PROBABILIDAD Y ESTADÍSTICA

Unidad IV
Estimación de Parámetros

Parte I

Mg. Luis Arenas

Facultad de Economía y Administración
UNIVERSIDAD NACIONAL DEL COMAHUE

Estimación de parámetros: Parte I

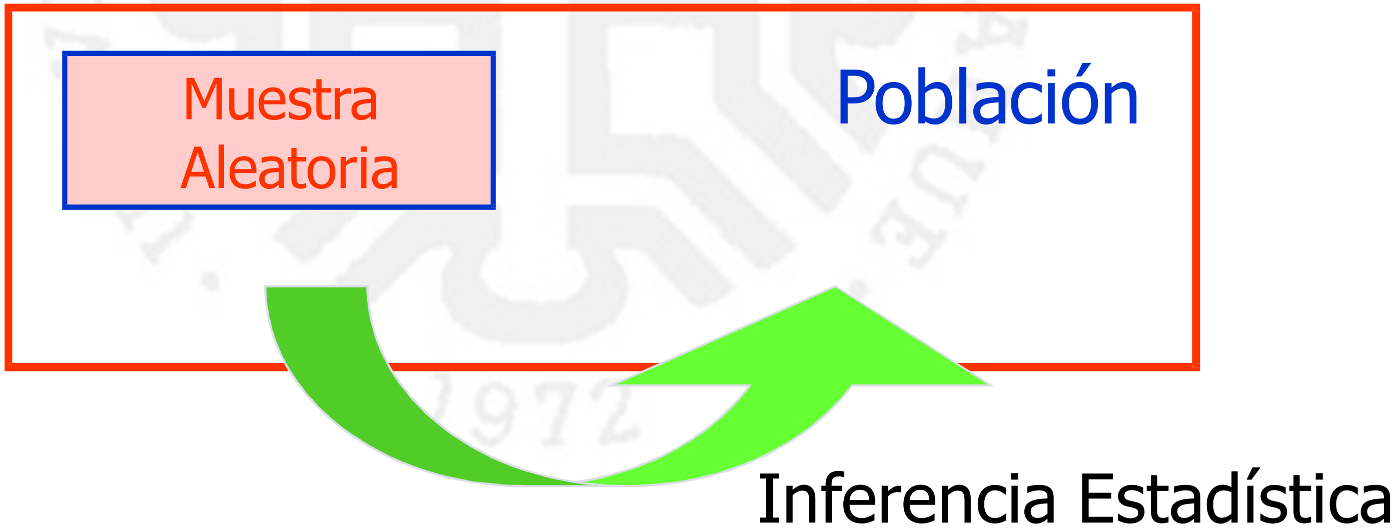
Estimación puntual y por intervalos

CONTENIDOS

- Inferencia Estadística.
- Estimación puntual: definición, ejemplos, desventajas.
- Intervalo de Confianza: definición, concepto.
- Estimación de la media en una población Normal con varianza conocida.
- Estimación de la media poblacional en una población normal con varianza desconocida. Distribución t-Student.
- Intervalo para una proporción.
- Intervalo de confianza para la varianza.

Inferencia Estadística

Colección de técnicas estadísticas que permiten a partir de una *muestra aleatoria* extraída de una población de interés realizar *estimaciones* de parámetros y *pruebas de hipótesis*, y que proporcionan mediante probabilidad una medida del *riesgo de error* de las estimaciones.



Estimación

La estimación estadística es el proceso mediante el cual se aproxima el valor de un parámetro de la población desconocido a partir de la información proporcionada por una muestra aleatoria.

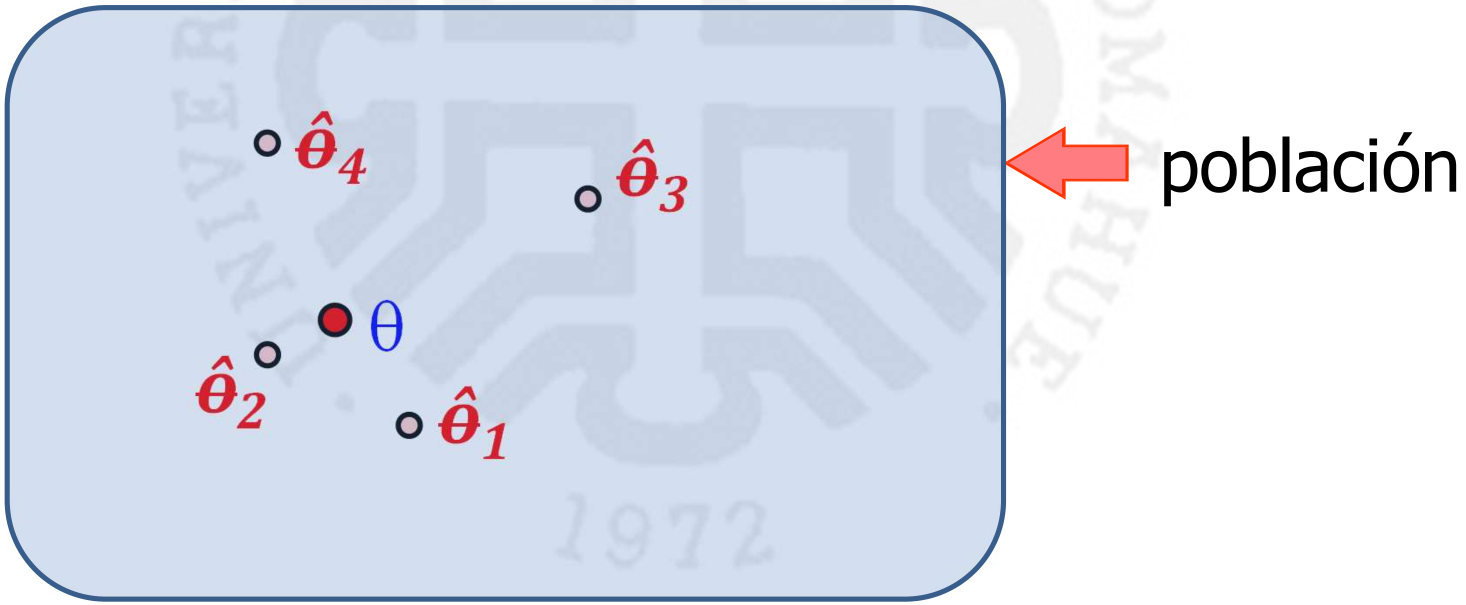
estimación

- puntual: produce un único valor como estimación del parámetro de interés.
- por intervalo: produce como estimación un intervalo de valores reales y se espera que el intervalo contenga al parámetro.

Estimación puntual: graficamente

Deseamos conocer el valor de un parámetro θ de la población, por ejemplo: el valor de la media μ.

Utilizaremos un estimador puntual $\hat{\theta}$ del parámetro por ejemplo: la media de la muestra $\overline{X}$



Estimación puntual

Un *estimador* $\hat{\theta}$ es una función de las variables aleatorias observables $(X_1, X_2, \ldots\ldots, X_n)$ que conforman la muestra aleatoria.

Un *estimador* $\hat{\theta}$ establece el mecanismo funcional por medio del cual se puede hacer una *estimación puntual* de θ una vez que se tienen las observaciones o realizaciones de la muestra.

La *estimación puntual* de θ puede determinarse cuando ya se dispone de la muestra observada, es decir, se conocen las observaciones $(x_1, x_2, \ldots\ldots, x_n)$ de una muestra particular.

Estimación puntual

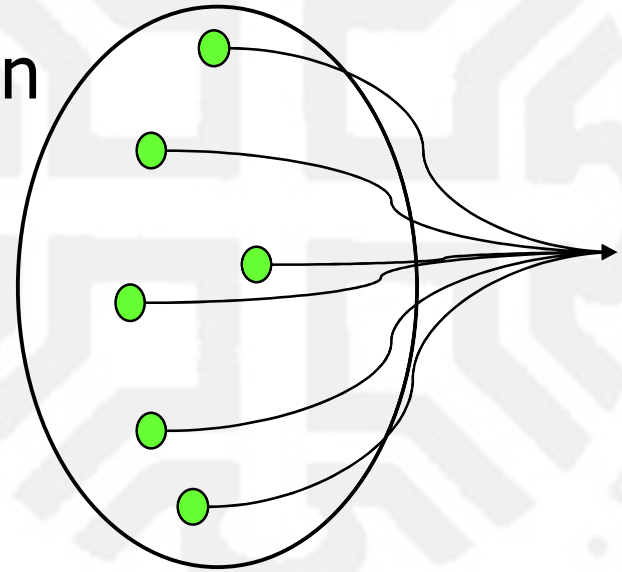
$\hat{\theta} = g(X_1, X_2, \ldots, X_n) \Leftarrow$ $\hat{\theta}$ es el estimador de θ

---

valor de $\hat{\theta}$ en la *muestra observada* $= g(x_1, x_2, \ldots, x_n) \Leftarrow$ es la estimación puntual de θ

Población

$X \sim ??(\mu; \sigma)$



$$\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i$$

Estimación puntual

La media como función de la muestra aleatoria es un *estimador puntual* de μ.

