D	55	55
E	45	56
F	49	50
G	54	57
H	53	55

- a) Establezca las hipótesis adecuadas.  
 b) Establezca la condición de rechazo y la regla de decisión  
 c) ¿Cuál es la conclusión de la prueba trabajando con un nivel de significación del 5%?
11. Una mandataria de taxis de la Capital está considerando el reequipamiento de su flota con una marca nueva de neumáticos radiales cuya promoción considera que mejora la economía de combustible. Para probarlo se equiparon 10 automóviles seleccionados al azar de la flota con los neumáticos de la nueva marca, los que recorrieron un determinado trayecto. Sin cambiar los choferes ni los vehículos recorrieron el mismo trayecto pero habiendo cambiado los neumático de la nueva marca por la marca anterior. Los datos que siguen están expresados en km /litro:

Automóvil	Kilómetros por litro	
	Nueva marca de neumáticos	Marca anterior de neumáticos
1	4,2	4,0
2	4,6	4,9
3	6,6	6,2
4	7,1	6,8
5	6,8	6,6
6	5,4	4,8
7	5,7	5,9
8	6,3	6,3
9	7,3	6,2
10	6,1	6,0

Con un nivel de significación del 5 %. ¿Puede inferirse que los automóviles equipados con la nueva marca de neumáticos consumen en promedio menos combustible que si estuvieran equipados con la anterior, en uso? Suponga que las poblaciones están normalmente distribuidas.



## **Comparación de poblaciones normales a través de sus proporciones**

### **Problemas propuestos**

1. La empresa industrial Santa Fe Hogar que fabrica electrodomésticos tiene dos líneas de producción, A y B dedicadas a la fabricación de cocinas. El gerente de producción considera necesario analizar la calidad con que se trabaja en ambas líneas. Para ello se selecciona una muestra aleatoria de 1000 unidades de la línea A en la que se encuentran 10 unidades defectuosas, mientras que en la muestra aleatoria de 1200 unidades seleccionadas de la producción de la línea B se encontraron 25 unidades defectuosas. Estime un intervalo de confianza del 99% respecto a la diferencia de unidades defectuosas producidas por las dos líneas.
2. Como parte de una investigación socioeconómica iniciada en dos ciudades se considera necesario comparar sus tasas de desocupación, ya que históricamente los niveles de desocupación de ambas ciudades han sido los mismos (11%). Se seleccionan muestras aleatorias de residentes de cada ciudad y se obtuvieron estos resultados:

$n_1 = 550$  residentes pertenecientes a la PEA (\*)       $X_1 = 55$  personas desocupadas

$n_2 = 750$  residentes pertenecientes a la PEA       $X_2 = 90$  personas desocupadas

( \* ) Población económicamente activa

Realice la prueba usando un nivel de significación del 5%. Interprete el resultado obtenido.

3. En un establecimiento industrial se proyecta adquirir una máquina nueva de alta precisión para su proceso de fabricación. En el mercado hay dos marcas de dicha máquina. A los fines de decidir cuál adquirir, se analizará el porcentaje de piezas defectuosas que produce cada máquina. Para ello se seleccionó una muestra de la producción de la marca Exis de 35 piezas que dio un 6% de piezas defectuosas y una muestra de la producción de la marca Selser del mismo tamaño que produjo un 10 % de piezas defectuosas. ¿Hay diferencias en la proporción de defectuosos en la producción de ambas marcas? Use un nivel de significación del 5%.
4. Una importante empresa de servicios públicos que suministra gas tiene intenciones de comprar una empresa más pequeña considerando que la opinión favorable de los accionistas es la misma, ya sean varones como mujeres. No obstante el asesor principal dice que la proporción de accionistas varones que están a favor de la compra es mayor que la proporción de accionistas mujeres. Se le encarga al estadístico de la empresa que ratifique o rectifique el informe del asesor principal. Así es que selecciona una muestra aleatoria de 100 accionistas (40 varones y 60 mujeres) y solicita su opinión por consulta mediante correo electrónico. Determina así que 22 varones y 32 mujeres están a favor de la compra. Realice usted una prueba con estos datos utilizando un nivel de significación del 5 %.
5. Un comerciante textil minorista decidió encarar un estudio acerca de la llegada de la publicidad de su comercio (un espacio nuevo en el periódico de cada una de dos ciudades donde tiene sucursales). Para ello realizó una consulta telefónica a 1000 personas seleccionadas al azar de cada ciudad. El resultado fue:  
  
proporción en la muestra de la ciudad 1 de personas que leyeron el anuncio= 0,20  
  
proporción en la muestra de la ciudad 2 de personas que leyeron el anuncio= 0,17  
  
¿A qué conclusiones arribó el comerciante? Use un  $\alpha = 5 \%$ .

### **Análisis de varianza a un sólo factor**

#### **(EJERCICIO RESUELTO)**

Se encara una investigación de mercado a los fines de detectar si hay diferencias en el consumo de agua mineral de marcas líderes. El siguiente cuadro muestra los resultados de un relevamiento del consumo de agua mineral de marcas líderes. El operativo consistió en registrar, durante cinco semanas, el consumo semanal medido en litros por persona (promedio). Se supone que las poblaciones son normales, independientes y homocedásticas.

- a) Con un nivel de significación del 5%. ¿Hay evidencia suficiente que indique diferencias significativas en el consumo de una u otra marca? Realice un análisis de varianza con los datos.

SEMANA	MARCA1 AGUA DULCE	MARCA2 CRISTAL	MARCA 3 CLARA Y PURA	MARCA 4 VILLA FIORITO
1	4	7	14	3
2	3	8	10	5
3	5	8	8	8
4	5	7	14	8
5	4	6	12	10

### Solución

$$r=4 \quad n_i=5$$

$$nt=20$$

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4$$

$$\bar{x}_1 = 4.2 \quad \bar{x}_3 = 11.6 \quad \bar{x} = 7.45 \quad SCE=141.35 \quad SCD=63.6$$

$$\bar{x}_2 = 7.2 \quad \bar{x}_4 = 6.8$$

$$\text{Valor empírico } F_e = \frac{\frac{SCE}{n-r}}{\frac{SCD}{r-1}} = \frac{\frac{141.35}{16}}{\frac{3}{63.6}} = 11.85$$

$$\text{Valor teórico } F_{(4-1; 20-4)}(0.95) = 3.24$$

Se rechaza  $H_0$ . Hay evidencia suficiente en los datos muestrales para rechazar la hipótesis nula de igualdad de los promedios de consumo de agua mineral de las 4 marcas. Se puede inferir que al menos, el promedio de consumo de una marca es distinto.

### Ejercicios para practicar

1. Dado el creciente aumento en el índice de consultas oftálmicas por parte de personas que trabajan como operadores de PC, se desea establecer si existe una incidencia sobre el número de consultas por la utilización de distintas marcas de monitores. Los datos son las consultas mensuales durante 4 meses de un centro oftalmológico. Efectúe la prueba con  $\alpha = 0.05$ .

MESES	MONITOR G	MONITOR V	MONITOR S	MONITOR P	MONITOR A
1	20	21	31	29	22
2	33	34	43	44	37
3	14	10	17	15	12
4	24	26	32	31	25

**Respuesta:  $F_c(4;15)=3.056$   $F_{calc}=0.588$  No se rechaza  $H_0$**



2. El gerente de comercialización de la empresa propietaria de la cadena de Supermercados Del Sur SA desea saber si las tres sucursales tienen el mismo promedio por compra en pesos. Selecciona una muestra de 6 compras en cada una de las sucursales. Los datos obtenidos se presentan en la siguiente tabla. Efectúe la prueba con un nivel de significación del 1%.

SUCURSAL A	SUCURSAL B	SUCURSAL C
12,00	16,00	9,50
24,00	18,00	7,00
14,5	19,50	10,00
26,00	21,00	8,00
17,50	14,60	12,00
20,00	22,00	6,80

**Respuesta:  $F_c(2;15) = 6.35$   $F_{calc} = 14.11$  Se rechaza  $H_0$**

3. El gerente de comercialización de Geselino SRL quiere determinar si existen diferencias en la vida promedio entre cinco marcas de televisores con pantallas estándar y de última generación. Se prueban muestras aleatorias de cinco televisores de cada marca. Los resultados (en horas) son:

MARCA 1	MARCA 2	MARCA 3	MARCA 4	MARCA 5
3520	4025	4520	3987	3620
3631	3901	4325	4123	3358
3954	3756	4189	3852	3428
3950	3760	4200	3800	4100
3780	3800	3900	3970	3900

Para un nivel de significación del 5%. ¿hay alguna diferencia significativa entre estas cinco marcas de televisores?

**Respuesta:  $F_c(4;20) = 2.87$   $F_{calc} = 5.117$  Se rechaza  $H_0$**

4. Carlos Arfelli, gerente de comercialización de Vapored SA quiere saber si hay diferencias entre el monto de las compras que realizan sus clientes mediante el uso de tres tarjetas de crédito: Nicard, Vis y Ccard. Se decide analizar compras con las tres tarjetas en los últimos diez días y los datos (en pesos) son los siguientes:

NICARD	VIS	CCARD
100	90	90
90	71	112
65	102	77
70	83	68
77	25	89
23	18	25
175	60	65
80	67	80
121	58	88
47	68	66

¿Puede concluir el Sr. Arfelli que hay diferencias en los promedios de compras realizadas con las tres tarjetas de crédito? Use un nivel de significación del 5%.

**Respuesta:  $F_c(2;27) = 3.35$   $F_{calc} = 1.0845$  No se rechaza  $H_0$**

5. El señor Arturo Gómez, gerente de ventas de “La buena suerte SA” que vende muebles de oficina, desea saber si existen diferencias significativas entre los promedios de ventas de las 3 sucursales que tiene, de las cuales la sucursal Barracas ha dejado de operar transitoriamente desde hace un mes y la sucursal Nuñez desde hace 2 meses. Construyó para ello el siguiente cuadro, con los montos expresados en cientos de miles de pesos. Considere un nivel de significación del 5%.

SUC. BARRACAS	SUC. CENTRO	SUC. NÚÑEZ
3	5	4
4	6	4
7	6	5
8	8	5
4	9	6
5	8	4
6	7	-
-	6	-

**Respuesta:  $F_c(2; 18) = 3.55$   $F_{calc} = 4.71$  Se rechaza  $H_0$**

6. La empresa MICROARTS vende cuatro clones diferentes de su sistema operativo. El gerente de ventas tiene los datos del número de unidades vendidas de cada marca en muestras de días de diferente tamaño. Con un nivel de significación del 5% quiere comprobar si hay diferencias significativas entre los promedios de ventas de las distintas marcas

DIA	MARCAS			
	CARO	LILA	SUKI	NINO
1	17	27	13	18
2	21	13	15	25
3	13	29	17	15
4	27	9	23	27
5	12	-	10	12
6	-	-	21	-

**Respuesta:**  $F_{\text{calc}} = 0.2284$   $F_c(3,17) = 3.24$  **No se rechaza  $H_0$**

7. Un analista financiero desea determinar si la tasa media de rendimiento de tres tipos de acciones: ROMA SA, WARTA SRL y BRINSKEN SA. arroja diferencias significativas. Utilice la siguiente información muestral para determinarlo con un nivel de significación del 5%.

ROMA SA	WARTA SRL	BRINSKEN SA
14	12	15
18	12	13
17	11	18
16	12	15
20	12	18
-	-	13

**Respuesta:**  $F_{\text{calc}} = 9.98$   $F_c = 3.81$  **Se rechaza  $H_0$**

8. En una estrategia de venta se compararon los efectos que sobre éstas tuvieron cuatro promos en 5 cadenas de supermercados. Determine si las promos producen diferentes efectos en las ventas. Utilice  $\alpha = 0.01$  La variable medida en el siguiente cuadro es ventas de igual producto en cientos de miles de unidades para el mismo mes.

