

## Caso de estudio

En la ciudad de Río Cuarto se ha desarrollado una aplicación para dispositivos móviles que permite el pago de estacionamiento medido en las calles de la ciudad. Para esto almacena información de los usuarios, registra inicio y fin de estacionamiento de autos registrados y administra el saldo de las cuentas de los usuarios, entre otros aspectos. Se ha realizado una muestra de 500 usuarios que utilizaron el programa en el mes de enero y febrero de 2021.

### Parte 1:

Informe En base al caso de estudio presentado, resolver las consignas bajo las siguientes pautas:

- Trabajar sobre la Base de Datos que acompaña al caso.
- Utilizar el software especializado InfoStat.
- Trabajar en todo momento con redondeo simétrico y 4 decimales de precisión.

Elaborar un informe con el formato indicado en el “Instructivo General”, y utilizarlo para responder los siguientes puntos:

1. Haciendo uso del vocabulario técnico, defina cada concepto, siempre con relación al caso propuesto.

a. **Población – Muestra:** La muestra que observamos dentro del enunciado son los 500 usuario que utilizaron el programa en el mes de enero y febrero del año 2021. Esta muestra es de carácter representativo ya que es un subconjunto de la población que proporciona una visión útil de la naturaleza de la población estudiada.

b. **Unidad Estadística:** presente en el enunciado es cada uno de los usuarios que se han registrado en la aplicación que en conjunto constituyen a la población de estudio.

2. Construya la/s tabla/s de Distribución de Frecuencias con agrupación en lista o por intervalos, según corresponda. Calcule todas las frecuencias que conozca y que sean necesarias de las siguientes variables:

a. Saldo (Saldo): construir la tabla de frecuencias de 5 intervalos, cerrados a izquierda, utilizando [-15; 620] como mínimo y máximo, respectivamente.

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Saldo	1	[ -15,00	112,00 )	48,50	13	0,03	13	0,03
Saldo	2	[ 112,00	239,00 )	175,50	124	0,25	137	0,27
Saldo	3	[ 239,00	366,00 )	302,50	232	0,46	369	0,74
Saldo	4	[ 366,00	493,00 )	429,50	117	0,23	486	0,97
Saldo	5	[ 493,00	620,00 ]	556,50	14	0,03	500	1,00

b. Notificaciones activas (Noti)- Aclaración: 1=Notificación habilitada. 0= Notificación deshabilitada

Variable	Clase	Categorías	FA	FR	FAA	FRA
Noti	1	0	369	0,74	369	0,74
Noti	2	1	131	0,26	500	1,00

3. Calcule, para las variables definidas en el punto (2), las medidas descriptivas de posición, dispersión, simetría y curtosis, interpretando sus resultados en términos del problema planteado. Realice los cálculos trabajando cada variable como serie simple (directamente sobre la BD utilizando las funciones estadísticas de InfoStat).

## Medidas de resumen del Saldo

Resumen	Saldo
n	500,0000
Media	299,1590
D.E.	99,8095
Var (n-1)	9961,9401
Var (n)	9942,0162
CV	33,3634
Mín	-10,0875
Máx	568,9956
Mediana	296,6297
Q1	228,4459
Q3	368,1682
Asimetría	0,0894
Kurtosis	-0,1894
P(10)	170,9439
P(90)	427,0807

- n: 500. Se han observado los saldos de 500 personas
- Media: 299,1590 . Saldo promedio. En promedio cada usuario percibe un saldo de 299,1590.
- D.E: 99,8095. En promedio el saldo de un usuario se diferencia del valor promedio ( 299,1590) en aproximadamente 99,8095 ,es decir, existen saldos que se diferencian del promedio en más del 99,8095 y también existen saldos que se diferencian del promedio en menos 99,8095.
- Varianza corregida: 9961,9401 . Promedio del cuadrado de los desvíos de los saldos de los usuarios con respecto al promedio de saldos
- Varianza: 9942,0162. Promedio del cuadrado de los desvíos de los saldos de los usuarios con respecto al promedio de saldos
- Coeficiente de variación: 33,3634 . El valor de la DE es el 33,36% del valor de la media (valor del coeficiente de variación  $(0,333634 \times 100)$ , esto indica que se observa mucha variabilidad pero tolerable (CV comprendido entre el 33% y el 50%) y los valores observados se distribuyen de manera muy heterogénea (CV mayor al 26%) con respecto al número de saldos promedio
- Mínimo: -10,0875. El menor número de saldo es -10,0875
- Máximo: 568,9956 .El mayor número de saldo es 568,9956
- Mediana: 296,6297 . El 50% de los usuarios tienen un saldo de 296,6297 o menos y el otro 50% tiene un saldo de 296,6297 o más