La función *g()* indica que para obtener la *estimación* de μ, se deben sumar los valores y dividir por n.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Cuando ya se cuenta con las n observaciones de la muestra observada, se puede calcular el valor de la media de la muestra, ese valor es una *estimación puntual* de μ.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ejemplo

En un relevamiento realizado en la formación Los Molles (Neuquén) se recolectaron 60 rocas de halita y se estudió la densidad de los cristales de cloruro de sodio (NaCl)(en gr/cm3). Los datos son:

2,1	2,14	2,32	2,38	2,39	2,46	2,48	2,5	2,53	2,55
2,57	2,58	2,59	2,62	2,67	2,69	2,73	2,74	2,75	2,75
2,76	2,76	2,77	2,79	2,82	2,84	2,84	2,89	2,89	2,91
2,91	2,95	2,95	2,95	2,99	2,99	2,99	2,99	3,01	3,03
3,03	3,07	3,08	3,08	3,08	3,08	3,08	3,09	3,1	3,11
3,13	3,14	3,18	3,18	3,19	3,21	3,23	3,24	3,26	3,35

Dado que no se conoce la media μ ni el desvío estándar $\sigma(x)$ de esta población, haremos una estimación puntual de esos parámetros.

Ejemplo

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{171,48}{60} = 2,858$$ estimación puntual de μ.

Se estima que los cristales de cloruro de Na en las rocas de halita tienen una densidad media de 2,858 gr/cm³.

$$s(x) = \sqrt{\frac{494,9 - \frac{(171,48)^2}{60}}{59}} = 0,286$$ estimación puntual de $\sigma(x)$

Se estima que la densidad de los cristales de cloruro de Na en las rocas de halita tienen un desvío estándar de 0,286 gr/cm³.

Parámetros, estimadores y estimación

parámetro	estimador	estimación
media μ	$\overline{X}$	$\overline{x}_{obs}$
proporción π	p	p_{obs}
variancia σ^2	$S^2(x)$	$s^2_{obs}(x)$
dif. de medias $\mu_1 - \mu_2$	$\overline{X}_1 - \overline{X}_2$	$(\overline{x}_1 - \overline{x}_2)_{obs}$
dif. de proporciones $\pi_1 - \pi_2$	$p_1 - p_2$	$(p_1 - p_2)_{obs}$
cociente de variancias σ_1^2/σ_2^2	$S_1^2(x) / S_2^2(x)$	$\left(s_1^2(x) / s_2^2(x) \right)_{obs}$

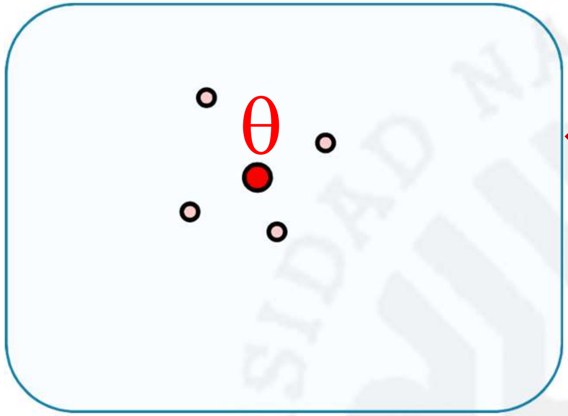
constantes desconocidas

variables aleatorias

calculados con la muestra observada

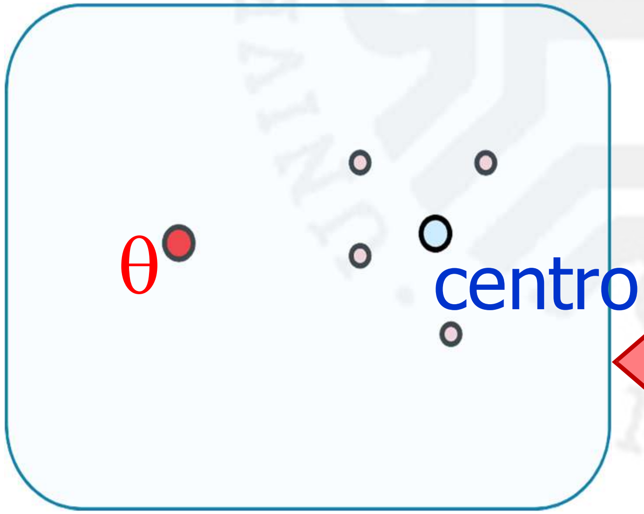
Propiedades de los estimadores

Estimador insesgado



estimador
insesgado

decimos que un estimador es insesgado si produce estimaciones que son correctas en promedio.

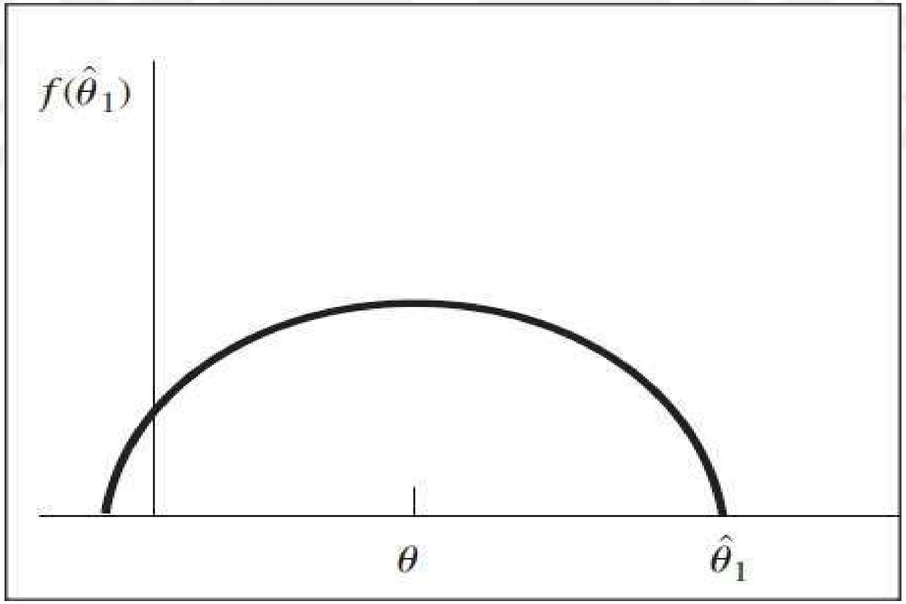


estimador
sesgado

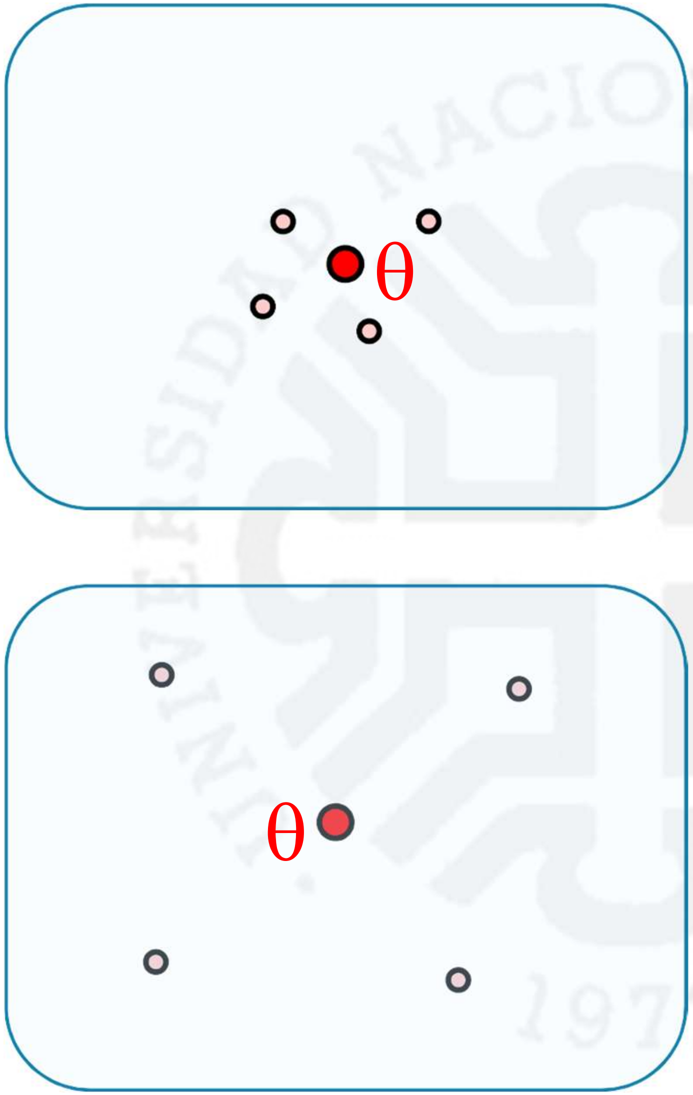
Propiedades de los estimadores: insesgado

Un estimador $\hat{\theta}$ se dice *insesgado* si la media de la distribución muestral de $\hat{\theta}$ es θ, es decir si: $E(\hat{\theta}) = \theta$

Ejemplo: la media muestral es un estimador *insesgado* de la media poblacional pues: $E(\bar{X}) = \mu$