Promoción	VENTAS (cientos de miles de unidades)				
	COTA	DISKA	KARREFU	OLMAR	CHUMBO
Muestra Gratis	78	87	81	89	85
20% más	94	91	87	90	88
Descuento del 10%	73	78	69	83	76
Otro Producto de regalo	79	83	78	69	81

**Respuesta:  $F_{\text{calc}} = 0.3438$   $F_c = 4.89$  No se rechaza  $H_0$ .**

9. A la empresa Méndez le interesa comprobar si hay diferencias en los promedios de tiempo que utilizan 3 máquinas de marcas distintas para mezclar un conjunto de materiales que se usan en la construcción. Tomadas las observaciones (en minutos) en el trabajo de las máquinas se obtuvo la siguiente tabla:

Marca 1	Marca 2	Marca 3
21	25	21
20	28	20
26	26	19
24	31	23
22	27	22

Realice la prueba utilizando un nivel de significación del 5 %.

**Respuesta:  $F_{\text{calc}} = 12,235$ ;  $F_c = 3,89$  Se rechaza  $H_0$**

**RESPUESTAS A LOS EJERCICIOS DEL TRABAJO PRACTICO N°2**

**Inferencia para dos poblaciones normales**

**Comparación de poblaciones normales a través de sus varianzas**

- 1)  $F_{calc} = 2,3354$  ;  $F_{c1} = 0,41536$  y  $F_{c2} = 2,3273$  **Rechazo  $H_0$**
- 2)  $F_{calc} = 0.64$  ;  $F_{c1} = 0,40054$  Y  $f_{c2} = 2,80$  **No Rechazo  $H_0$**
- 3)  $F_{calc} = 2,6985$  ;  $F_{c1} = 0,1041$  y  $f_{c2} = 9,6044$  **No Rechazo  $H_0$**

**Comparación de poblaciones normales e independientes a través de sus promedios**

- 1)  $T_{calc} = -2,6982$  ;  $T_c = -1,664$  **Rechazo  $H_0$**
- 2)  $T_{calc} = 1,0435$  ;  $T_c = 1,761$  **No Rechazo  $H_0$**
- 3)  $T_{calc} = -1,697$  ;  $T_c = +-1,653$  **Rechazo  $H_0$**
- 4)  $T_{calc} = 3,462$ ;  $T_c = +-2,41$  **Rechazo  $H_0$**
- 5)  $T_{calc} = -1,3056$  ;  $T_c = +-2,629$  **No Rechazo  $H_0$**
- 6)  $T_{calc} = -3,88$  ;  $T_c = +-1,996$  **Rechazo  $H_0$**
- 7)  $T_{calc} = -3,15$  ;  $T_c = +-2,017$  **Rechazo  $H_0$**
- 8)  $T_{calc} = -3,51$  ;  $T_c = +-2,101$  **Rechazo  $H_0$**

**Comparación de dos poblaciones normales a través de sus promedios en muestras pareadas**

- 9)  $T_{calc} = -1,28$  ;  $T_c = 2,447$  **No Rechazo  $H_0$**
- 10)  $T_{calc} = 2,392$  ;  $T_c = 1,89$  **Rechazo  $H_0$**
- 11)  $T_{calc} = 1,8818$  ;  $T_c = 1,833$  **Rechazo  $H_0$**

**Comparación de proporciones**

- 1) IC: (0; 0,0028) al 99%
- 2)  $Z_{calc} = -1,132$  y  $Z_{crit} = +-1,96$  **No rechazo  $H_0$**
- 3)  $Z_{calc} = -0,617$  y  $Z_{crit} = +-1,96$  **No rechazo  $H_0$**
- 4)  $Z_{calc} = 0,1638$  y  $Z_{crit} = 1,64$  **No rechazo  $H_0$**
- 5)  $Z_{calc} = 1,73$  y  $Z_{crit} = 1,64$  **Rechazo  $H_0$**



# PRÁCTICO N° 3

## REGRESION LINEAL

### SIMPLE

### Y

### CORRELACION

### LINEAL





## Regresión lineal simple y correlación



### Problema resuelto

El siguiente cuadro presenta el número de reservas hoteleras realizadas por internet durante la temporada alta y los montos efectivamente facturados por tales reservas en los 5 hoteles mas lujosos de la ciudad de Tandil. Se pide: a) Hallar el modelo de regresión lineal que permita estimar los montos facturados en función de las reservas observadas por internet. b) ¿Qué porcentaje de las variaciones en los montos facturados queda explicado por el modelo obtenido en el punto anterior? c) ¿Es válido el modelo obtenido en el punto a)? Considere un  $\alpha=0,05$  d) Estimar al 95% el monto que se facturará en un hotel de similares características que haya tenido 5500 reservas.

Nº de reservas hoteleras (en miles)	Montos facturados en millones de \$
X	Y
1	4
2	2
4	5
7	7
10	9

### Solución:

Obs. Nº	x	y	x · y	x <sup>2</sup>	y <sup>2</sup>
1	1	4	4,00	1,00	16,00
2	2	2	4,00	4,00	4,00
3	4	5	20,00	16,00	25,00
4	7	7	49,00	49,00	49,00
5	10	9	90,00	100,00	81,00
Sumas	24	27	167	170	175

$$n = 5 \quad S_{xy} = \sum xy - n \cdot \bar{x} \cdot \bar{y} = 167 - 5 \cdot 4,8 \cdot 5,4 = 37,4$$

$$\bar{x} = 4,8 \quad S_{xx} = \sum x^2 - n \cdot \bar{x}^2 = 170 - 5 \cdot 4,8^2 = 54,8$$

$$\bar{y} = 5,4 \quad S_{yy} = \sum y^2 - n \cdot \bar{y}^2 = 175 - 5 \cdot 5,4^2 = 29,2$$

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{37,4}{54,8} = 0,6825$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 5,4 - 0,6825 \cdot 4,8 = 2,124$$

$$R^2 = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} = \frac{37,4^2}{54,8 \cdot 29,2} = \frac{41}{54,8 \cdot 29,2} = 0,8742$$

a)  $\hat{y} = 2,124 + 0,6825 \cdot x$

Interpretación de las componentes del modelo:

$b_0=2,124$  Aún sin reservas por internet, se estima una facturación de 2,12 millones de \$

$b_1=0,6825$  Por cada mil reservas adicionales, se estima un incremento en la facturación de 682 mil pesos aproximadamente o por cada reserva adicional se estima un incremento de 682 \$ (esto se logra si dividimos por 1000 ambos valores de las variables)

b)  $R^2 = 0.8742$  El 87,42 % de las variaciones observadas en los montos facturados, está explicado por el modelo hallado en a), es decir se halla explicado a través de las variaciones observadas en la variable explicativa x: # de reservas realizadas por internet.

c) El modelo es válido para hacer pronósticos, pues  $R^2$  es mayor a 0,50

$H_0 : \beta_1 > 0$  El modelo NO es válido

$H_1 : \beta_1 > 0$  El modelo es válido y directo

C.R. : Si  $t_{calc} > t_{n-2;1-\alpha} = t_{3;0,95} = 2,353$

R.D. : Si Rechazo  $H_0$ , entonces el modelo es válido y directo.

$$S^2 = \frac{1}{n-2} * [S_{yy} - b_1^2 * S_{xx}] = \frac{1}{3} * [29.2 - 0.6825^2 * 54.8] = 1.22506$$

$$t_{calc} = \frac{b_1 - \beta_1}{\frac{\sqrt{S^2}}{\sqrt{S_{xx}}}} = \frac{0.6825 - 0}{\sqrt{\frac{1.22506}{54.8}}} = 4.5646$$

Conclusión: Como  $t_{calc} = 4.5646$  es mayor a 2.353 Rechazo  $H_0$  y el modelo es válido y directo.

d) Aquí se pide un intervalo de predicción para un valor particular de la variable y, dado el valor de  $x = 5,5$  miles de reservas

$$\left[ \hat{y}(x_0 = 5,5) \pm t_{n-2;1-\alpha/2} * \sqrt{\left( S^2 * \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right] \right)} \right]$$

$$\left[ 2,124 + 0,6825 * 5,5 \pm t_{3;0,975} * \sqrt{\left( 1,2251 * \left[ 1 + \frac{1}{5} + \frac{(5,5 - 4,8)^2}{54,8} \right] \right)} \right]$$

$$[5,878 \pm 3,182 * 1,217] = [2 ; 9,75]$$

Con un 95% de confianza para un nivel de 5500 reservas, se pronostica una facturación de entre 2 y 9,75 millones de pesos.



### Problemas propuestos

1. Una empresa vial relevó durante una semana el número de vehículos que pasaban por día por su autopista así como el número de accidentes que se producían de manera de establecer cuál es la relación entre dichas variables.

Día	Número de autos	Número de accidentes
Lunes	4500	20
Martes	5200	25
Miércoles	5000	42
Jueves	5300	35
Viernes	6200	40
Sábado	3800	30
Domingo	2500	8

Calcule:

- El modelo de regresión lineal muestral.
  - ¿Es válido el modelo?
  - ¿Cuántos accidentes se espera que se produzcan si pasan 6100 vehículos?
  - ¿Cuál es la intensidad de la relación entre las variables?
2. Dadas las siguientes variables X: Volumen físico de la producción. Y: Kw/h consumidos. Explicar el significado del resultado obtenido.

Meses	X	Y
1	91	94
2	78	82
3	100	103
4	99	100
5	101	105
6	102	107
7	104	109
8	105	110
9	100	102
10	101	105
11	100	110
12	106	120

- a) Determinar la recta de regresión que explica las variaciones en los Kw/h consumidos ante las variaciones en el volumen físico de la producción.
  - b) ¿Qué porcentaje de las variaciones observadas en los Kw/h consumidos queda explicado por el modelo?
  - c) Determinar si el modelo es válido al 5%
  - d) Construir un intervalo de confianza al 95% para predecir el valor medio de Kw/h consumidos si se sabe que el volumen de la producción ha sido de 103 unidades.
3. Una empresa supone que su línea de producción opera con mayores defectos al sobrecargarla de materia prima. Para ello observó durante 20 días el ciclo productivo, cuyos datos figuran en la siguiente tabla:

Materia prima (Kg)	Merma obtenida (Kg)
80	2
85	2
87	3
90	3
92	3
95	5
97	5
100	6
105	8
115	11

- a) Obtenga el modelo lineal que pronostique la merma obtenida en función a la materia prima ingresada.
  - b) Verifique si el modelo es válido al 5%.
  - c) Estimar con un 90% de confianza la merma obtenida si en el próximo ciclo se ingresan 100kg de materia prima.
4. El supervisor de mantenimiento de una empresa de radiotaxis quiere determinar la relación entre el costo anual de mantenimiento de los vehículos y los años de uso de los mismos. Para ello reúne datos de una muestra aleatoria (10 vehículos del total de la flota que posee la empresa).

Vehículo	Costo de mantenimiento(\$)	Años de uso
1	900	8

2	850	8
3	1100	10
4	720	6
5	790	7
6	1050	10
7	1200	11
8	910	9
9	345	3
10	330	3

- Determine la ecuación de la recta de regresión del costo “y” en función de los años de uso “x”.
  - Valide el modelo al 10%
  - Estime al 95% de confianza la variación estimada en el costo por cada año que transcurre
- 5) Los siguientes datos corresponden a las ventas en miles de pesos de ocho meses de la empresa textil Deymond SA propietaria de la marca de paraguas FGL y los gastos en publicidad de la empresa.