- Primer Cuartil. Q1 : 228,4459. El 25% de los usuarios tienen un saldo de 228,4459 o menos y el 75% restante tiene un saldo mayor a 228,4459
- Tercer Cuartil: Q3 : 368,1682 . El 75% de los usuarios tienen un saldo de 368,1682 o menos y el 25% restante tiene un saldo mayor a 368,1682
- Asimetría: 0,0894 . Debido que el coeficiente de asimetría es positivo(0,0894) y que su media aritmética es mayor a la mediana presenta una leve asimetría derecha o positiva.
- Kurtosis: -0,1894 Al ser la kurtosis menor a cero nos indica que el grado de concentración alrededor de los valores centrales de la variable es reducido, por lo tanto, su distribución es platicúrtica
- P(10):170,9439. El 10% de los usuarios tienen un saldo de 170,9439 o menos y el restante 90% tiene un saldo mayor a 170,9439.
- P(90): 427,0807. El 90% de los usuarios tienen un saldo de 427,0807 o menos y el restante 10% tiene un saldo mayor a 427,0807.
- 

#### Medidas de resumen de Notificaciones

Resumen	Noti
n	500,0000
Media	0,2620
D.E.	0,4402
Var(n-1)	0,1937
Var(n)	0,1934
CV	168,0012
Mediana	0,0000
Asimetría	1,0858
Kurtosis	-0,8282

- n: 500. Se han observado el estado de la cuenta de 500 personas
- Media: 0,2620. cuentas habilitadas en promedio. En promedio hay un 0,2620 de cuentas habilitadas.
- D.E: 0,4402. En promedio las cuentas habilitadas se diferencia del valor promedio (0,2620) en aproximadamente 0,4402 ,es decir, existen varias cuentas habilitadas que se diferencian del promedio en más del 0,4402 y también existen varias cuentas habilitadas que se diferencian del promedio en menos 0,4402
- Varianza corregida: 0,1937 . Promedio del cuadrado de los desvíos de los saldos de los usuarios con respecto al promedio de saldos

- Varianza: 0,1934. Promedio del cuadrado de los desvíos de los saldos de los usuarios con respecto al promedio de saldos
- Coeficiente de variación: 168,0012 . El valor de la DE es el 168,0012% del valor de la media (valor del coeficiente de variación ( $168,0012 \times 100$ ), esto indica que se observa una variabilidad excesiva ( $CV > 50\%$ ) y los valores observados se distribuyen de manera muy heterogénea (CV mayor al 26%) con respecto al promedio de cuentas habilitadas
- Asimetría: 1,0858 . Debido a que el coeficiente de asimetría es positivo(1,0858) y que su media aritmética es mayor a la mediana presenta una leve asimetría derecha o positiva.
- Kurtosis: -0,8282 Al ser la kurtosis menor a cero nos indica que el grado de concentración alrededor de los valores centrales de la variable es reducido, por lo tanto, su distribución es platicúrtica

4. La empresa desarrolladora desea estimar con un nivel de confianza del 90% la media del “Saldo” que es el monto de dinero que actualmente posee el usuario en la cartera del producto. Realice un análisis estadístico adecuado para el parámetro de interés indicando lo siguiente:

a. ¿Cuál es el parámetro de interés y cuál es la variable de estudio?

El parámetro de interés en este análisis es la media poblacional ( $\mu$ ) y la variable de estudio es el promedio de dinero que actualmente posee el usuario en la cartera del producto en la muestra ( $\bar{x}$ ).

b. Exprese la fórmula del estadístico de prueba, su distribución y enumere los supuestos teóricos subyacentes al estadístico utilizado.

Estadístico:

$$\bar{X} = \left( \sum_{i=0}^n x_i \right) / n$$

Distribución:

$$K(\bar{x}; \mu) = (\bar{x} - \mu) / \hat{s} / \sqrt{n} \sim N(0; 1)$$

Como la muestra es de gran tamaño, es decir que  $n > 30$ , además la varianza poblacional es desconocida, se debe utilizar una distribución normal para calcular la media muestral. La distribución normal se describe como:

$$N(0;1). \hat{\theta} = \bar{x} : \text{Media Muestral} ; \quad \bar{x} = \left( \sum x_i \right) / n = (x_1 + x_2 + \dots + x_i) / n$$

Los valores de la media muestral se distribuyen alrededor de una Esperanza [ $E(\bar{x})$ ], que se ha verificado que coinciden con el valor medio  $\mu$  de la variable  $X$ , a esta relación entre la esperanza del estimador media muestral y el valor medio de la variable  $X$  en la población se denomina **Ins sesgabilidad**.

$$V(\bar{x}) = (\sigma^2) / n ; \sigma^2 : \text{Varianza poblacional} , n : \text{Tamaño de muestra}$$

Se ha verificado que coincide con el resultado del cociente ( $\sigma^2$ ) de la variable X de acuerdo a sus valores de la población y el tamaño de muestra (n).