Estimador eficiente



estimador
eficiente

es insesgado
y tiene
varianza
menor

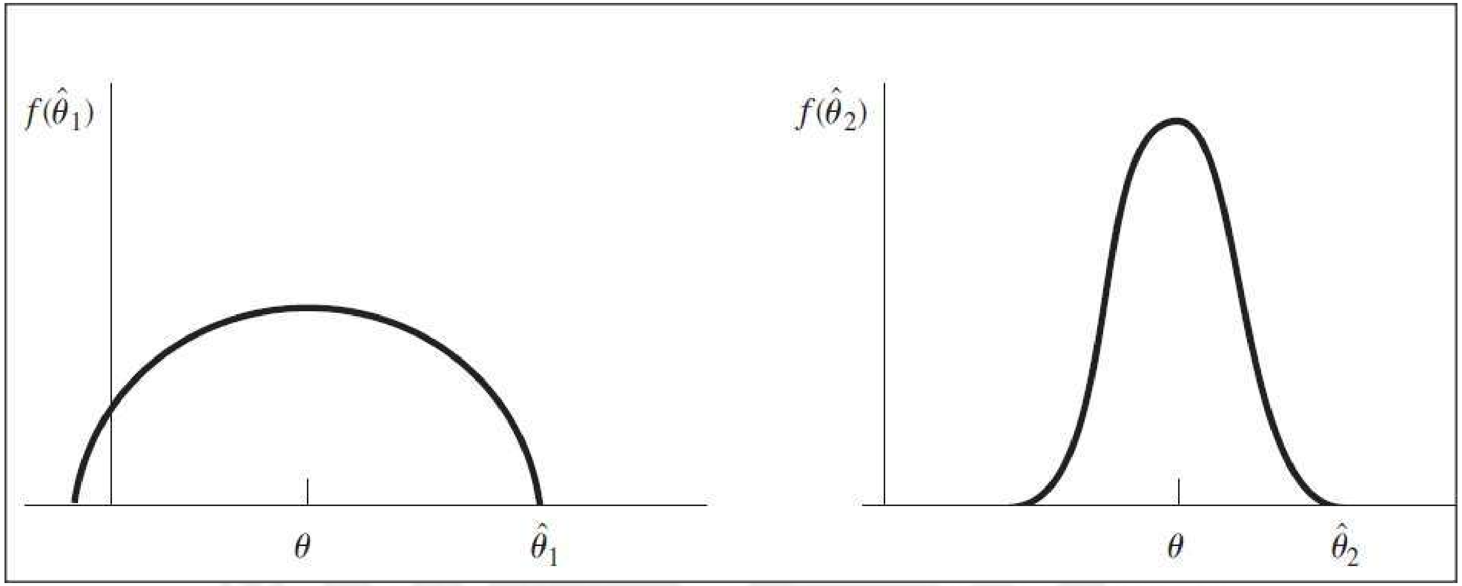
Eficiente

Si $\hat{\theta}_j$ es un estimador insesgado de θ y no hay otro estimador insesgado de θ con varianza menor, se dice que $\hat{\theta}_j$ es el estimador más eficiente de θ.

Si $V(\hat{\theta}_j) < V(\hat{\theta}_i) \quad \forall\ \hat{\theta}_i$ estimador insesgado de $\theta \Rightarrow \hat{\theta}_j$ es el estimador más *eficiente* de θ.

Considerando todos los estimadores insesgados de θ, el que tiene varianza mínima se llama estimador *eficiente* de θ o estimador de *mínima varianza* de θ.

Eficiente



Como $V(\hat{\theta}_2) < V(\hat{\theta}_1) \Rightarrow \hat{\theta}_2$ es un estimador de θ más *eficiente* que $\hat{\theta}_1$.

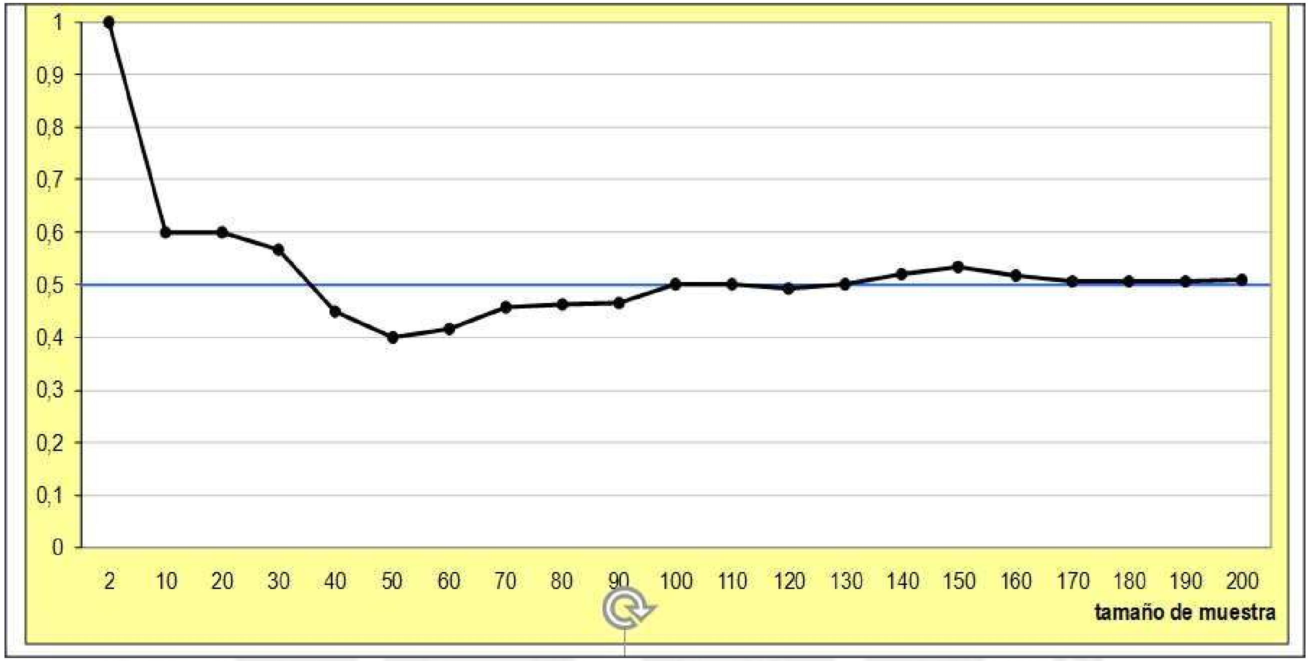
Propiedades de los estimadores: consistente

Sea $\hat{\theta}_n$ un estimador de θ basado en una muestra de tamaño n y sea $\varepsilon > 0$ arbitrariamente

pequeño, $\hat{\theta}_n$ es un estimador *consistente* de θ si:

$$P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) \xrightarrow[n \to \infty]{} 1$$

Estimador consistente: ejemplo



A medida que aumenta el tamaño de la muestra n, la proporción muestral de caras p, se aproxima a la proporción teórica de caras π.

$$P(|p_n - \pi| < \varepsilon) \xrightarrow[n \to \infty]{} 1$$

Error de muestreo

Se define como la diferencia en valor absoluto entre el *estimador* y el *parámetro* :

$$Error = |\hat{\theta} - \theta|$$

Ejemplo

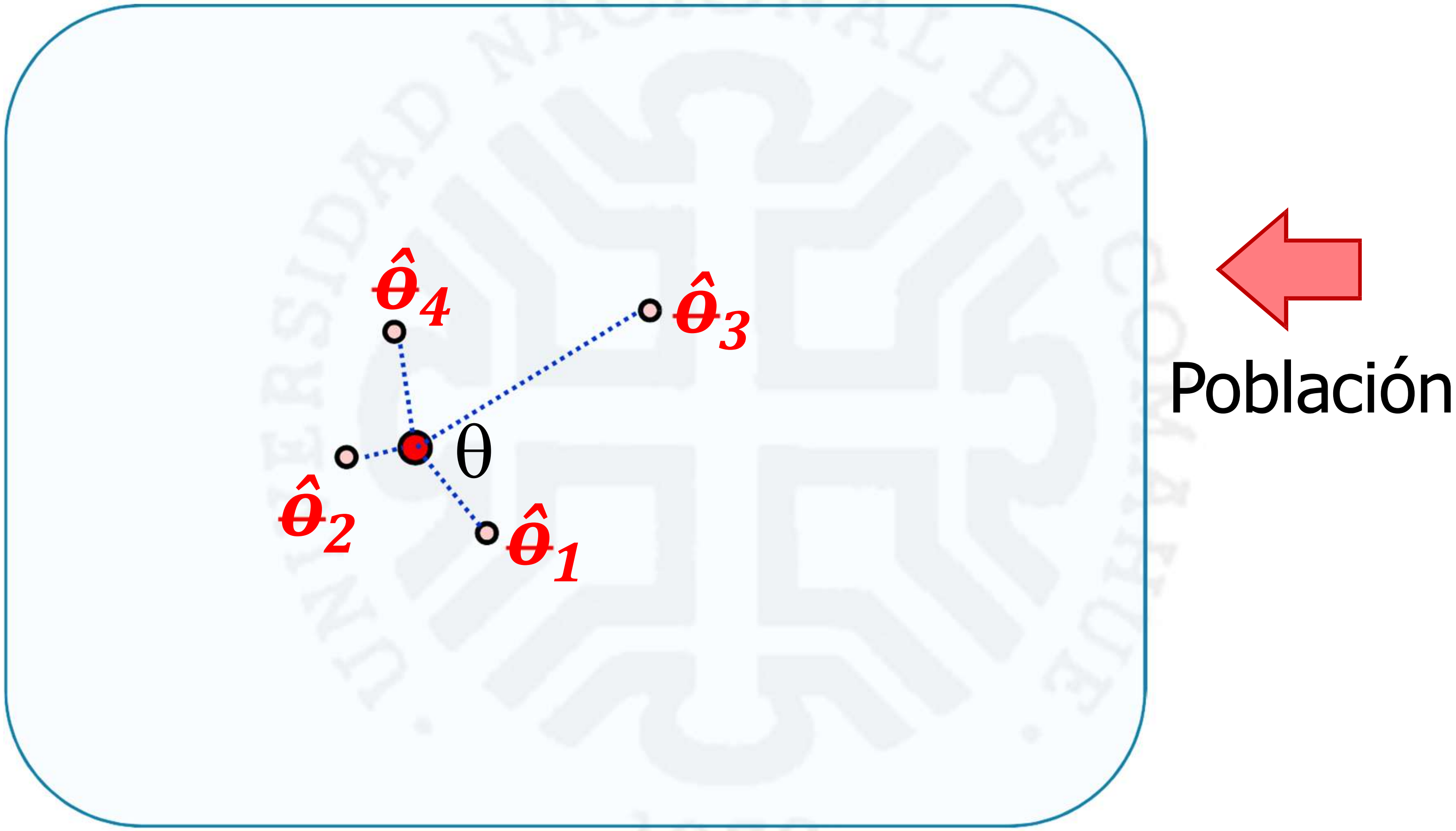
Si estimamos la media de la población μ, el error de muestreo es:

$$Error = |\bar{X} - \mu|$$

Si estimamos la varianza $\sigma^2(x)$ de la población el error de muestreo es:

$$Error = |S^2(x) - \sigma^2(x)|$$

Error de muestreo

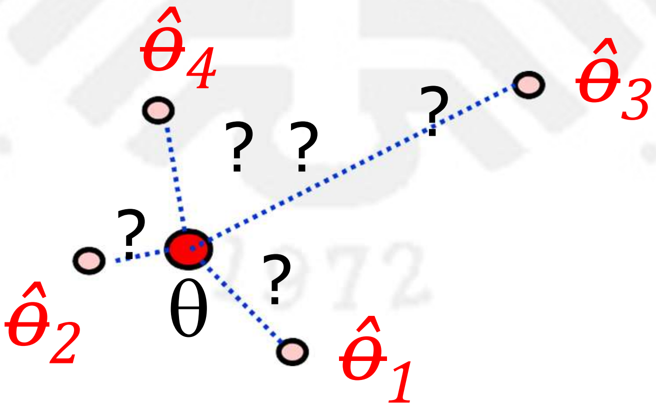


Estimación puntual: desventajas

Como no se conoce el error de muestreo, no se sabe cuán cerca está la estimación del verdadero valor del parámetro.

La estimación puntual no da información de la *precisión* ni de la *confiabilidad* de la estimación.

En la estimación puntual, no se sabe si la estimación obtenida está o no próxima al verdadero valor de θ.



Ejemplo: Error de muestreo

¿cuál es el Error de muestreo en la estimación de la densidad media μ de cloruro de sodio?

$$Error = |\bar{X} - \mu|$$

Para la muestra de 60 cristales de halita el Error de muestreo sería:

$$Error = |2{,}858 - \mu| \quad ?$$

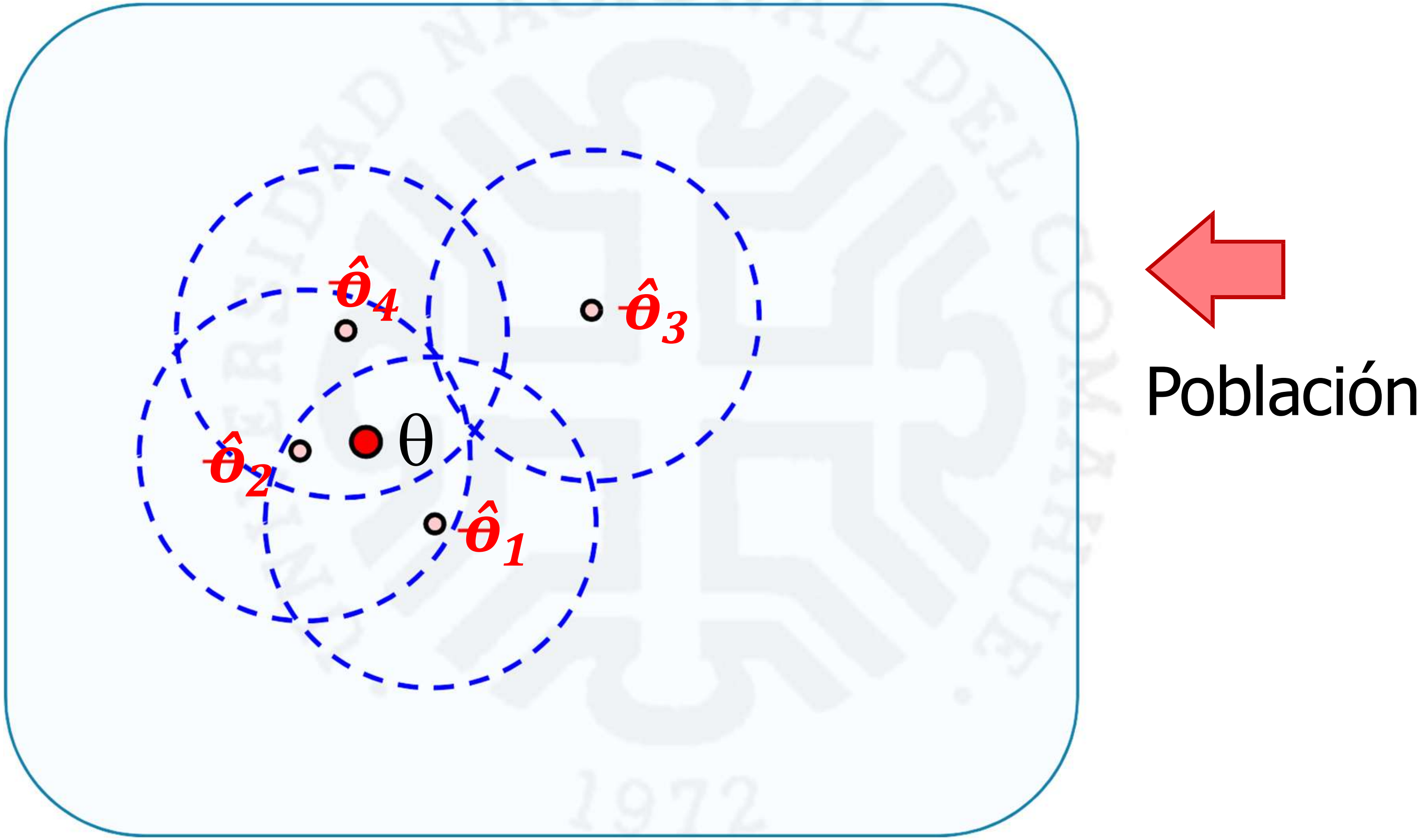
No se puede determinar el valor del Error de muestreo porque no se conoce el valor verdadero de la densidad media μ de cloruro de sodio.