Gastos Publicidad (miles de \$)	Ventas (miles de \$)
36	1391
283	5401
310	5600
280	2600
40	1020
50	1100
90	1800
100	2200

- Hallar el modelo que mejor explique la variación en las ventas en función de los gastos en publicidad. Interprete los componentes del modelo.
  - Valide el modelo al 10%
  - Estimar al 95% la venta promedio del mes en el que se inviertan 70000\$ en publicidad
- 6) Un auditor del gobierno municipal de una ciudad desea estudiar la relación entre los impuestos municipales y la edad de las casas familiares en dicha ciudad. Cuenta con los datos de una muestra aleatoria de 11 casas familiares que se presentan en la siguiente tabla:

Edad de las casas (en años)	Impuesto (en \$)
1	925

2	870
4	809
4	720
5	694
8	630
10	626
10	562
12	546
15	523
20	480

- a) Halle la recta que explique la variación en el impuesto en función a la edad de las casas.
  - b) ¿Qué porcentaje de las variaciones observadas en los montos impositivos se hallan explicadas por el modelo?
  - c) Estime al 95% y de ser posible el impuesto que pagará una casa cualquiera que tiene 10 años.
- 7) Los siguientes datos se refieren al ingreso (X) por las entradas a partidos de fútbol de un estadio de Capital Federal, durante 10 meses y las ventas (Y) en comestibles y bebidas en dichos partidos. Efectúe un análisis de correlación y pruebe la significatividad del coeficiente de correlación de Pearson para explicar la relación entre las variables. Utilice  $\alpha = 0.05$ . Interprete los resultados obtenidos.

Partido	Entradas (miles de \$)	Ventas (miles de \$)
1	340	85
2	350	90
3	350	87
4	400	91
5	305	70
6	210	60
7	280	80
8	300	59
9	380	71
10	385	70

- 8) En el análisis del funcionamiento de los sistemas de transporte urbano de pasajeros es necesario poder estimar el tiempo de viaje esperado entre distintos lugares. Habiéndose recolectado datos sobre los tiempos y distancias de viajes entre dos lugares de una ciudad para los automotores de pasajeros de la empresa La Continental SA se realizó un análisis de regresión lineal simple para un conjunto de 10 datos ( $y =$  tiempo de viaje en minutos,  $x =$  distancia entre los dos puntos en kilómetros) y los resultados son:

$$Y = 2.50 + 1.93 X \quad r^2 = 0.676$$

- a) ¿Hay pruebas suficientes que indiquen que la distancia entre dos puntos tiene una relación lineal con el tiempo de viaje urbano de automotores de pasajeros? Utilice  $\alpha = 0.05$
- b) Prediga el tiempo de viaje para un automotor de pasajeros que recorre una distancia de 4 kilómetros.
- 9) Para una muestra de 15 pares de observaciones, se conoce que el error estándar de estimación es  $S = 0.25$  y la Suma de Cuadrados Total de  $y$  es de  $S_{yy} = 5000$ . ¿Es alto el valor del coeficiente de determinación?
- 10) El coeficiente de correlación de una muestra de datos es  $r = -0.88$  y la SCT = 1320. Si el error estándar de estimación es de  $S = 13.28$ , ¿Cuántos pares de observaciones hay?
- 11) El gerente de ventas de La Buena Suerte está analizando por qué han disminuido las ventas de lapiceras Tony; para ello ha centrado su análisis en la variable precio de la marca competidora Deidei, sospechando que la caída de éstos en los últimos meses habrían afectado de manera significativa las ventas de la primera.  
Ha tomado para el análisis los datos de 15 semanas consideradas representativas.

Precio de Deidei	Ventas en miles de \$
3.30	190
3.28	160
3.29	170
3.25	160
3.20	150
3.18	140
3.17	130
3.20	120
3.18	125
3.16	120
3.15	110
3.15	105
3.10	100
3.00	90
3.00	85

Se pide determinar:

- a) La recta de regresión que explica las variaciones en las ventas de Tony ante variaciones en el precio de la marca Deidei.  
b) El coeficiente de determinación.  
c) A un nivel de significación del 5%, si un descenso en el precio de la marca Deidei explica un descenso en las ventas de Tony.  
d) Construya un intervalo de confianza del 90% para el coeficiente de regresión.
- 12) Una empresa de servicios públicos viene soportando en el último bimestre un número relevante de caídas de su sistema informático de atención al cliente.  
El personal de sistemas ha informado que paralelamente se han detectado ataques de hackers a dicho sistema y en la sospecha de que éstos pudieran ser causa de tales caídas, le solicitan a ud. que obtenga conclusiones al respecto.  
Para ello se ha relevado la información de las últimas diez semanas.

X : N° de ataques de hackers

Y : N° de caídas del sistema

X	Y
2	6
3	8
1	4
4	12
2	5
3	2
2	3
3	7
2	5
4	9

Se pide determinar:

- a) La recta de regresión que explica las variaciones en el número de caídas del sistema ante variaciones en el número de ataques de hackers. Explicar su significado.  
b) El coeficiente de determinación. Explicar su significado.  
c) A un nivel de significación del 5%, si un aumento en el número de ataques de hackers explica un aumento en el número de caídas del sistema.
- 13) El supermercado “Don Fulgencio” no tiene un número fijo de empleados asignados al delivery, por lo tanto continuamente recibe quejas telefónicas de sus clientes por las demoras en las entregas. Desea saber si es que realmente existe una relación entre estas dos variables.  
Para ello ha recopilado la información de los últimos 15 días corridos.



X : N° de empleados afectados a delivery por día  
Y : N° de quejas diarias relacionadas con el delivery

X	Y
5	15
4	20
5	12
3	28
2	35
1	40
4	10
0	25
3	32
2	18
0	45
5	8
7	1
4	12
5	12

Se pide determinar:

- La recta de regresión que explica las variaciones en el número de quejas diarias ante variaciones en el número de empleados afectados a delivery.
  - El coeficiente de determinación.
  - A un nivel de significación del 5%, si un aumento en el número de empleados afectados a delivery explica un descenso en el número de quejas diarias.
- 14) Un Banco privado desea establecer la relación existente entre el número de cajas habilitadas al público en hora pico y el número de clientes sentados en su salón a la espera de ser atendidos; todo esto con el objetivo de analizar la reducción del número de empleados de caja para afectarlos a otras funciones.  
Para ello lo ha contratado a ud de modo que efectúe los análisis pertinentes y le acerca los datos de las últimas quince jornadas.

X : N° de cajas habilitadas en hora pico  
Y : N° de clientes sentados en el salón del banco y a la espera de ser atendidos

X	Y
6	8
3	15
2	32
4	20
1	42

5	6
2	30
3	25
6	5
4	18
2	30
1	35
3	28
4	20
5	12

Se pide determinar:

- La recta de regresión que explica las variaciones en el número de personas a la espera de ser atendidos ante variaciones en el número de cajas habilitadas.
  - El coeficiente de determinación.
  - A un nivel de significación del 5%, si un aumento en el número de cajas habilitadas explica un descenso en el número de clientes a la espera de ser atendidos
- 15) La Compañía Financiera “Finado” conoce por experiencia que cuánto mayor es el monto que adeuda un cliente, más alta es la tasa de interés que le cobra. Pero a los fines de presentar un informe ante la autoridad de control, necesita contar con una herramienta estadística sólida que demuestre tal aseveración.  
Para ello solicita a su consultora que realice las mediciones pertinentes.  
Ud, como encargado del trabajo decide tomar los datos de quince cuentas seleccionadas al azar.

X : Monto de dinero adeudado

Y : Tasa efectiva mensual en porcentaje

X	Y
1000	3
1500	3.1
2000	3.2
2500	3.4
3000	3.6
3100	3.8
3300	4.0
3500	4.2
4000	4.3
4500	4.35
5000	4.37

5500	4.40
5700	4.41
5800	4.42
6000	4.45

Se pide determinar:

- a) La recta de regresión que explica las variaciones en la TEM ante variaciones en los montos de las deudas.
  - b) El coeficiente de determinación.
  - c) A un nivel de significación del 5%, si un aumento en el monto de deuda explica un aumento en la TEM.
- 16) La empresa “Turronegli”, líder en la fabricación de turronegli de maní desea saber si el precio del maní es determinante en el costo de fabricación del turronegli.  
Para ello ha solicitado a la cámara empresaria del sector que realice un relevamiento entre todas las industrias de la actividad.  
Dicha cámara, a través de una consultora externa relevó los datos de quince empresas que fabrican el mismo tipo de turronegli.

X : Precio de la Tn de maní en miles de pesos  
Y : Costo unitario de fabricación del turronegli en pesos

X	Y
3.2	1.7
3.3	1.72
3.4	1.75
3.2	1.73
3.1	1.65
3.15	1.69
3.22	1.71
3.23	1.74
3.25	1.80
3.18	1.70
3.12	1.60

3.15	1.65
3.14	1.61
3.14	1.63
3.15	1.67

Se pide determinar:

- a) La recta de regresión que explica las variaciones en el costo de fabricación ante variaciones en el precio del maní.
  - b) El coeficiente de determinación.
  - c) A un nivel de significación del 5%, si un aumento en el precio del maní explica un aumento en el costo de fabricación del turrón.
- 17) En la facultad de arquitectura de una universidad privada se desea estudiar el grado de inscripción que los alumnos que se reciben de arquitecto mantienen en el único posgrado que ofrece, al año siguiente de recibidos. Para ello se relevaron quince ciclos sucesivos de inscripción.

X : Número de egresados en el año t

Y : Número de inscriptos en posgrado en el año t+1, habiendo egresado en el año t

X	Y
38	10
64	23
72	25
28	5
39	12
42	12
40	11
52	15
55	8
68	19
72	32
83	28
39	42

42	24
45	20

Se pide determinar:

- La recta de regresión que explica las variaciones en el número de inscriptos a posgrado ante variaciones en el número de egresados como arquitecto.
  - El coeficiente de determinación.
  - A un nivel de significación del 5%, si un aumento en el número de egresados como arquitecto explica un aumento en el número de inscriptos a posgrado.
- 18) En una empresa de mensajería están analizando si el hecho de atender un mayor número de viajes diarios redundaría en un mayor número de motos que sufren problemas mecánicos por día, a fines de iniciar acciones preventivas. Para ello se han recolectado los datos de los últimos quince días.

X : Número total de viajes atendidos por día

Y : Número de motos con problemas mecánicos por día

X	Y
40	4
42	3
30	3
35	2
25	1
28	0
32	0
35	1
40	2
42	3

Se pide determinar:

- La recta de regresión que explica las variaciones en el número de motos con problemas mecánicos ante variaciones en el número de viajes por día.
- El coeficiente de determinación.
- A un nivel de significación del 5%, si un aumento en el número de viajes por día explica un aumento en el número de motos con problemas mecánicos.