$DE(\bar{x}) = \sigma / \sqrt{n}$ ;  $\sigma$  : Varianza poblacional, n: tamaño de muestra También se verifica que la desviación estándar del estimador media muestral resulta del cociente entre la desviación estándar ( $\sigma$ ) de la variable X de acuerdo a sus valores en la población y la raíz cuadrada del tamaño de la muestra (n). Por lo tanto podemos concluir que el estimador es **Eficiente**

El estimador media muestral tiene la propiedad de **Consistencia** porque al seleccionar muestra de mayor tamaño es más probable que su valor se aproxime al valor de parámetro (Ley de Números Grandes).  $\lim_{n \rightarrow \infty} \Pr(|\bar{x} - \mu| < d) = 1$

Además de lo expresado, podemos observar que este estimador cumple con la condición de **Suficiencia** porque utiliza toda la información contenida en la muestra, respecto al promedio a estimar.

c. Procese los datos y pegue el análisis realizado.

### Intervalos de confianza

*Bilateral  
Estimación paramétrica*

Variable	Parámetro	Estimación	E.E.	n	LI(90%)	LS(90%)
Saldo	Media	299,1590	4,4636	500	291,8033	306,5146

Conclusiones: El saldo medio de la muestra estimada es de \$299,1590 para una muestra de 500 usuarios, con un nivel de confianza de 0,90. El cálculo presentado nos deja ver los límites de nuestra estimación por intervalos.

d. Interprete en términos estadísticos y en el **contexto del problema** las conclusiones.

En base a los resultados expuestos podemos determinar que con un 90% de nivel de confianza los saldos promedios en la población se encuentran entre 291,80 y 306,51.

e. Realice una comparación de la estimación del parámetro de interés, con una confianza del **95 %** y obtenga conclusiones.

### Intervalos de confianza

*Bilateral*

*Estimación paramétrica*

Variable	Parámetro	Estimación	E.E.	n	LI (95%)	LS (95%)
Saldo	Media	299,1590	4,4636	500	290,3892	307,9288

A partir de la comparación realizada podemos observar que el tamaño del intervalo crece al mismo tiempo que aumentamos el nivel de confianza. Esto se traduce como una disminución de la precisión de la estimación, pero es necesario para que aumente la probabilidad de que el valor real se encuentre dentro del intervalo.

5. Un investigador contrasta la hipótesis, afirmando que el Saldo promedio de los usuarios es mayor a \$400.

a. En base a la información relevada, concluya acerca de la afirmación del investigador con un nivel  $\alpha=0,05$ .

$$\bar{x} = 299,1590$$

$$H_0: \mu_0 = 400$$

$$H_1: \mu_1 > 400 : \text{lateral derecha}$$

$$\text{nivel de significación} = 5\% \rightarrow \alpha = 0,05$$

$$\text{nivel de confianza} = 95\% \rightarrow 1 - \alpha = 0,95$$

$$\sigma^2 = 9961,9401 \rightarrow 99,8095$$

$$DE = 99,8095$$

$$z^* = 1,6448$$

$$\bar{x}^* = 407,3418$$

Nuestra hipótesis alternativa incluye los valores mayores que el supuesto en la hipótesis nula, por lo que tenemos una dódima lateral derecha. Partiendo de un nivel de significancia de 0.05 se establece un nivel de confianza del 95%. Además, al tratar con una muestra grande ( $>30$ ) podemos aproximar una distribución normal para trabajar en la prueba de hipótesis.



b. Procese los datos y pegue el análisis realizado.

$$\begin{aligned}\bar{x}^* &= \mu_0 + (z^*) \cdot \hat{s} / \sqrt{n} \\ &= 400 + 1,6448 \cdot 99,8095 / \sqrt{500} \\ \bar{x}^* &= 407,3418\end{aligned}$$

#### Prueba t para una media

Valor de la media bajo la hipótesis nula: 400

Variable	n	Media	DE	LI (95)	T	p(Unilateral D)
Saldo	500	299,1590	99,8095	291,8033	-22,5918	>0,9999

c. Interprete en el contexto del problema y justifique.

Como la Hipótesis alternativa planteada es mayor a la Hipótesis nula, se presenta una prueba lateral derecha, quedando el coeficiente de significación  $\alpha$  del lado derecho. Luego a partir de la gráfica, podemos determinar el punto crítico que en este caso es  $\bar{x}^* = 407,3418$  es este valor quien divide la zona de rechazo de la de aceptación. Zona de Aceptación:  $(-\infty ; 407,3418]$ , Zona de Rechazo:  $[407,3418 ; \infty)$  En base a la evidencia muestral y con un nivel de significación del 5%, no se rechaza la Hipótesis Nula ya que  $\bar{x} = 299,1590$  queda dentro de la zona de aceptación.