Estimación por intervalo

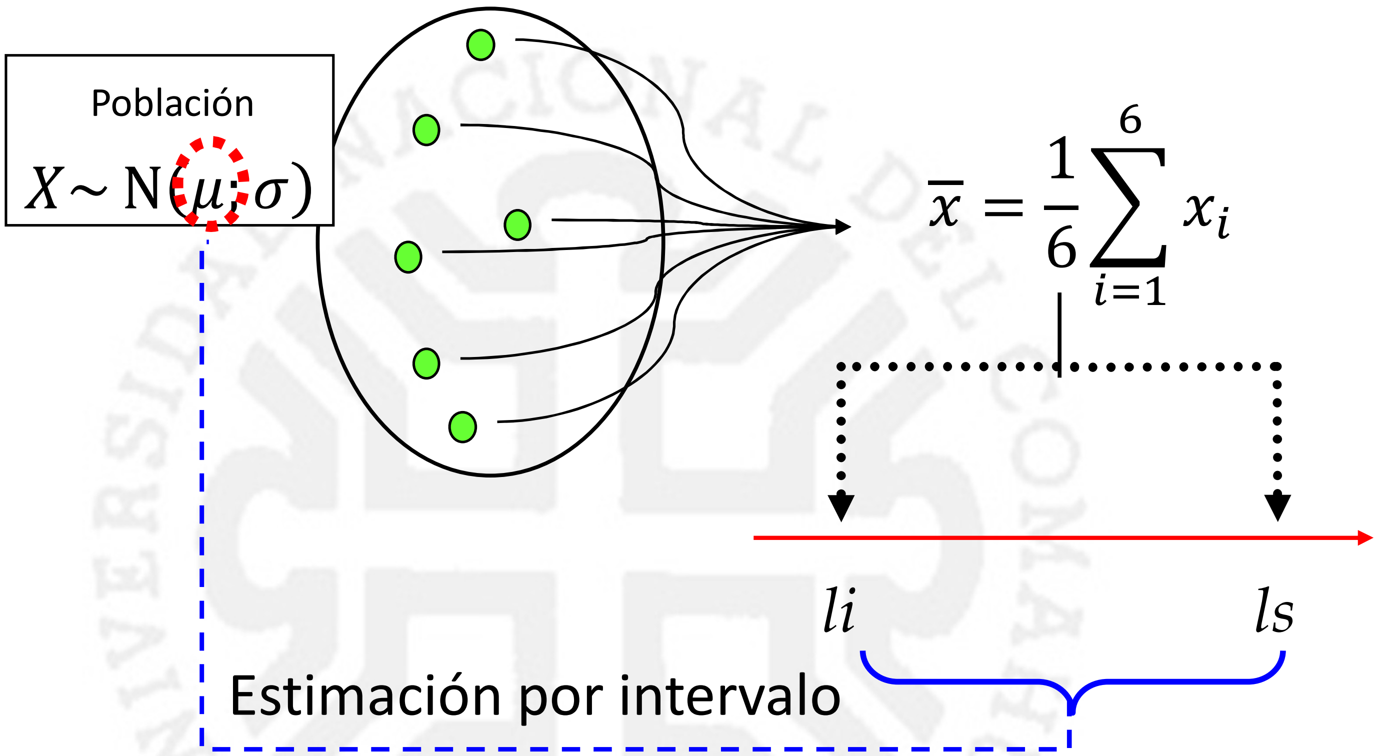
La solución es utilizar la estimación puntual para construir un intervalo de valores reales $(li\ ;\ ls)$ que *confiamos* incluya al verdadero valor del parámetro.

Este intervalo llamado *intervalo de confianza* se construye determinando el límite inferior li y el límite superior ls a partir de la estimación puntual y utilizando el procedimiento de la variable pivotal.

Estimación por intervalo: graficamente



Estimación por intervalo



En cada muestra habrá estimaciones puntuales diferentes del parámetro y valores distintos de li y ls, es decir, de una muestra a otra habrá distintos *intervalos de confianza*.

Estimación por intervalo

Los límites del intervalo son funciones de la información muestral, por lo tanto son variables aleatorias LI y LS respectivamente tales que:

$$P(LI \leq \theta \leq LS) = 1 - \alpha$$

Donde 1-α recibe el nombre de *nivel de confianza*.

El intervalo antes de determinar el valor de sus límites para una muestra particular es un intervalo aleatorio pues depende de los valores que resulten en la muestra aleatoria.

Nivel de confianza

El nivel de confianza 1- α debe interpretarse como la probabilidad de que el intervalo aleatorio incluya al verdadero valor del parámetro θ antes de extraer la muestra, es decir:

$$P[\theta \in (LI; LS)] = 1 - \alpha$$

Una vez obtenida la muestra, la probabilidad de que intervalo contenga al parámetro es 0 si no lo contiene y 1 si contiene a θ.

En esta situación no es correcto decir que 1- α es la probabilidad de que el intervalo contenga a θ,sino que tenemos una *confianza* de (1-α)% que el intervalo contenga a θ.

Concepto de confianza en muestreo repetido

- Si en una población se extrae un número infinito de muestras aleatorias independientes del mismo tamaño n.
- En cada muestra, se calcula un intervalo de confianza del $100(1-\alpha)\%$ para el parámetro θ.

Se espera que $100(1-\alpha)\%$ de esos intervalos contengan el verdadero valor de θ.

Es decir que el nivel de confianza es el porcentaje esperado de intervalos que contienen el verdadero valor de θ.

Concepto de confianza en muestreo repetido

- En la práctica solo se tiene una muestra aleatoria y el intervalo $li \leq \theta \leq ls$ es una estimación de θ y por lo tanto no es razonable asociar a él una probabilidad.

- La proposición correcta es que el intervalo *contiene a θ con una confianza del $100(1-\alpha)\%$.*

No se sabe si para la muestra que tenemos intervalo contiene al parámetro o no. Sin embargo, podemos afirmar que el método utilizado para obtener el intervalo proporciona intervalos que contienen a θ el $100(1-\alpha)\%$ de las veces.

Probabilidad inicial y confianza

El siguiente ejemplo fue extraído del libro Fundamentos de Inferencia Estadística por Ruiz Maya-Pérez y Martín Pliego.

Un estudiante rinde con un programa que tiene 100 temas de los cuales hay 1 que no sabe. No recuerda exactamente el número de ese tema, tendrá que mirarlo en el programa. El examen consiste hacer girar un bolillero para extraer una bolilla con el número del tema y desarrollarlo ante los profesores. Si lo sabe aprueba, sino desaprueba.

El estudiante tiene una probabilidad de 0,99 de aprobar antes de sacar la bolilla.

Probabilidad inicial y confianza

Supongamos que está ante los profesores y saca una bolilla con el tema:

- la probabilidad que tiene de aprobar no sigue siendo 0,99
- ahora la probabilidad es 0 si sacó el tema que no sabe y 1 si sacó algunos de los temas que sabe.
- mientras consulta el programa él tiene una *confianza* del 99% de aprobar porque antes de sacar la bolilla tenía una probabilidad de 0,99 de sacar un tema que sabía.

Variable pivotal

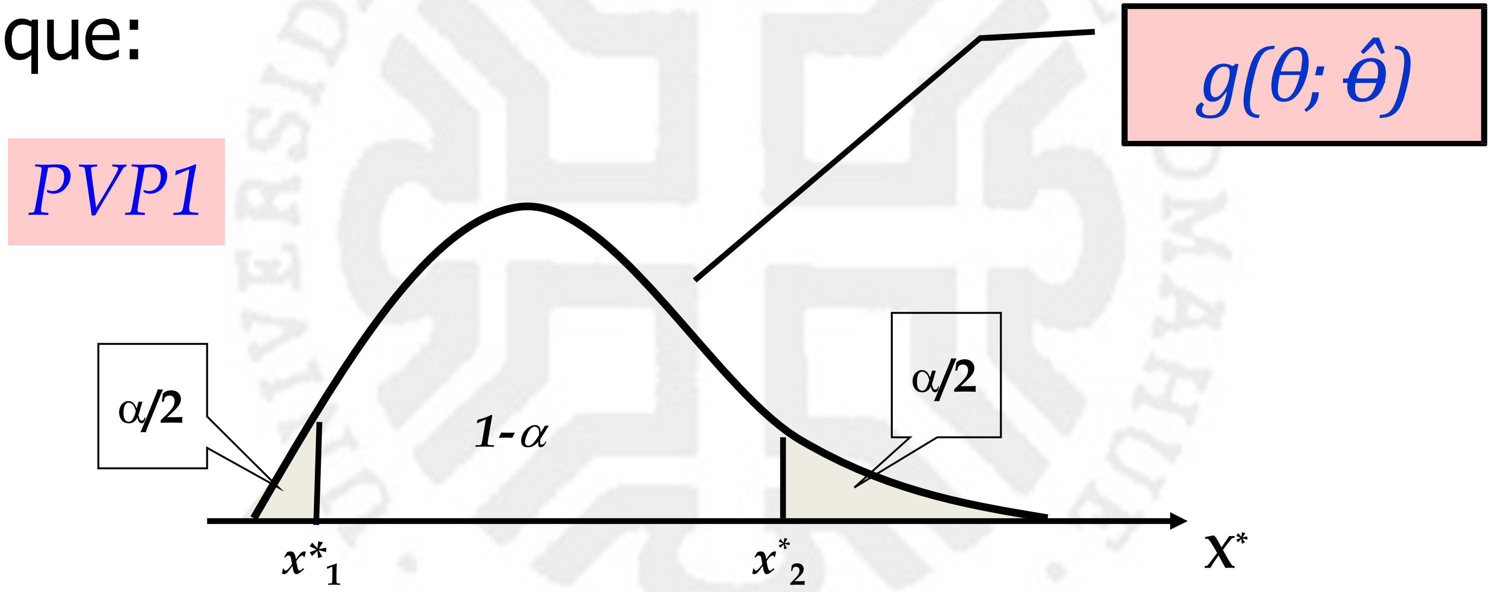
Para construir el intervalo se utiliza una nueva variable que llamamos *variable pivotal* $X^*=g(\theta;\hat{\theta})$.

La *variable pivotal* es una función que no depende del parámetro, (ni de otros parámetros) pero que vincula al parámetro y al estimador.

Además tiene una distribución de probabilidad conocida, con valores de probabilidad tabulados o que se pueden hallar utilizando tablas o programas estadísticos como R Commander.