### RESPUESTAS A LOS EJERCICIOS DEL TRABAJO PRACTICO N°3

#### Regresión Lineal Simple y Correlación

- 1)
  - a)  $Y = -9,0029 + 0,0081X$
  - b) Si, pues  $R^2 = 0,6384$
  - c) Aprox 40 accidentes
  - d)  $r = 0,799$
- 2)
  - a)  $Y = -9,9563 + 1,1512X$
  - b) 86,84%
  - c) Si, pues  $T_{calc} = 1,812$
  - d) IC = [105,98 ; 111,25] al 95%
- 3)
  - a)  $Y = -21,1990 + 0,2748X$
  - b) Si, pues  $R^2 = 0,9496$   $T_{calc} = 12,2709$   $T_c = 1,860$
  - c) IC = [4,92 ; 7,65] al 90%
- 4)
  - a)  $Y = 42,6383 + 103,5816X$
  - b) Si, pues  $R^2 = 0,9863$   $T_{calc} = 23,1225$   $T_c = 1,397$
  - c) IC = [93,25 ; 113,91] al 95%
- 5)
  - a)  $Y = 611,8974 + 13,6390X$
  - b) Si, pues  $R^2 = 0,7909$   $T_{calc} = 4,7644$   $T_c = 1,440$
  - c) IC = [604,79 ; 2528,46] al 95%
- 6)
  - a)  $Y = 862,99 - 23,16X$
  - b) 85,12%
  - c) IC = [588,76 ; 674,02] al 95%
- 7)  $r = 0,5574$ ;  $T_{calc} = 1,8989$   $T_c = 2,306$  (no existe relación entre las variables)
- 8)
  - a) No hay datos para el segundo criterio de validación.
  - b) 10 min y 13 segundos
- 9) Si,  $r^2 = 0,9998375$
- 10)  $n = 4$
- 11)
  - a)  $Y = -866,85 + 314,17X$
  - b)  $r^2 = 0,8564$
  - c)  $T_{calc} = 8,8036$   $T_c = 1,771$  **Rechazo  $H_0$**
  - d) IC = [250,97 ; 377,37] al 90%

12)

- a)  $Y = 0,4048 + 2,1905X$
- b)  $r^2 = 0,4982$
- c)  $T_{calc} = 2,8182$   $T_c = 1,860$  **Rechazo  $H_0$**

13)

- a)  $Y = 38,7349 - 5,3605X$
- b)  $r^2 = 0,7170$
- c)  $T_{calc} = -5,739$   $T_c = -1,771$  **Rechazo  $H_0$**

14)

- a)  $Y = 43,743 - 6,4734X$
- b)  $r^2 = 0,8978$
- c)  $T_{calc} = -10,69$   $T_c = -1,771$  **Rechazo  $H_0$**

15)

- a)  $Y = 2,76 + 0,000312X$
- b)  $r^2 = 0,8874$
- c)  $T_{calc} = 10,12$   $T_c = 1,771$  **Rechazo  $H_0$**

16)

- a)  $Y = 0,0053 + 0,5272X$
- b)  $r^2 = 0,5493$
- c)  $T_{calc} = 3,98$   $T_c = 1,771$  **Rechazo  $H_0$**

17)

- a)  $Y = 4,4031 + 0,2824X$
- b)  $r^2 = 0,2043$
- c)  $T_{calc} = 1,83$   $T_c = 1,771$  **Rechazo  $H_0$**

18)

- a)  $Y = -3,2575 + 0,1478X$
- b)  $r^2 = 0,4276$
- c)  $T_{calc} = 2,44$   $T_c = 1,860$  **Rechazo  $H_0$**





# **PRÁCTICO N° 4**

## **ESTADISTICA**

### **NO**

## **PARAMETRICA**



**Tema: Estadística no paramétrica**

**Sub-tema: Prueba de corridas**

**⇒ Ejercicios propuestos**

- 1) Un profesor prepara un examen con 25 preguntas de verdadero o falso. La secuencia de respuestas correspondiente al orden de las preguntas es la siguiente:

V V F F F V V V V F F F V V F F V F V F F F F V V

y desea saber con un 5% de nivel de significación, si los dos tipos de respuestas aparecen en forma aleatoria.

- 2) Dos personas discutían sobre si las caras de una moneda que poseían se presentaban al azar, o producto del desgaste esto no ocurría. Para comprobarlo, decidieron arrojar 14 veces la moneda y anotaron la siguiente secuencia, escribiendo una “C” cada vez que aparecía una cara y “X” cuando aparecía cruz.

C X X X C C C X X X C C C C

Con las herramientas no paramétricas que usted maneja, podría ayudarlos a dilucidar el problema. Considere  $\alpha=0.05$

- 3) Un lote de 20 productos, sometidos a un estricto control de calidad fueron calificados en una escala de 1 a 10 puntos, considerándose como aptos para la venta a aquellos de 7 o más puntos. Tenga en cuenta la siguiente información y determine, con un 95% de confianza, si los productos, aptos y no aptos se presentaron al azar.

Prod. N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Calif.	7	4	5	8	8	6	9	7	7	6	1	4	3	7	7	8	8	9	7	2

- 4) Una empresa que fabrica ciertos instrumentos de medición de alta precisión quiere evaluar los desajustes que se producen cuando envía dichos instrumentos desde la planta elaboradora hasta el lugar de venta. Para ello se enviaron 48 de tales instrumentos por tren, desde la planta que los fabrica hasta el punto de venta distante 120 Km. Al arribar al mismo se revisaron y se marcó (B) a aquellos que presentaban desajustes. Al resto se los marcó (A). Usando  $\alpha=0.05$  pruebe si existe aleatoriedad en la presencia de los desajustes provocados por el traslado.

A B A A A B B A B B A A B A B A A B B B A B B  
 A A A B A B A A B B A B B A A A B B A A B A A A



- 2) Se evaluó el peso en kg. de 10 personas sometidas a un control dietario durante una semana.

Persona N°	1	2	3	4	5	6	7	8	9	10
Peso inicial	74	74	73	71	78	70	74	75	78	79
Peso final	72	73	72	68	69	70	73	77	82	80

Al 5% de nivel de significación, ¿Piensa Usted que surtió efecto el control dietario como para determinar una reducción efectiva del peso?

- 3) Se encuestó a 140 personas sobre su preferencia a vivir en casa o en departamento. Del total de la muestra 80 respondieron que prefieren vivir en casa, 50 en departamento y el resto no se definió. Considerando  $\alpha=0.05$ ; ¿Existen diferencias significativas entre las preferencias por el tipo de vivienda?
- 4) Debido a los grandes cambios climáticos, se ha dicho muchas veces que las zonas subtropicales están cambiando y cada vez son menos lluviosas. Sin embargo algunos científicos no lo piensan así. Entre los estudios a realizar se decidió analizar la precipitación acumulada en los días de otoño y primavera de los años 1999 y 2000 para la ciudad de Cochabamba (Bolivia) de manera de tener registros de una ciudad con dicho clima y en meses donde las precipitaciones no son extremas.  
 Así, se seleccionaron al azar 15 fechas de otoño y primavera, consignándose los siguientes datos:

Fecha	5/4	21/4	8/5	11/6	15/6	20/6	25/9	30/9	5/10	11/10	8/11	23/11	6/12	10/12	18/12
1999	58	45	56	75	52	49	48	67	54	74	69	80	82	59	50
2000	57	70	46	67	60	47	45	71	48	79	60	74	79	72	50

Probar con un nivel de significación del 5% si el año 2000 fue menos lluvioso que el año 1999.

- 5) A una muestra aleatoria de 60 estudiantes de un instituto educativo se le aplicó un programa especial con el fin de mejorar la organización del tiempo dedicado al estudio y el aprovechamiento en el aprendizaje. Después de pasados dos meses de finalizar el programa se preguntó si lo aprendido en el mismo fue eficaz. Un total de 32 respondieron afirmativamente, 25 negativamente y el resto no emitió opinión. Usando  $\alpha = 0.05$  ¿Puede concluirse que el programa es efectivo?

- 6) Un deportista principiante de arquería probaba suerte en un club cuando se le acercó un experto y ofreció enseñarle en forma gratuita, indicándole que sólo le cobraría las clases si quedaba demostrado que el principiante mejoraría su destreza luego de las clases. Para ello el principiante realizó 13 tiros anotándose su puntuación en la siguiente tabla.

Tiro N°	1	2	3	4	5	6	7	8	9	10	11	12	13
Puntaje	3	4	2	1	5	4	2	2	3	3	1	4	2

Luego se sometió a las clases del experto durante una semana y repitió otros 13 intentos anotándose la siguiente puntuación.

Tiro N°	1	2	3	4	5	6	7	8	9	10	11	12	13
Puntaje	4	4	4	3	5	3	3	5	3	2	2	3	1

El experto asegura que debe cobrar las clases. El principiante piensa que no mejoró. Para mediar lo consultan a usted a fin de que realice la prueba. Utilice un nivel de significación del 3%.

**Sub-tema: Prueba de Wilcoxon**



**Ejercicios propuestos**

- 1) Se realizó un mismo test de cociente intelectual, dos veces a 20 individuos con una diferencia de media hora entre cada test. Se desea averiguar al 95% de confianza, si el 2° test arroja mejores resultados que el 1°(tomando como posible causa la repetición de los mismos ejercicios)

Persona N°	1	2	3	4	5	6	7	8	9	10
Test 1	101	115	96	93	80	102	95	125	130	111
Test 2	104	117	107	92	90	102	94	131	132	120

Persona N°	11	12	13	14	15	16	17	18	19	20
Test 1	110	125	104	110	104	106	100	107	122	114
Test 2	114	128	109	111	104	104	107	115	118	118

- 2) El siguiente cuadro muestra el número de chocolates de cierta marca vendidos en 10 kioscos, antes y después de colgar publicidad estática en sus frentes, durante una semana. Verifique al 95% de confianza si las ventas han mejorado con dicha publicidad.

Kiosco N°	1	2	3	4	5	6	7	8	9	10
Antes	20	10	18	30	20	22	40	45	20	23
Después	23	12	15	24	30	28	32	36	18	24

- 3) Se evaluó la velocidad de acceso al disco rígido (en milisegundos) de una PC antes y después de cambiar la memoria RAM por una más veloz. Se realizaron 9 intentos, en ambas oportunidades.

Verifique al 95% de confianza, si la nueva memoria RAM acelera o no el acceso al disco rígido.

Intento N°	1	2	3	4	5	6	7	8	9
Antes	40	42	43	40	41	45	42	40	43
Después	41	38	36	39	38	37	37	41	40

- 4) Hace un par de meses se instaló cerca de un barrio residencial un polígono de tiro. Las quejas de los vecinos por los continuos estampidos se han multiplicado y han llevado a la comuna a solicitar al polígono la construcción de muros más altos y paneles acústicos.

El polígono ha colocado temporariamente paneles acústicos y desea saber, antes de construir el costoso muro si las quejas han disminuido en los siguientes 13 días cotejándolos con los 13 días anteriores a la instalación de tales paneles. Efectúe una prueba y explique su resultado utilizando un nivel de significación del 2%.

Sin paneles acústicos	27	15	20	24	13	18	30	46	15	29	17	21	18
Con paneles acústicos	26	23	19	12	25	9	16	12	28	20	16	14	11

- 5) Se ha tomado el número de lesionados por tipo de lesión para el año 2002 de las Ligas Nacional de Basketball (LN) y del Torneo Nacional de Ascenso (TNA) obteniéndose los siguientes datos:

Lesión	LN	TNA
Hombro	43	19
Cuello	6	3
Tórax	4	2
Codo	25	23
Dedos mano	11	9
Muslo	10	7
Ingle	6	2
Rodilla	26	28
Espalda	15	12
Muñeca	8	0
Cadera	3	3
Mano	5	3
Tobillo	7	5
Pie	2	5
Dedos pie	0	1
otros	14	8

¿Se puede afirmar que en la LN hubo más lesionados que en el TNA durante 2002 con un nivel de significación del 5% ?



## Pruebas de Bondad de Ajuste

### Sub-tema: Prueba de Ji cuadrado ( $\chi^2$ )

#### Ejercicio resuelto

La Secretaría de Transporte de una ciudad desea observar la cantidad de vehículos que violan la luz roja en determinado cruce. Para ello ha instalado previamente un sensor bajo la cinta asfáltica. Se relevó en 500 oportunidades el mismo semáforo un mismo día de semana (Martes) y en hora pico, durante 5 semanas. Se obtuvo la siguiente información:

Cantidad de autos que violan la luz roja por semáforo	Frecuencia observada
0	11
1	40
2	72
3	88
4	102
5	85
6	45
7	32
8	18
9 y más	7

El estudio se encaró para saber con un nivel de significación del 5%, si la información recopilada proviene de una población que responde a una distribución de Poisson con media igual a 4 vehículos por semáforo.