Sin embargo, se podría estar cometiendo el error tipo 2.

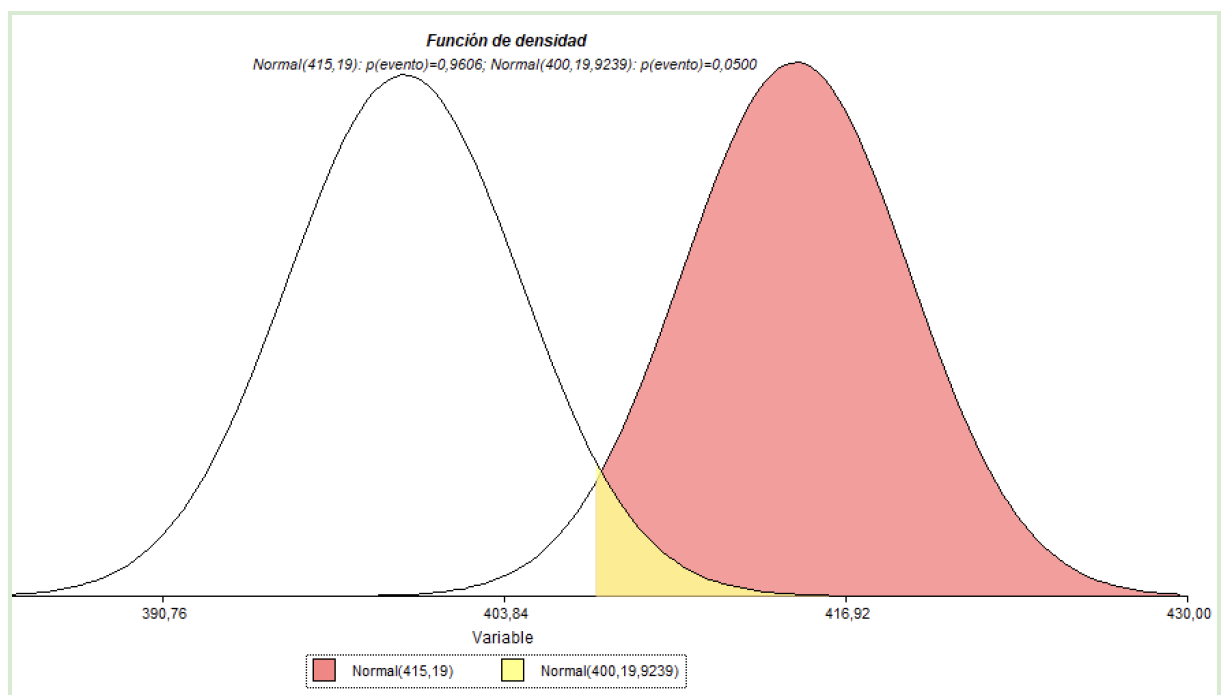
6. En base a lo planteado en la actividad 5 calcule y grafique la potencia de la prueba si se supone que el parámetro de interés es igual a \$415, considerando una varianza poblacional de \$9500. Interprete en el contexto del problema y justifique la distribución elegida.

$$\begin{aligned} H_1 &= 415 \\ \sigma^2 &= 9500 \\ \sigma &= 97,4679 \end{aligned}$$

$$z = \bar{x}^* - \mu_1 / \sigma / \sqrt{n} = (407,3418 - 415) / (97,4679 / \sqrt{500}) = -1,7569$$

$$B = \Pr \{ z < -1,7569 \} = 0,0394$$

$$\begin{aligned} -B &= \Pr \{ z > -1,7569 \} = 0,9606 \\ &\Pr \{ \text{Rechazar } H_0 / H_0 \text{ Falsa} \} \end{aligned}$$



7. Analizando la variable “Notificaciones activas”, se desea tomar una muestra con un riesgo del 2% y un error máximo aceptable de 0,03:

a. ¿Qué tamaño de muestra es el adecuado para estos requerimientos?

Utilizar el estimador puntual obtenido en el punto 2.

riesgo = 2%

error = 0,03

$z(0,01) = 2,3263$

alfa = 0,02

$$n = \frac{Z^2 \hat{P}(1 - P)}{e^2}$$

$n \geq 1162$

b. ¿Qué tamaño debería ser la muestra si se sabe que en otras ciudades que implementaron el mismo sistema, se tiene un nivel de notificaciones activas del 25% y se asume el mismo comportamiento?

riesgo = 2%

error = 0,03

$z(0,01) = 2,3263$

alfa = 0,02

$$n = \frac{Z^2 \hat{P}(1 - P)}{e^2}$$

$n \geq 1127$