Procedimiento de la variable pivotal

1. A partir de la distribución de la variable pivotal y para un *nivel de confianza* $1\text{-}\alpha$ es posible seleccionar dos valores críticos x_1^* y x_2^* tales que:



$$P(x_1^* \leq X^* \leq x_2^*) = 1 - \alpha \quad 0 \leq \alpha \leq 1$$

Procedimiento de la variable pivotal $PVP2$

2. A partir de la variable pivotal X^* se efectúan pasos algebraicos miembro a miembro hasta despejar el parámetro θ, es decir, se arriba a una expresión como:

$$P(LI \leq \theta \leq LS) = 1 - \alpha$$

Donde los límites LI y LS son funciones del estimador $\hat{\theta}$ y por lo tanto variables aleatorias.

Además LI y LS proporcionan la expresión matemática para hallar los límites del intervalo una vez que se disponga de la muestra observada.

Procedimiento de la variable pivotal

PVP3

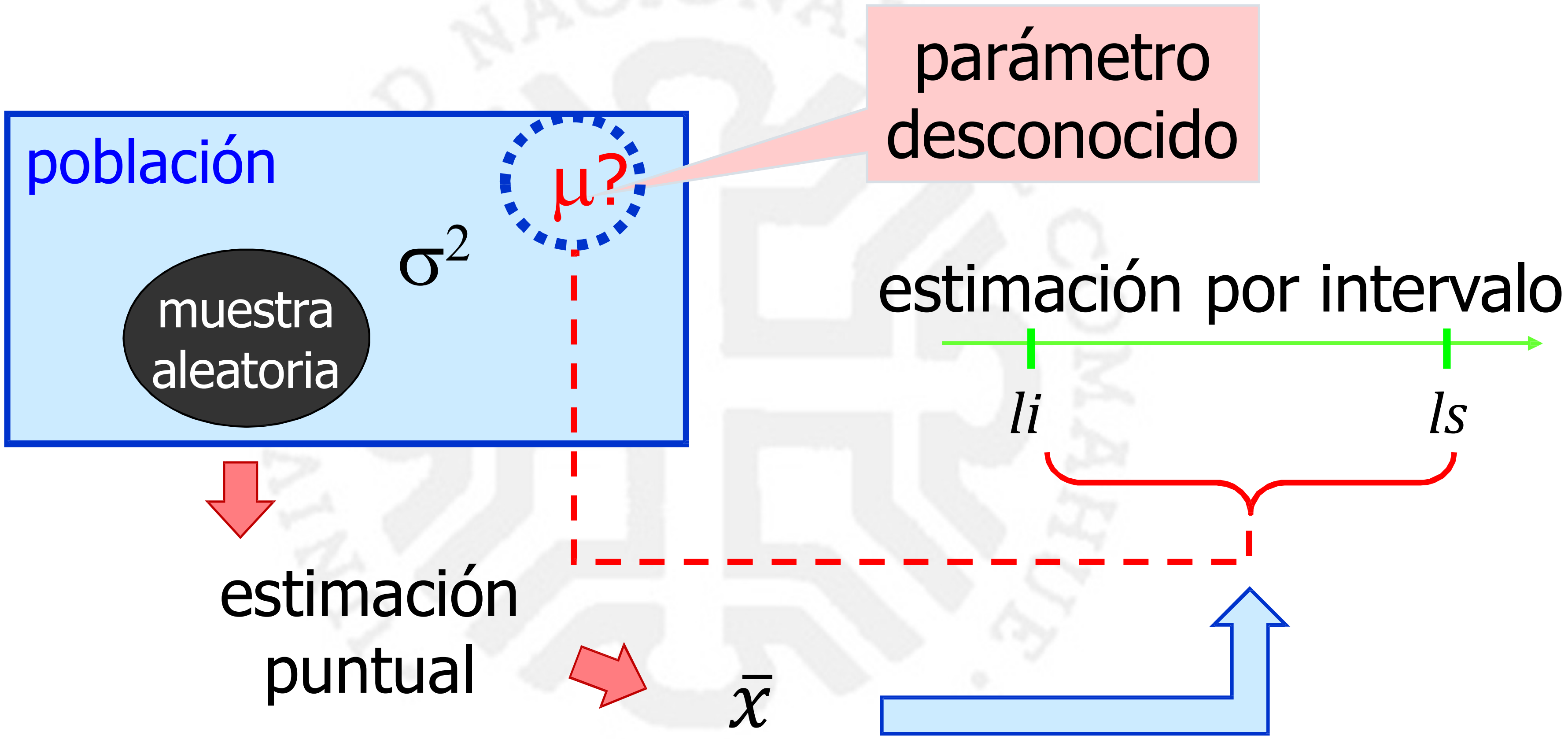
3. Luego, mediante un procedimiento que garantice la aleatoriedad, se extrae la muestra de la población y con los datos observados se calculan los límites del intervalo de confianza cuya expresión es:

$$li \leq \theta \leq ls$$

Este intervalo de valores reales es una *estimación por intervalo* de θ.

Hay una confianza del $100(1-\alpha)\%$ de que este intervalo contenga el valor verdadero de θ.

Intervalo para la media con σ^2 conocida



Intervalo para la media con σ^2 conocida

La distribución en el muestreo de la media es:

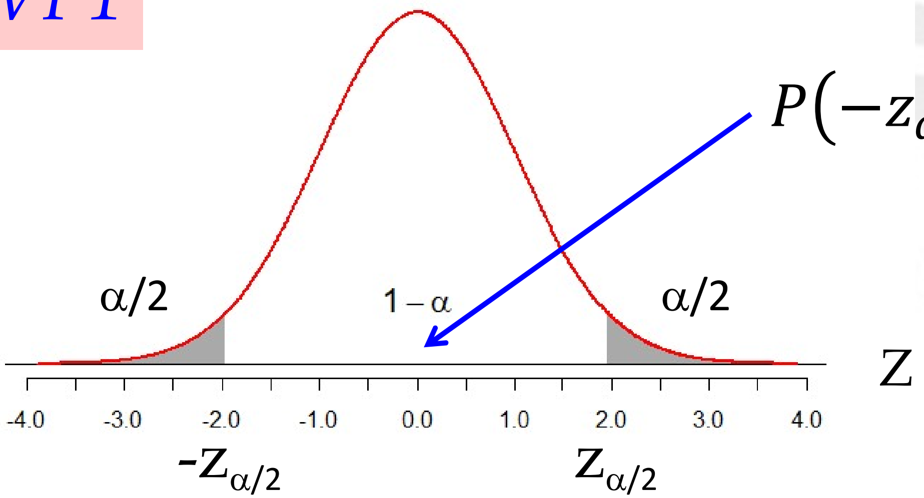
$$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

Para construir el intervalo de confianza para μ, se utiliza como variable pivotal la variable **Z**.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

PVP1

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$



Intervalo para la media con σ^2 conocida

pasos algebraicos

PVP2

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2}\,\sigma/\sqrt{n} \leq \overline{X}-\mu \leq z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$$P\left(-\overline{X}-z_{\alpha/2}\,\sigma/\sqrt{n} \leq -\mu \leq -\overline{X}+z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$$P\left(\overline{X}+z_{\alpha/2}\,\sigma/\sqrt{n} \geq \mu \geq \overline{X}-z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$$P\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Intervalo para la media con σ^2 conocida

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Se selecciona la muestra aleatoria de la población; y a partir de las observaciones o datos se calcula el valor de media de la muestra y se reemplaza en la expresión del intervalo aleatorio:

PVP3

$$\bar{x}_{ob} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}_{ob} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

estimación de μ mediante un intervalo del $100(1-\alpha)\%$ de confianza

Intervalo para la media con σ^2 conocida

límite inferior li

media poblacional

límite superior ls

$$\underbrace{\overline{x}_{ob} - z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}}_{li} \leq \mu \leq \underbrace{\overline{x}_{ob} + z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}}_{ls}$$

media muestral observada

valor crítico de la variable normal estándar Z que deja una probabilidad $\alpha/2$ a la derecha

desvío estándar poblacional (conocido)

Observaciones

$$\overline{x}_{ob} - z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}} \leq \mu \leq \overline{x}_{ob} + z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}$$

- El intervalo es simétrico y está centrado en el valor de la media de la muestra.
- En la medida que el tamaño n de la muestra crece, la amplitud del intervalo de confianza disminuye.
- En la medida que el nivel de confianza $100(1-\alpha)\%$ crece, la amplitud del intervalo de confianza aumenta.
- Si la dispersión de la población $\sigma(x)$ es mayor, la amplitud del intervalo de confianza será mayor.

Ejemplo

En un relevamiento realizado en la formación Los Molles se recolectaron 60 rocas de halita y se estudió la densidad de los cristales de cloruro de sodio (en gr/cm3).

2,1	2,14	2,32	2,38	2,39	2,46	2,48	2,5	2,53	2,55
2,57	2,58	2,59	2,62	2,67	2,69	2,73	2,74	2,75	2,75
2,76	2,76	2,77	2,79	2,82	2,84	2,84	2,89	2,89	2,91
2,91	2,95	2,95	2,95	2,99	2,99	2,99	2,99	3,01	3,03
3,03	3,07	3,08	3,08	3,08	3,08	3,08	3,09	3,1	3,11
3,13	3,14	3,18	3,18	3,19	3,21	3,23	3,24	3,26	3,35

Suponiendo que el desvío estándar poblacional $\sigma(x) = 0,3$ gr/cm3, obtenga una estimación por intervalo de la media poblacional μ. Utilice una confianza del 95%.