### Solución

Cantidad autos que violan la luz roja por semáforo	Frecuencia observada $f_o$	Frecuencia relativa esperada $f_{re}$	Frecuencia esperada $F_e=n*f_{re}$	$\frac{(f_o - f_e)^2}{f_e}$
0	11	0.0183	9	0.44
1	40	0.0733	37	0.24
2	72	0.1465	73	0.01
3	88	0.1954	98	1.02
4	102	0.1954	98	0.16
5	85	0.1563	78	0.63
6	45	0.1042	52	0.94
7	32	0.0595	30	0.13
8	18	0.0298	15	0.60
9 y más	7	0.0213	11	1.45
Totales	500	1	500	$5.62 = \chi_e^2$

Ho: La distribución de Poisson es una buena descripción de la información.

Ha: La distribución de Poisson no es una buena descripción de la información.

Punto crítico:  $\chi^2_{10-1-1}(0.95)=15.51$

**Regla de decisión: Rechazar Ho para todo valor de  $\chi_e^2 \geq 15.51$ .**

**Conclusión: Como  $\chi_e^2 = 5.62 \Rightarrow$  No se rechaza Ho. Los datos provienen de una población que responde a una distribución de Poisson.**



### Ejercicios propuestos

- 1) En una línea de montaje, que trabaja al 100% de su capacidad, se espera que todas las máquinas entreguen la misma cantidad de unidades del producto por hora, según las especificaciones estándar, esto es 200 unidades del producto por máquina. En una hora seleccionada al azar se obtuvo la siguiente información:

Máquina	Cantidad de productos
A	180
B	250
C	150
D	220

Considerando un nivel de significación del 5% ¿Cree Ud. Que existan diferencias significativas entre la producción programada y la real?

- 2) Miguel Hernández, supervisor contable en Erlistán SA, quiere saber si el proceso contable de la empresa se puede simular usando una distribución de Poisson para describir la magnitud de errores. Para ello toma una muestra aleatoria de 400 cuentas de los registros de contabilidad y resume el n° de errores encontrados en cada cuenta en la siguiente tabla. Utilice  $\alpha = 0.05$

N° de errores	0	1	2	3	4 y más	Total
N° de cuentas	102	140	75	52	31	400

- 3) En un casino se controla periódicamente el estado de los dados para analizar su desgaste, el que no debe incidir en los resultados cuando se los utiliza. En este caso en particular se analizó un dado seleccionado al azar y se lo lanzó 300 veces, obteniéndose los siguientes datos:

Cara obtenida	Número de veces
As	52
2	51
3	50
4	48
5	49
6	50

¿Encuentra Ud. argumentos suficientes que hagan afirmar que el dado se ha desgastado como para que haya perdido la equidad en la aparición de cada una de las caras? Considere  $\alpha=0.05$

- 4) La empresa Aeropuertos 4000 quiere controlar los arribos de taxis por minuto al aeropuerto de Ezeiza para las horas pico, de manera de hacer un modelo matemático. Para un día seleccionado al azar anotó lo siguiente considerando como variable

$X_i$  : Número de arribos entre las 19:10 y las 20:00

Se pregunta si dichos datos responden a una distribución de Poisson con media igual a 6 taxis por minuto considerando un  $\alpha = 0.05$

$X_i$	1	2	3	4	5	6	7	8	9	10	11
$F_o$	5	3	2	6	6	2	6	10	4	4	2

- 5) El siguiente cuadro muestra el número de piezas defectuosas fabricadas por una determinada máquina. Mediante herramientas de análisis, indique si la distribución normal es una buena descripción de la información. Considere  $\alpha=0.05$

Número de piezas Defectuosas	Frecuencia observada
0-12	9
12-24	20
24-36	66
36-48	40
48-60	75
60-72	40

- 6) A los fines de iniciar un análisis del  $n^\circ$  de pacientes que se reciben en la guardia de un hospital a causa de accidentes de tránsito se han tomado los datos de los últimos 210 días. Se quiere probar si los mismos pueden ser analizados a través de una distribución de probabilidad de Poisson. Utilice para ello un nivel de significación del 5%.

N° de pacientes recibidos por accidente de tránsito	N° de días
0	25
1	55
2	65
3	35
4	20
5 y más	10
Total	210

- 7) Con el objeto de realizar determinadas acciones para mejorar la atención al público, la Dirección de Rentas de la Municipalidad desea determinar, en primer lugar, si el tiempo de espera de los contribuyentes en el lugar de pago se puede explicar por una distribución de Poisson. La información para determinarlo se presenta en el siguiente cuadro. Use un nivel de significación del 10 %.

Tiempo de espera (en minutos)	Cantidad de personas
0	400
1	250
2	150
3	75
4	25
5	10

**Sub-tema: Prueba de Kolmogorov-Smirnov**



**Ejercicio resuelto**

El mismo ejemplo de  $\chi^2$  resuelto por Kolmogorov.

Cantidad de autos que violan la luz roja por semáforo	Frecuencia observada $f_o$	Frecuencia relativa observada $f_{ro}$	Frecuencia relativa observada acumulada $F_o$	Frecuencia relativa esperada $f_{re}$	Frecuencia relativa esperada acumulada $F_e$	$ F_e - F_o $
0	10	0.0220	0.022	0.0183	0.0183	0.0037
1	40	0.0800	0.102	0.0733	0.0916	0.0104
2	72	0.1440	0.246	0.1465	0.2381	0.0079
3	88	0.1760	0.422	0.1954	0.4335	0.0115
4	102	0.2040	0.626	0.1954	0.6289	0.0029
5	85	0.1700	0.796	0.1563	0.7852	0.0108
6	45	0.0900	0.886	0.1042	0.8894	0.0034
7	32	0.0640	0.950	0.0595	0.9489	0.0011
8	18	0.0360	0.986	0.0298	0.9787	0.0073
9 y más	7	0.0140	1	0.0213	1	0
Totales	500	1		1		

$H_o$  : La distribución de Poisson con  $\lambda = 4$  es una buena descripción de la información.

$H_a$  : La distribución de Poisson con  $\lambda = 4$  no es una buena descripción de la información.

El estadístico  $D_n = 0.0115$  que es el máximo de los valores de la última columna de la derecha.

El punto crítico es, al 95%  $\frac{1.36}{\sqrt{500}} = 0.0608$

**La regla de decisión es rechazar  $H_o$  para todo valor de  $D_n \geq 0.0608$**

**Por lo tanto, como  $0.0115 < 0.0608 \Rightarrow$  No se rechaza  $H_o$**

**Conclusión: La distribución de Poisson con  $\lambda = 4$  es una buena descripción de la información.**



**Ejercicio resuelto**

- 1) Verifique si la distribución normal es una buena aproximación de los datos presentados en la siguiente serie. Considere  $\alpha=0.05$ . La variable X representa la calificación obtenida por un grupo de 25 alumnos de un mismo curso.

$X_i$	$f_o$
0-2	3
2-4	5
4-6	12
6-8	4
8-10	1

**Respuesta:  $P_c=0.27$   $D_n=0.22074$  No se rechaza  $H_0$**

- 2) Una empresa encomendó a una consultora la realización de un estudio de mercado sobre la cantidad de clientes que admitirían ser “target” de su producto en el primer mes, separados en franjas de edad. Asimismo la empresa relevó luego de dicho mes la cantidad de clientes de cada franja que obtuvo.  
Verifique al 95% de confianza si es que existen o no diferencias significativas entre la cantidad esperada de clientes y la cantidad real obtenida.

Franja De edad	Cantidad esperada	Cantidad real
15-20	10	7
20-25	21	16
25-30	26	33
30-35	29	34
35 y más	20	16

**Respuesta:  $P_c=0.1321$   $D_n=0.07546$  No se rechaza  $H_0$**

- 3) Resuelva el ejercicio N° 3 de  $\chi^2$  por Kolmogorov.

- 4) La siguiente serie presenta los datos de gastos en publicidad en miles de pesos por mes de las 125 sucursales de una empresa argentina en Sudamérica.  
¿Se ajustan bien estos datos por la función normal? Utilice  $\alpha = 0.10$

Gastos	fo	fe
25-30	9	6
31-36	22	17
37-42	25	32
43-48	30	35
49-54	21	18
55-60	12	13
61-66	6	4

**Respuesta:  $P_c = 0.1216$   $D_n = 0.064$  No se rechaza  $H_0$ .**

**Sub-tema: Prueba de Mann-Whitney**



**Ejercicio resuelto**

La variable X medida, es el tiempo de respuesta de dos robots similares ante una misma tarea, medido en minutos. Se solicitó la realización de dicha tarea en 6 oportunidades.

Robot A	1	3	2	1	2	4
Robot B	4	3	5	6	4	1

$H_0$ : No existe diferencia significativa entre el tiempo que tarda uno u otro modelo de robot.

$H_a$ : Existe diferencia

**Solución**

X	Robot	Rango	Ra	Rb
1	A	2	2	
1	A	2	2	
1	B	2		2
2	A	4.5	4.5	
2	A	4.5	4.5	
3	A	6.5	6.5	
3	B	6.5		6.5
4	A	9	9	
4	B	9		9
4	B	9		9
5	B	11.5		11.5
6	B	11.5		11.5
Sumas			28.5	49.5

$$\begin{aligned}
 n_1 &= 6 & U &= 6*6 + \frac{6*7}{2} - 28.5 = 28.5 & \sigma_U &= \sqrt{\frac{6*6*(6+6+1)}{12}} = 6.245 \\
 n_2 &= 6 & m_U &= \frac{6*6}{2} = 18 & Z_E &= \frac{28.5 - 18 + 0.5}{6.245} = 1.76 \\
 R_1 &= 28.5 & & & & \\
 R_2 &= 49.5 & & & & 
 \end{aligned}$$

**Punto crítico al 0.95, tabla de distribución normal es 1.96 .Por lo tanto, no se rechaza Ho, luego se puede inferir que no existen diferencias significativas entre el tiempo que tarda el robot A y el robot B para realizar la misma tarea.**



**Ejercicios propuestos**

- 1) La empresa “La Buena Suerte S.A.” proporcionó una determinada capacitación, como parte de su plan de mejoras para su empresa, a la mitad de sus vendedores. Para analizar los efectos de dicha capacitación reunió los datos sobre las ventas de la empresa durante 15 semanas posteriores a la misma. De acuerdo a tales datos y trabajando con un nivel de significación del 5% determine si hay diferencias significativas entre el número de ventas concretadas por los vendedores que recibieron la capacitación (Ventas A) para efectuar sus tareas y los que no la recibieron (Ventas B), que avalen la continuidad de dicha capacitación.

Ventas A	20	30	45	18	15	22	30	25	23	17	8	12	31	25	23
Ventas B	23	31	29	20	19	22	25	22	45	15	15	30	12	8	45



- 2) La empresa “Redondo y Hnos.” quiso averiguar si había diferencias significativas entre los montos de venta (en miles de \$), durante 12 meses para los dos tipos de ventas que utiliza. Utilice un nivel de significación del 10% para realizar la prueba.

Venta telefónica	5	2	3	8	4	6	7	9	2	2	3	4
Venta por correo	1	3	2	2	3	3	4	5	2	3	3	3

- 3) Un renombrado hospital privado contrata a sus médicos solamente si son egresados de dos de las universidades más importantes del país. Durante el año anterior decidieron además analizar si había diferencias en el nivel de tales universidades y para ello tomaron un examen a los médicos recién graduados que se postularon para trabajar en el hospital. Se obtuvieron los datos de la tabla siguiente que son las calificaciones obtenidas por los examinados en una escala del 0 al 100. Realice la prueba con un nivel de significación del 6%.

Universidad A	97	69	73	84	76	92	90	88	84	87	93	--	--
Universidad B	88	99	65	69	97	84	85	89	91	90	87	91	72

- 4) La empresa “Rivara & Prieto Ltd.” enfrenta una situación financiera no muy favorable. Como primera medida efectuó una reestructuración en el área del personal de ventas, reduciendo el plantel de vendedores a la mitad. Reuniendo la información de ventas (en miles de pesos) correspondientes a 10 meses antes de la reducción de la cantidad de vendedores y la correspondiente a los 10 meses posteriores, la empresa desea probar si los efectos de tal reducción se materializaron en diferencias significativas en las ventas que le indiquen la necesidad de rever la medida.

Trabaje con un nivel de significación del 5%.

Ventas antes de la reducción	2.5	2.3	2	2.8	2.4	1.8	1.6	1.4	1.2	1.4
Ventas luego de la reducción	1.8	1.6	1.6	1.8	1.2	1.2	1	0.8	1	1

- 5) Luis Zapia, presidente de la empresa elaboradora de jugos de frutas "La naranja madura" pide al estadístico de la empresa que analice la efectividad de las muestras gratis. El analista releva datos de ventas en días normales y en días en los que se entregan muestras gratis en los supermercados que venden los jugos. Los resultados en cientos de unidades son los siguientes. Pruebe con un nivel de significación del 5 % si en los días de muestras gratis se incrementan las ventas.

Ventas en días normales	4	5	4,4	6	4	2,7	4	4,6	5	3,5	3	3
Ventas en días de entrega de muestras gratis	4	5	7	7	4,5	5,5	6	6,4	4	7	6	5

**Sub-tema: Prueba de Kruskal-Wallis**



**Ejercicio resuelto**

La variable “X” medida es el tiempo total de decantación de un químico en un proceso industrial de fabricación de telas por 3 métodos diferentes (medido en minutos). Se llevaron a cabo 7 observaciones para la aplicación de cada método, siendo el resultado expresado en la tabla siguiente. El ingeniero químico desea saber si existen o no diferencias significativas entre los 3 métodos.

Método A	2	3	2	2	1	1.5	3
Método B	4	3	3	4	3	3	4
Método C	0.8	1	1	0.7	0.5	1	1.2

Ho: No existen diferencias significativas entre los 3 métodos.

Ha: alguno al menos es diferente

**Solución**

Orden	Método	Rango	Ra	Rb	Rc
0.5	C	1			1
0.7	C	2			2
0.8	C	3			3
1	C	5.5			5.5
1	C	5.5			5.5
1	C	5.5			5.5
1	A	5.5	5.5		
1.2	C	8			8
1.5	A	9	9		
2	A	11	11		
2	A	11	11		
2	A	11	11		
3	A	15.5	15.5		
3	A	15.5	15.5		
3	B	15.5		15.5	
3	B	15.5		15.5	
3	B	15.5		15.5	
3	B	15.5		15.5	
4	B	20		20	
4	B	20		20	
4	B	20		20	
Sumas			78.5	122	30.5

$$K = \frac{12}{21 \times (21+1)} \times \left[ \frac{78.5^2}{7} + \frac{122^2}{7} + \frac{30.5^2}{7} \right] - 3 \times (21+1) = 15.55$$

**Punto crítico de  $\chi^2_{(3-1)}=5.99$**

**Se rechaza la Ho. Conclusión: Hay evidencia significativa para rechazar la Ho, por lo tanto se puede inferir que existen diferencias entre los métodos.**



**Ejercicios propuestos**

- 1) Este cuadro muestra el número de ventas concretadas (en cien unidades) por una casa de electrodomésticos, según modalidad de pago. La empresa propietaria desea saber con un nivel de riesgo de significación del 5%, si existen diferencias significativas entre el número de ventas concretadas y la modalidad de pago. La información fue relevada durante los últimos 6 meses.

Meses	1	2	3	4	5	6
Forma de pago						
Efectivo	8	7	6	7	8	5
A Plazo	6	8	7	6	9	8
Con tarjeta	12	10	9	7	7	8

- 2) La empresa “Simondet y Hnos.”, busca determinar, al 5% de significación si existen o no diferencias significativas entre los kilogramos de materia prima que deben reponerse en las 4 plantas productoras que posee. Para ello relevó dichos datos durante 8 meses seguidos, lo cual se volcó en la tabla que sigue.

Meses	1	2	3	4	5	6	7	8
Planta								
A	200	250	240	205	220	215	223	218
B	195	192	190	190	175	200	180	190
C	205	250	195	190	190	240	200	218
D	215	175	190	220	190	223	180	192

- 3) El siguiente cuadro muestra, en minutos, el número de espacios publicitarios contratados en horario pico por un canal de cable en sus 3 señales durante 7 semanas. Verifique al 5% de significación si es que existen o no diferencias significativas entre el número de minutos contratados por señal.

Semanas	1	2	3	4	5	6	7
Señal							
X	35	30	32	34	28	26	35
Y	20	22	25	25	20	18	25
Z	30	30	35	32	34	35	25

- 4) Muestras independientes de tres marcas distintas de tubos para hornos de microondas se sometieron a pruebas de esfuerzo y se registró la cantidad de horas que cada uno operó sin necesidad de alguna reparación. Realice la prueba con un nivel de significación del 5% para determinar si se justifica la conclusión de que no hay diferencias en la duración según las distintas marcas.

Marca 1	Marca 2	Marca 3
46	48	70
50	30	35
8	60	120
60	10	60
50	40	40

- 5) La pescadería “El Salmón” posee 3 sucursales en los shoppings de la ciudad. A los fines de iniciar una investigación acerca de las ventas que realizan decidió encarar una prueba sobre la afluencia de clientes. Por lo tanto se registró la cantidad de clientes que diariamente efectuaron alguna compra durante un período de 10 días en las 3 sucursales. Al 5%, ¿ puede Usted afirmar que las sucursales no tienen diferencias en la cantidad de clientes diarios?.

Shopping P	99	64	101	85	79	88	97	95	90	100
Shopping L	83	102	125	61	91	96	94	89	93	75
Shopping A	89	98	56	105	87	90	87	101	76	89

- 6) En la siguiente tabla se muestran los precios minoristas de tres marcas del mismo modelo de impresora color. Determinar si existen diferencias significativas entre los precios de estas marcas considerando un  $\alpha = 0.01$

Marca A	89	90	92	81	76	88	85	95	97	86	100
Marca B	78	93	81	87	89	71	90	96	82	85	---
Marca C	80	88	86	85	79	80	84	85	90	92	---

**Sub-tema: Tablas de contingencia**



**Ejercicio resuelto**

El director de un diario de la ciudad capital analiza la relación entre la edad de sus lectores y la sección que leen primero. Para una muestra aleatoria de lectores se obtuvo la siguiente información:

Edad (en años)	Generales	Deportes	Humor
Menos de 20	120	124	100
21 - 40	170	112	88
41 y más	165	90	90

Al nivel de significación del 5%, ¿se puede concluir que existe una relación entre la edad y la sección del diario que lee primero?

**Solución**

Fo	Fe	Fo	Fe	Fo	Fe	Totales
120	148	124	106	100	90	344
170	159	112	114	88	97	370
165	148	90	106	90	91	345
455		326		278		1059

$\chi_e^2 = 15.47$   $\chi^2_{(3-1)(3-1)} = \chi^2_{(0.95)} = 9.49$  **Se rechaza Ho. Existe relación entre las edades de los lectores y la sección leída.**



**Ejercicios propuestos**

- 1) Una investigación referente a la relación entre la ubicación en la empresa La Luna SA y el nivel de presión que el personal siente respecto de su trabajo obtuvo la siguiente información de una muestra aleatoria Al nivel de significación del 5%. ¿Existe alguna relación entre ambas variables.

Gerencia	Nivel de presión		
	Bajo	Medio	Alto
Administrativo-Financiera	22	20	24
De Recursos Humanos	48	44	42
De Ventas	58	64	60
De RelacionesPúblicas	35	45	45

- 2) ¿La propensión a accidentes en el trabajo de los obreros depende del tiempo que ha estado trabajando ese día cada obrero? Un análisis efectuado en la empresa Modessi SA de 500 accidentes dio los resultados que se presentan en la tabla siguiente para la distribución de lesiones en 8 períodos de una hora de cada turno.

Hora de cada turno	1	2	3	4	5	6	7	8
Nº de accidentes	80	70	75	88	98	100	72	105

- a) ¿Es mayor la probabilidad la probabilidad de que un obrero sufra un accidente es mayor en algunos turnos que en otros?. Pruebe con nivel de significación del 10%.
- b) ¿Proporcionan los datos pruebas suficientes que indiquen que la probabilidad de un accidente durante las últimas cuatro horas del turno es mayor que durante las primeras cuatro horas?. Pruebe con nivel de significación del 10%.
- 3) Una encuesta de opinión anual que tiene como objetivo analizar un conjunto amplio de variables sobre la población de una ciudad dio los siguientes resultados respecto del nivel de vida de las familias.

Período	Estándar de vida			Total
	Muy bueno	Bueno	Regular	
1998	70	150	80	300
1999	70	130	100	300
2000	50	100	50	200
2001	40	110	50	200

¿Hay evidencia significativa que las proporciones de familias dentro de cada categoría de nivel de vida son las mismas para cada uno de los cuatro períodos? Utilice un nivel de significación del 5%.

- 4) El gerente de una compañía de seguros considera que los conductores de automóviles que usan celular tienen más accidentes y por lo tanto deberían pagar más por el seguro de sus autos. Para analizar el tema reunió los datos que se presentan en la siguiente tabla de accidentes de los últimos tres años. Utilice  $\alpha=0.05$ .

Celular	Con accidentes	Sin accidentes
Posee	95	300
No poseen	65	325

- 5) La empresa de publicidad Metropolitana SA investiga la relación entre el tipo favorito de mensaje comercial y el nivel de ingresos de consumidores potenciales. A una muestra aleatoria de consumidores potenciales le solicitó su opinión. Los datos obtenidos son:

Ingreso	Comercial favorito		
	A	B	C
Bajo	25	40	70
Medio	30	30	30
Alto	45	20	10

¿Tendrá necesidad la empresa de adecuar el tipo de mensajes según los niveles de ingresos de los consumidores? ¿Por qué? Use un nivel de significación del 5 %.

- 6) El gerente de ventas de una franquicia de heladerías está preocupado puesto que la misma podría estar mal distribuida en las cuatro ciudades importantes del país. Se muestreó a 100 potenciales clientes a los fines de recabar su opinión para probar si hay relación entre la ubicación y demanda de adquisición. Para la realización de la prueba considere un  $\alpha = 0.05$ .

	Capital	Rosario	Córdoba	Mendoza
Adquirirían	40	55	45	50
No adquirirían	60	45	55	50

- 7) Para ver si la tendencia del MerVal es independiente del volumen de acciones negociadas en la Bolsa de Bs. As. Se han recogido datos correspondientes a una semana y clasificado previamente el volumen en 4 categorías. Utilice  $\alpha = 0.10$

Volumen negociado	Merval		
	Subió	S/C	Bajó
Muy alto	20	7	3
Alto	30	40	30
Medio	20	8	2
Bajo	30	5	5

- 8) Un reconocido diario trata de determinar qué tipo de público lo lee. Analiza para ello una encuesta realizada a 269 personas clasificadas por nivel educativo y frecuencia semanal de lectura. Establezca si son independientes ambas categorías considerando un  $\alpha = 0.10$

Frec. de lectura	Nivel Educativo			
	Universitario Completo	Secundario Completo	Primario Completo	Primario Incompleto
Nunca	10	17	11	21
A veces	12	23	8	5
Matutino sólo y a diario	35	38	16	7
Ambas ediciones a diario	28	19	6	13

**Sub-tema: Coeficiente de correlación de Spearman**



**Ejercicio resuelto**

- 1) La empresa “Darth Deimon inc.” consultó a 20 personas sobre la calificación que le otorgarían a su producto y sobre la habitualidad en el consumo de dicha producto, para determinar si existe correlación directa entre consumo y opinión de su producto, con un 95% de confianza. Para ello construyó las siguientes escalas de referencia.