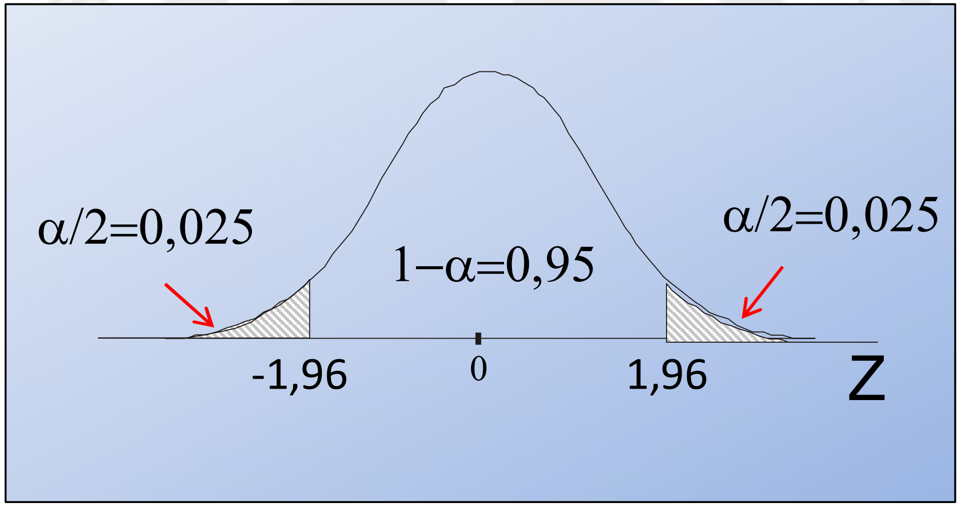
Ejemplo

El nivel de confianza es 95%: $\alpha = 0{,}05 \Rightarrow \alpha/2 = 0{,}025$

La expresión del intervalo es:

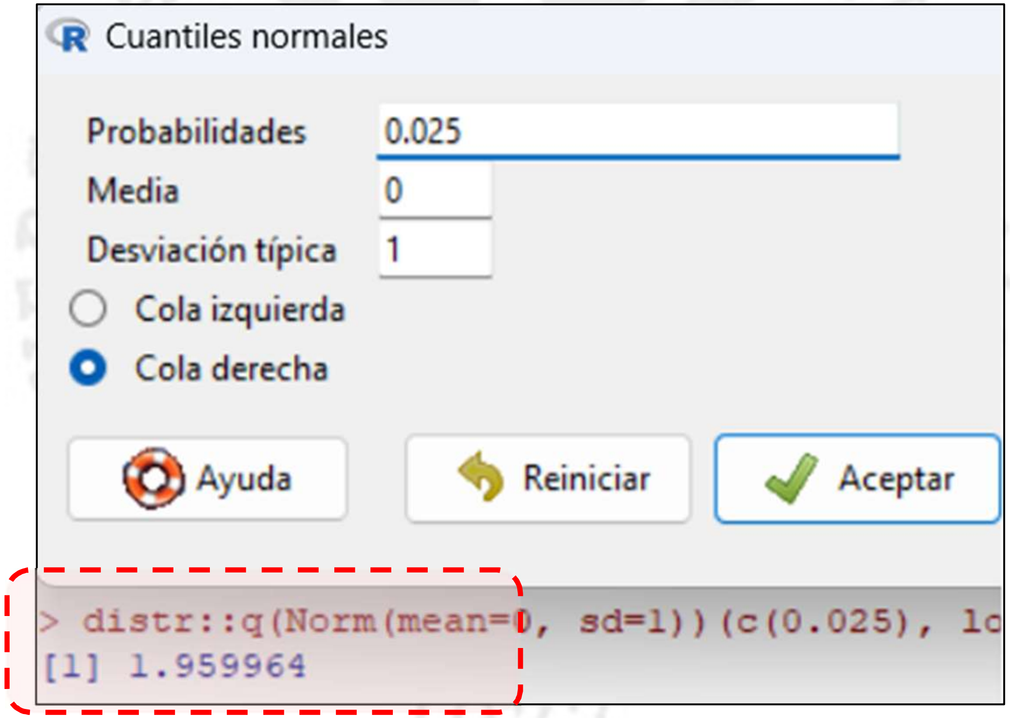
$$\bar{x}_{ob} - z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}$$

Para hallar el valor crítico busco el valor de Z que deja una probabilidad 0,025 por arriba o por debajo:



Uso de R Commander

Para hallar el valor crítico busco el valor de Z que deja una probabilidad 0,025 por arriba o por debajo:



$\bar{x} = 2{,}858$ estimación puntual de μ.

Ejemplo

Reemplazando en la expresión del intervalo:

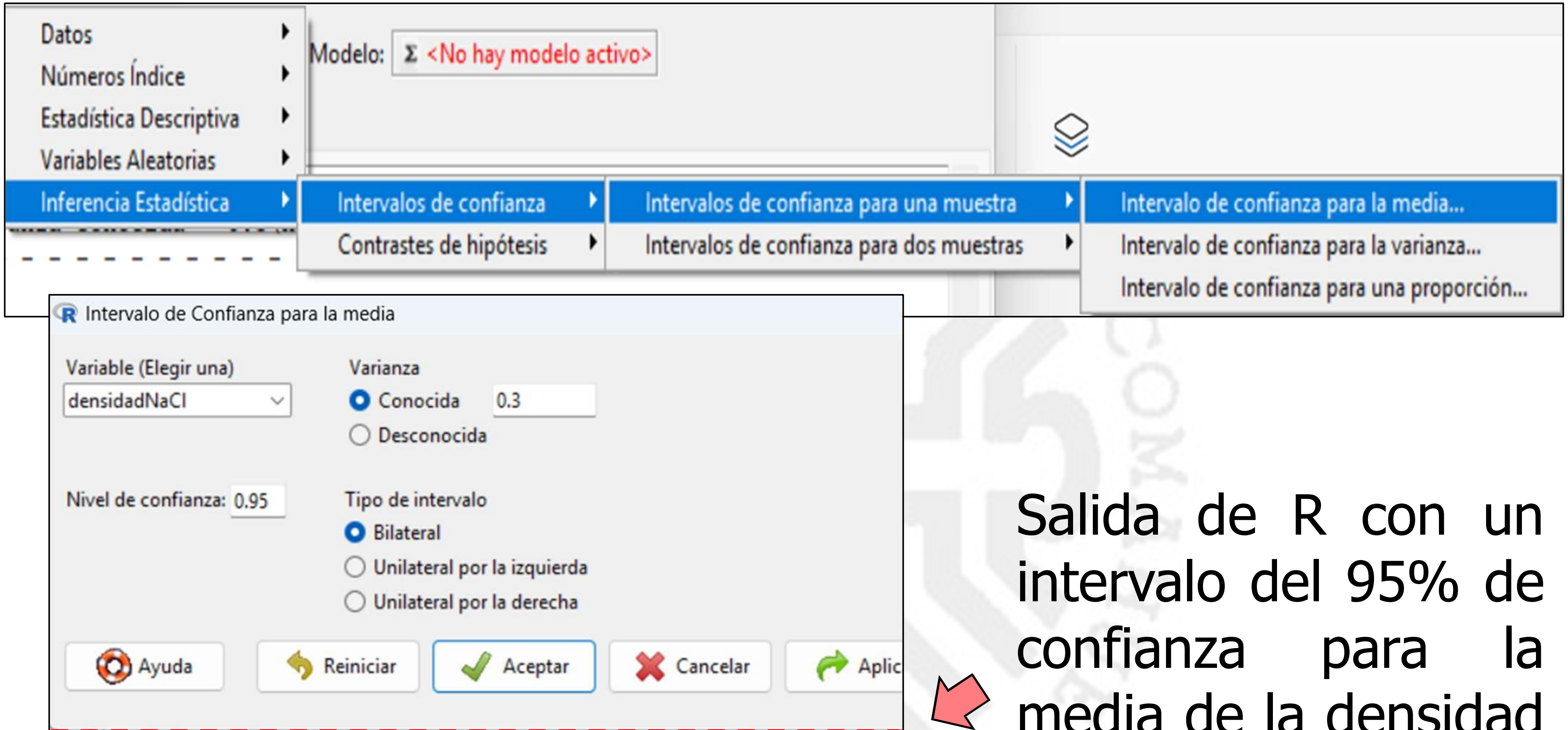
$$\bar{x}_{ob} - z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}$$

$$2{,}858 - 1{,}96\frac{0{,}3}{\sqrt{60}} \leq \mu \leq 2{,}858 + 1{,}96\frac{0{,}3}{\sqrt{60}}$$

$$\boxed{2{,}782 \leq \mu \leq 2{,}933}$$

Interpretación: hay un 95% de confianza que el intervalo (2,782 ; 2,933) contiene a la *media* poblacional de la densidad de cloruro de sodio.