Opinión	Frecuencia de consumo
1: Malo	1: Nunca
2: Regular	2: Poca
3: Bueno	3: Casi siempre
4: Muy bueno	4: Siempre

**Solución**

Ho : No existe correlación entre las variables opinión y frecuencia de consumo.

Ha : Existe correlación directa entre las variables opinión y frecuencia de consumo.

Encuestado N°	Opinión	Consumo	Di <sup>2</sup>
1	1	1	0
2	1	1	0
3	1	2	1
4	1	2	1
5	1	3	4
6	2	1	1
7	2	2	0
8	2	2	0
9	2	1	1
10	2	1	1
11	3	2	1
12	3	3	0
13	3	2	1
14	3	3	0
15	3	4	1
16	3	4	1
17	4	3	1
18	4	3	1
19	4	4	0
20	4	4	0
Suma			15

$$r_s = 1 - \frac{6 * 15}{20 * (20^2 - 1)} = 0.9887$$

Punto crítico (Pc) = 0.4451

**Conclusión:** Se rechaza Ho, por lo tanto existe correlación directa entre opinión y consumo.  
 A mejor opinión, mayor frecuencia de consumo.



**Ejercicios propuestos**

- 1) El siguiente cuadro muestra la ubicación de 10 equipos de fútbol de la 1° D durante los campeonatos de 1988 y 1989 en Argentina. ¿Existe correlación al 95%?

Año	1990	1991
Equipo		
Lugano	12	10
Muñiz	14	9
Yupanqui	3	2
San Carlos	4	1
Fénix	8	13
J.J.Urquiza	5	11
Claypole	16	7
Fc. Urquiza	9	8
Midland	7	6
San Martín	6	5

- 2) En un instituto de enseñanza de la computación se realizó un seguimiento a 8 alumnos que cursaron los 2 módulos que allí se dictan. Se tomó nota de sus puntajes y se desea saber al 95% de seguridad, si existe correlación entre ambos puntajes.

Alumno	Módulo A	Módulo B
A	8	9
B	9	7
C	10	9
D	9	7
E	8	7
F	7	8
G	8	7
H	7	9

- 3) En el mismo instituto anterior se relevó también las horas de estudio semanales y la calificación obtenida por 5 alumnos que cursaron íntegramente un seminario de Internet. Se desea saber si existe correlación directa al 95% de confianza.

Hs. De estudio	Calificación
1	4
2	2
4	7
7	5
10	9

- 4) Un grupo de consumidores probó 9 marcas de hornos a microondas para verificar su calidad total. Los rangos asignados por el grupo y los precios de distribución sugeridos son los siguientes:

Fabricantes	Rango del panel	Precio sugerido (en pesos)
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 del microondas? Utilice un nivel de significación del 5%.

- 5) Se desea averiguar si entre los 8 empleados de una Pyme existe alguna correlación entre las horas extra trabajadas en el último mes y el sueldo neto percibido en dichos meses. Considere  $\alpha = 0.02$

N° de Hs. Extras	40	60	25	35	32	52	20	45
Sueldo Neto en \$	300	500	450	350	800	850	450	400

- 6) Se desea averiguar si existe correlación inversa entre los gastos (en miles de \$) en materia de seguridad y el número de accidentes producidos durante el último año en 11 compañías de altos hornos. Considere  $\alpha = 0.01$

Cía.	A	B	C	D	E	F	G	H	I	J	K
Gtos.	60	37	30	20	24	42	39	54	48	58	26
Accidentes	2	7	6	9	7	4	8	2	4	3	8

- 7) Con la finalidad de verificar la calidad total de 10 marcas de multiprocesadoras de alimentos, se les solicitó a 10 consumidores la prueba de multiprocesadoras de tales marcas. El rango asignado por el grupo de consumidores y los precios de distribución sugeridos son los siguientes:

Marca	Calificación	Precio sugerido (\$)
A	6	270
B	8	250
C	3	380
D	8	245
E	5	290
F	2	480
G	6	265
H	5	285
I	4	360
J	1	600

¿Existe una relación significativa entre la calidad y el precio de la multiprocesadora? Utilice un nivel de significación del 5%.

### **Sub-tema: Coeficiente Tau de Kendall**



#### **Ejercicio resuelto**

En un concurso de belleza, hubo 2 jurados que asignaron puntajes a las 8 participantes en una escala de 0 a 100. Verificar al 90% de seguridad si existe correlación entre los puntajes asignados por ambos jurados.

Participante	Calificación del Jurado "M"	Calificación del jurado "N"
A	98	100
B	100	98
C	95	94
D	92	93
E	94	89
F	85	90
G	88	87
H	90	95

#### **Solución**

Ho: No existe asociación entre los rangos de las variables estudiadas.

Ha: Existe asociación directa entre los rangos de las variables estudiadas.

Participante	Calificación del Jurado "M"	Calificación del jurado "N"	Rm	Rn
A	98	100	7	8
B	100	98	8	7
C	95	94	6	5
D	92	93	4	4
E	94	89	5	2
F	85	90	1	3
G	88	87	2	1
H	90	95	3	6

Ordenamiento de los rangos, por el orden natural de M

Partic.	Rm	Rn
F	1	3
G	2	1
H	3	6
D	4	4
E	5	2
C	6	5
A	7	8
B	8	7

$$U=5+6+2+3+3+2+0+0=21$$

$$\text{Total de pares ordenados} = \frac{8 \times (8-1)}{2} = 28$$

$$V = \text{Total de pares} - U = 7$$

$$\tau = \frac{U - V}{U + V} = 0.5 \Rightarrow \text{Asociación directa.}$$

Por tabla  $n=8$  y  $\tau=0.50$  la  $P=0.054$

**Por lo tanto con  $\alpha=0.10$  se rechaza la  $H_0$  y existe asociación directa.**

## RESPUESTAS A LOS EJERCICIOS DEL TRABAJO PRACTICO N°4

### Estadística no paramétrica

#### *Prueba de Corridas*

- 1)  $Z_{\text{calc}} = -0,81$   $Z_c = \pm 1,960$  **No Rechazo Ho**
- 2) Límites 3 y 12  $r = 5$  **No Rechazo Ho**
- 3) Límites 6 y 16  $r = 8$  **No Rechazo Ho**
- 4)  $Z_{\text{calc}} = 0,7884$   $Z_c = \pm 1,960$  **No Rechazo Ho**
- 5) Límites 4 y 14  $r = 8$  **No Rechazo Ho**
- 6)  $Z_{\text{calc}} = 0,0593$   $Z_c = \pm 1,645$  **No Rechazo Ho**
- 7)  $Z_{\text{calc}} = -0,8190$   $Z_c = \pm 1,960$  **No Rechazo Ho**

#### *Prueba del signo*

- 1) Si.  $P = 0,0547 * 2 = 0,1094$  (éxito: reducción del colesterol) **No Rechazo Ho**
- 2) No.  $P = 0,2539$  (éxito: reducción del peso) **No Rechazo Ho**
- 3) Si.  $Z_{\text{calc}} = 0,2,7189$   $Z_c = \pm 1,960$  (éxito: vivir en casa) **Rechazo Ho**
- 4)  $P = 0.2120$  **No Rechazo Ho**
- 5) No.  $Z_{\text{calc}} = 1,06$   $Z_c = 1,645$  **No Rechazo Ho**
- 6)  $P = 0,3770$  **No Rechazo Ho**

#### *Prueba de Wilcoxon*

- 1) Límites 47 y 124;  $T_+ = 152$   $T_- = 19$  **Rechazo Ho**
- 2) Límites 10 y 45;  $T_+ = 24,5$   $T_- = 30,5$  **No Rechazo Ho**
- 3) Límites 8 y 37;  $T_+ = 41$   $T_- = 4$  **Rechazo Ho**
- 4) Límites 16 y 75;  $T_+ = 64,5$   $T_- = 26,5$  **No Rechazo Ho**
- 5) Límites 30 y 90;  $T_+ = 15$   $T_- = 105$  **Rechazo Ho**

#### *Prueba de Ji cuadrado*

- 1)  $Q = 29$   $\chi_c^2 = 7,815$  **Rechazo Ho**
- 2)  $Q = 91,175$   $\chi_c^2 = 9,488$  **Rechazo Ho**
- 3)  $Q = 0,20$   $\chi_c^2 = 11,07$  **No Rechazo Ho**
- 4)  $Q = 13.35$   $\chi_c^2 = 11.07$  **Se rechaza Ho.**
- 5)  $Q = 31.15$   $\chi_c^2 = 7.81$  **Se rechaza Ho.**
- 6)  $Q = 2.03$   $\chi_c^2 = 9.48$  **No se rechaza Ho**
- 7)  $Q = 64.46$   $\chi_c^2 = 6.251$  **Se rechaza Ho**

### ***Prueba de Mann-Whitney***

- 1)  $Z_{\text{calc}} = 0,083$   $Z_c = \pm 1,960$  **No Rechazo Ho**
- 2)  $Z_{\text{calc}} = -1,70$   $Z_c = \pm 1,645$  **Rechazo Ho**
- 3)  $Z_{\text{calc}} = 0,2027$   $Z_c = \pm 1,881$  **No Rechazo Ho**
- 4)  $Z_{\text{calc}} = -2,46$   $Z_c = \pm 1,960$  **Rechazo Ho**
- 5)  $Z_{\text{calc}} = 2,86$   $Z_c = 1,645$  **Rechazo Ho**

### ***Prueba de Krustal Wallis***

- 1)  $K = 3,66$   $\chi_c^2 = 5,991$  **No Rechazo Ho**
- 2)  $K = 12,03$   $\chi_c^2 = 7,815$  **Rechazo Ho**
- 3)  $K = 12,56$   $\chi_c^2 = 5,991$  **Rechazo Ho**
- 4)  $K = 2,02$   $\chi_c^2 = 5,991$  **No Rechazo Ho**
- 5)  $K = 0,196$   $\chi_c^2 = 5,991$  **No Rechazo Ho**
- 6)  $K = 2,58$   $\chi_c^2 = 9,21$  **No Rechazo Ho**

### ***Tablas de contingencia***

- 1)  $Q = 2,38$   $\chi_c^2 = 12,592$  **No Rechazo Ho**
- 2)  $Q = 7,18$   $\chi_c^2 = 12,592$  **No Rechazo Ho**
- 3)  $Q = 9,22$   $\chi_c^2 = 12,592$  **No Rechazo Ho**
- 4)  $Q = 6,59$   $\chi_c^2 = 3,841$  **Rechazo Ho**
- 5)  $Q = 45,41$   $\chi_c^2 = 9,488$  **Rechazo Ho**
- 6)  $Q = 5,01$   $\chi_c^2 = 12,592$  **No Rechazo Ho**
- 7)  $Q = 34,59$   $\chi_c^2 = 10,645$  **Rechazo Ho**
- 8)  $Q = 32,85$   $\chi_c^2 = 14,684$  **Rechazo Ho**

### ***Coefficiente de correlación de Spearman***

- 1)  $r_s = 0,4787$   $P_c = \pm 0,6364$  **No Rechazo Ho**
- 2)  $r_s = -0,0595$   $P_c = \pm 0,7143$  **No Rechazo Ho**
- 3)  $r_s = 0,8$   $P_c = 0,8$  **No Rechazo Ho (depende del criterio)**
- 4)  $r_s = -0,47$   $P_c = \pm 0,6833$  **No Rechazo Ho**
- 5)  $r_s = 0,1845$   $P_c = \pm 0,8095$  **No Rechazo Ho**
- 6)  $r_s = -0,8591$   $P_c = 0,7$  **Rechazo Ho**
- 7)  $r_s = -0,9424$   $P_c = \pm 0,6364$  **Rechazo Ho**



# **PRÁCTICO N° 5**

## **DECISION**

## **ESTADISTICA**

## **BAYESIANA**



## **Tema: Decisión Bayesiana**

**Teorema de Bayes:** “Si la realización de un acontecimiento aleatorio  $A$  depende necesariamente que se produzca uno de los acontecimientos excluyentes  $B_1, B_2, \dots, B_n$  y se sabe que  $A$  se ha cumplido, la probabilidad que sea uno determinado, por ejemplo  $B_i$ , el que se haya cumplido conjuntamente con  $A$ , está dada por”:

$$P(B_i / A) = \frac{P(B_i) \cdot P(A / B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A / B_i)}$$

donde  $P(B_i)$  son las probabilidades a priori. Representan las probabilidades de los estados del suceso antes de tomar información adicional.

$P(A/B_i)$  es la verosimilitud de la hipótesis. Representa la probabilidad condicional de observar la información adicional  $A$  cuando el estado del suceso que se presente sea  $B_i$ .

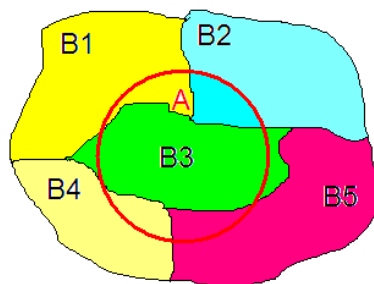
$P(B_i/A)$  es la probabilidad a posteriori. Representa la probabilidad de ocurrencia del estado del suceso  $B_i$  dada la información adicional  $A$ .

Desde otra óptica también se puede enunciar el Teorema del siguiente modo:

“Sea  $B = \{B_1, B_2, \dots, B_n\}$  un conjunto de “ $n$ ” sucesos mutuamente excluyentes y colectivamente exhaustivos y  $A$  un suceso particular compatible con los sucesos del conjunto  $B$ , entonces:

$$P(B_i / A) = \frac{P(B_i) \cdot P(A / B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A / B_i)}$$

Gráficamente



La información a priori puede ser objetiva o subjetiva. La información adicional se puede adquirir o elaborar por cuenta propia. Esta información adicional se utiliza para revisar las probabilidades a priori.

Un ejemplo concreto:

Una empresa fabrica tornillos a través de 3 máquinas iguales (A,B,C)

El 60% de la producción es fabricado por la máquina A, el 30% por la B y el resto por la C. Asimismo se sabe que la proporción de tornillos defectuosos es del 5% para la máquina A, del 3% para la B y del 2% para la C.

Se tiene un tornillo fallado a la vista y se desea saber cuál es la probabilidad de que haya sido fabricado por cada una de las máquinas.

Prob. a priori.

$$P(A) = 0,6$$

$$P(B) = 0,3$$

$$P(C) = 0,1$$

Verosimilitud de la hipótesis (Prob. condicionales)

$$P(\text{def} / A) = 0,05$$

$$P(\text{def} / B) = 0,03$$

$$P(\text{def} / C) = 0,02$$

Prob. conjuntas

$$P(A \text{ y def}) = P(A) * P(\text{def} / A) = 0,6 * 0,05 = 0,03$$

$$P(B \text{ y def}) = P(B) * P(\text{def} / B) = 0,3 * 0,03 = 0,009$$

$$P(C \text{ y def}) = P(C) * P(\text{def} / C) = 0,1 * 0,02 = 0,002$$

Por lo tanto

$$P(A \text{ y no def}) = P(A) * P(\text{no def} / A) = 0,6 * 0,95 = 0,57$$

$$P(B \text{ y no def}) = P(B) * P(\text{no def} / B) = 0,3 * 0,97 = 0,291$$

$$P(C \text{ y no def}) = P(C) * P(\text{no def} / C) = 0,1 * 0,98 = 0,098$$

Prob. totales

$$P(\text{def}) = P(A \text{ y def}) + P(B \text{ y def}) + P(C \text{ y def}) = 0,03+0,009+0,002 = 0,041$$

$$P(\text{no def}) = P(A \text{ y no def}) + P(B \text{ y no def}) + P(C \text{ y no def}) = 0,57+0,291+0,098 = 0,959$$

Prob. a posteriori

$$P(A / \text{def}) = 0,03 / 0,041 = 0,73171$$

$$P(B / \text{def}) = 0,009 / 0,041 = 0,21951$$

$$P(C / \text{def}) = 0,002 / 0,041 = 0,04878$$

$$P(A / \text{no def}) = 0,57 / 0,959 = 0,59437$$

$$P(B / \text{no def}) = 0,291 / 0,959 = 0,30344$$

$$P(C / \text{no def}) = 0,098 / 0,959 = 0,10219$$

Más fácil de ver es aún si volcamos los valores en una tabla como sigue:

Si se observó un tornillo defectuoso

	Prob. a priori	Prob. condicional	Prob. conjuntas	Prob. A posteriori
A	0,6	0,05	0,03	<b>0,73171</b>
B	0,3	0,03	0,009	<b>0,21951</b>
C	0,1	0,02	0,002	<b>0,04878</b>
Sumas	1		0,041	1

Si se observó un tornillo no defectuoso

	Prob. a priori	Prob. condicional	Prob. conjuntas	Prob. A posteriori
A	0,6	0,95	0,57	<b>0,59437</b>
B	0,3	0,97	0,291	<b>0,30344</b>
C	0,1	0,98	0,098	<b>0,10219</b>
Sumas	1		0,959	1



### Ejercicios propuestos sobre Teorema de Bayes

#### **DECISIÓN BAJO CONDICIONES DE RIESGO Estrategias Bayesianas. Árboles de decisión.**

(EN EL ARCHIVO EXCEL RESBAYES.XLS QUE ESTA EN LA PAGINA WEB TIENEN RESUELTO EL EJERCICIO N°4 DE ESTA GUIA)

1) Para la siguiente matriz de beneficios con datos de probabilidad:

P(F)	0,1	0,6	0,2	0,1
	F1	F2	F3	F4
A1	50	25	5	20
A2	20	-4	60	10
A3	60	15	-20	52

a) Indique la estrategia óptima mediante el criterio del VME.

b) Calcule el VEIP.

a) A1 con VME= 23; b) 15,2.

2) En un edificio de propiedad horizontal, los propietarios han verificado maniobras dolosas por parte de la Administración. Han pagado muchas facturas elevadas por trabajos de plomería pero las cinco últimas son realmente increíbles. El administrador sustentó una sospechosa defensa del plomero y presionó para que los propietarios sigan pagando; como no se encontró solución al problema, decidieron despedir a ambos. Ahora el plomero ha entablado juicio al consorcio para cobrar las facturas. El monto del juicio, incluidos los gastos, se ha estimado en \$55.000, que es la suma que deberían pagar en caso de perder ( $\theta_1$ ). Un fallo intermedio ( $\theta_2$ ) les costará \$12.500, pero si ganan el juicio ( $\theta_3$ ) calculan cobrar \$15.000. En la audiencia de conciliación, el abogado del actor manifestó que aceptaba concluir las acciones por un monto de \$20.000.

- a) *Si  $P(N_2) = 2P(N_1) = 3P(N_3)$ . Aplique el criterio del VME para decidir la estrategia óptima.*  
 b) *Calcule el VEIP.*

a) Seguir el juicio; b) VEIP = \$ 9.545.

3) Una fábrica de juguetes considera la posibilidad de lanzar al mercado un nuevo producto, con 70% de probabilidad de éxito, lo cual redundará en un beneficio de \$8.000. Si el producto fracasa, se ha estimado una pérdida de \$2.500 y si no se lanza, no habrá pérdidas ni ganancias.

- a) *Indique la estrategia óptima.*  
 b) *Calcule el máximo pago a un pronosticador 100% confiable.*  
 c) *¿Cuánto puede gastarse como máximo por la presentación del producto en una feria del juguete que pronostica correctamente el éxito o fracaso con 90% de probabilidad?*

a) Lanzar el producto; b) \$750; c) \$115.

4) Una empresa introducirá un nuevo producto en el mercado con los resultados de éxito total ( $\theta_1$ ), éxito parcial ( $\theta_2$ ) o fracaso ( $\theta_3$ ) cuyas probabilidades a priori no se han evaluado todavía. El problema consiste en establecer qué tipo de campaña publicitaria es más conveniente, intensa o moderada. Se ha logrado establecer la siguiente matriz de beneficios:

	Éxito total	Éxito parcial	Fracaso
Campaña Intensa	100	50	-50
Campaña Moderada	70	50	-20

- a) *Suponga que  $P(F_2) = 2P(F_1) = 6P(F_3)$ , indique la estrategia óptima y el máximo pago por una investigación de mercado perfecta.*  
 b) *Calcule el máximo pago por una investigación de mercado cuyo pronóstico es correcto con 80% de probabilidad.*

a) Intensa, VEIP=3; b) 1,5.

5) En un comercio de librería aledaño a la facultad se considera la posibilidad de adquirir una moderna fotocopiadora de alta velocidad. El éxito de la inversión depende de la cantidad de estudiantes del próximo año, que es desconocida y de pronóstico incierto. Se han considerado, en principio, tres posibilidades: Disminución (F1), igualdad (F2) y aumento (F3) de dicha cantidad. Los beneficios correspondientes serán -6.000, 3.000 y 10.000, respectivamente. Una segunda posibilidad es adquirir una máquina de menor capacidad pero más barata que arrojará beneficios de -1.500, 1.500 y 4.500, respectivamente. Finalmente, si no se adquiere nada, no habrá pérdidas ni ganancias.

- a) ¿Cuál sería la decisión óptima si las probabilidades son  $P(F1)=0,4$  y  $P(F2)=0,5$ .*
- b) ¿Cuánto podría pagarse al mejor pronosticador?*
- c) ¿Cuánto puede gastarse por una información muestral 80% confiable?*
- d) En caso de hacerse, ¿cuál sería la estrategia óptima?*
- e) Construya el árbol de decisión.*

a) MP; b) 1.900; c) 1.170; d) Si pronostica F1 no adquirir nada, y si pronostica F2 ó F3 adquirir la MG.

---

## **BIBLIOGRAFIA DE CONSULTA:**

- HANKE, J. E. Y REITSCH, A. G.: "Estadística para Negocios". Irwin. Madrid, 1996
- LEVIN, R. I. Y RUBIN, D.S.: "Estadística para Administradores", sexta edición. Prince Hall. México, 1996
- MENDENHAL, W. Y REINMUTH: "Estadística para Administración y Economía". Segunda edición. Grupo Editorial Latinoamericano. México, 1996
- HINES, W. Y MONTGOMERY, D.: "Probabilidad y Estadística para Ingeniería y Administración". Compañía Editorial Continental S.A.
- WALPOLE Y MYERS: "Probabilidad y Estadística". Cuarta edición. McGRAW-HILL.
- FERNANDEZ LOUREIRO DE PEREZ, Emma: "Estadística no paramétrica. A modo de introducción". Ediciones Cooperativas. Agosto de 2000.
- FERNANDEZ LOUREIRO DE PEREZ, Emma: "Decisión Estadística Bayesiana. A modo de introducción". Ediciones Cooperativas. Agosto de 2000.