Uso de RcmdrPlugin.TeachStat




```
Intervalo de Confianza para la media con varianza conocida = 0.3
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Tipo de intervalo: Bilateral
Nivel de confianza: 95%
Variable: densidadNaCl
Estimador muestral: mean of Dataset$densidadNaCl 2.858
Intervalo: ( 2.71941 , 2.99659 )
```

Salida de R con un intervalo del 95% de confianza para la media de la densidad de cloruro de sodio (varianza conocida)

Observaciones

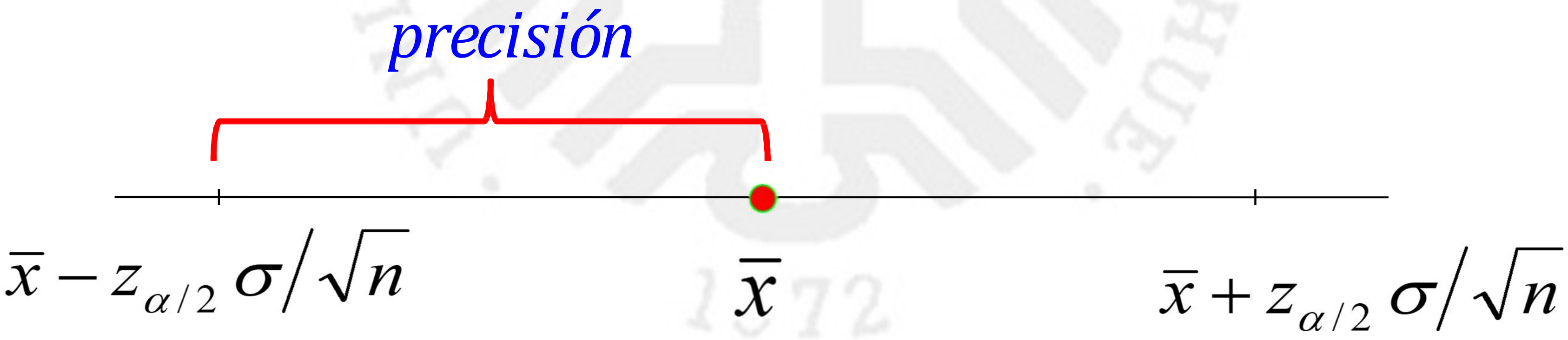
- Debe observarse que la media poblacional de densidad de cloruro de sodio puede ser:
  - cualquiera de los valores dentro del intervalo de confianza.
  - un valor fuera del mismo.
- Sin embargo, en virtud del método utilizado, tenemos una confianza del *100(1-α)%* de que este intervalo contiene a μ..
- La estimación por intervalo es considerablemente más rica que la estimación puntual, dado que informa la *confiabilidad* de la estimación (1-α)% y también la *precisión* de la estimación.

Precisión

Se llama *precisión* a la *semiamplitud* del intervalo de confianza. Menor *semiamplitud* implica mayor *precisión*.

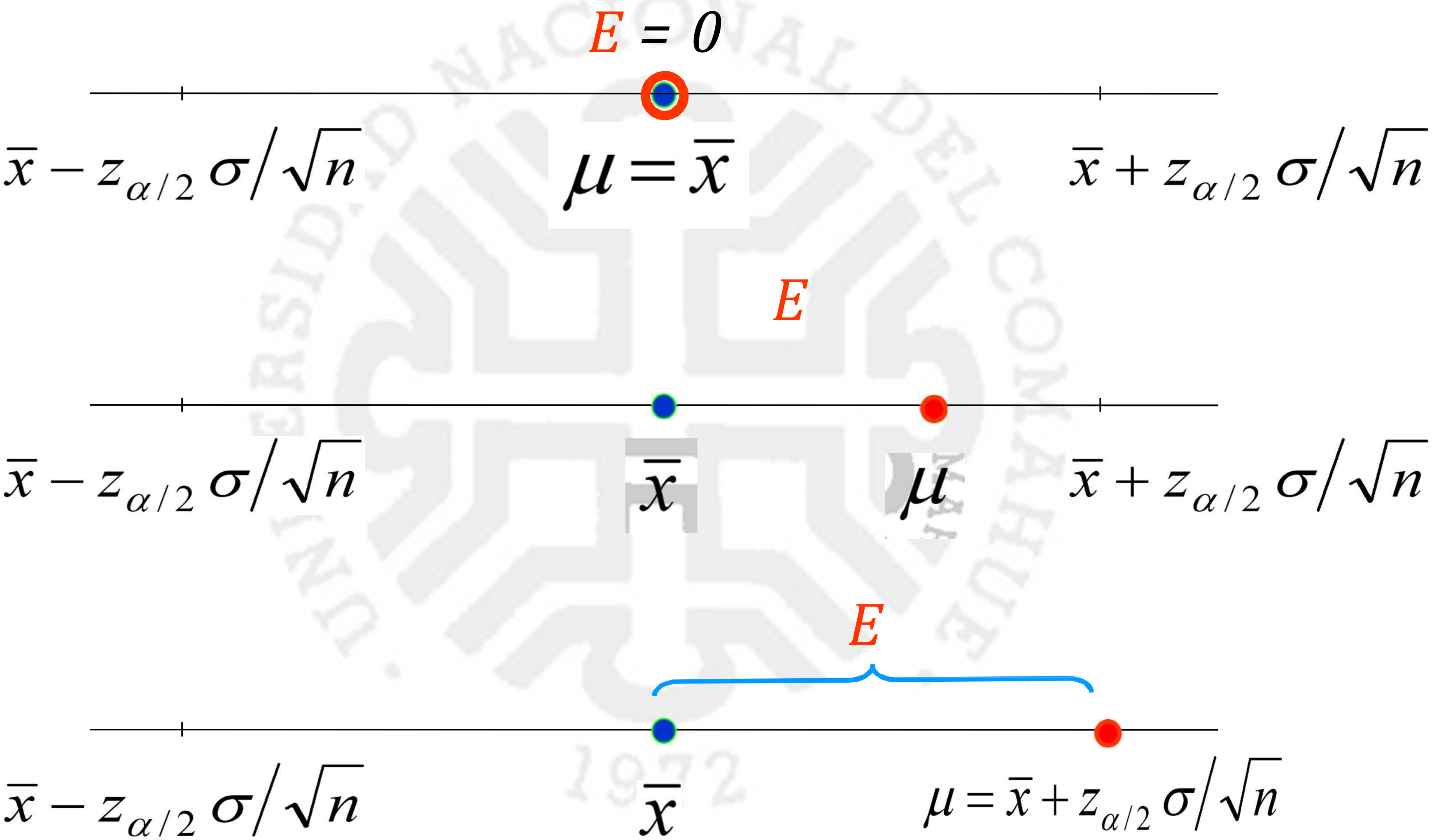
$$z_{\alpha/2} \cdot \frac{\sigma(x)}{\sqrt{n}}$$

$$\bar{x}_{ob} - z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + z_{\alpha/2}\frac{\sigma(x)}{\sqrt{n}}$$



Precisión y error de muestreo

$$E = |\overline{X} - \mu|$$



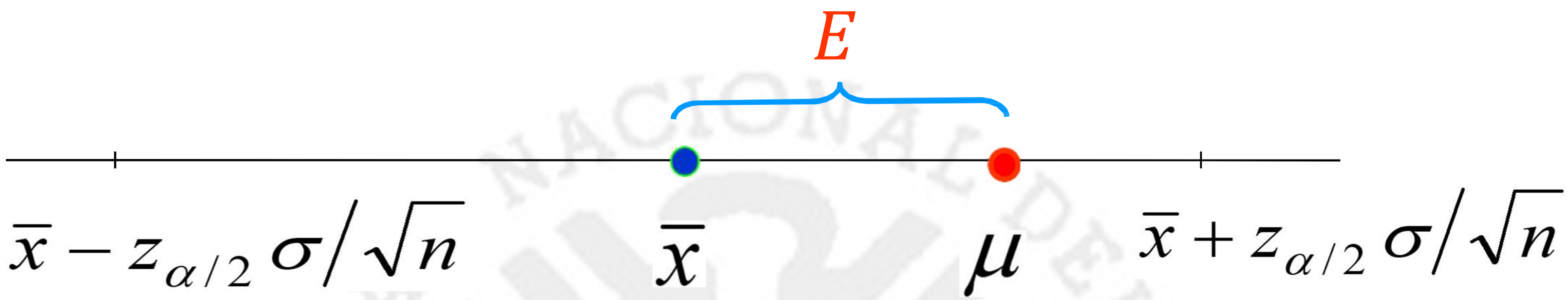
Precisión y error de muestreo

$$\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n} \qquad \overbrace{\bar{x} \qquad\qquad \mu}^{E} \qquad \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n}$$

Interpretación

- Se espera que en el *100(1- α%)* de las muestras el *error de muestreo* sea menor o igual a la *semiamplitud*.
- Hay un *100(1- α%)* de confianza de que el *error de muestreo* es como máximo igual a la *semiamplitud*.

Precisión y error de muestreo



- En el ejemplo de la concentración de cloruro de sodio, el intervalo del 95% de confianza para la media es: (2,782 ; 2,933)
- la *semiamplitud* se puede calcular como: (*ls*-*li*)/2 = (2,933-2,782)/2 = 0,0755
- con un 95% de confianza, esperamos que el error de muestreo sea como máximo de 0,0755 gr/cm3.

Intervalo para la media con σ^2 desconocida

Condiciones

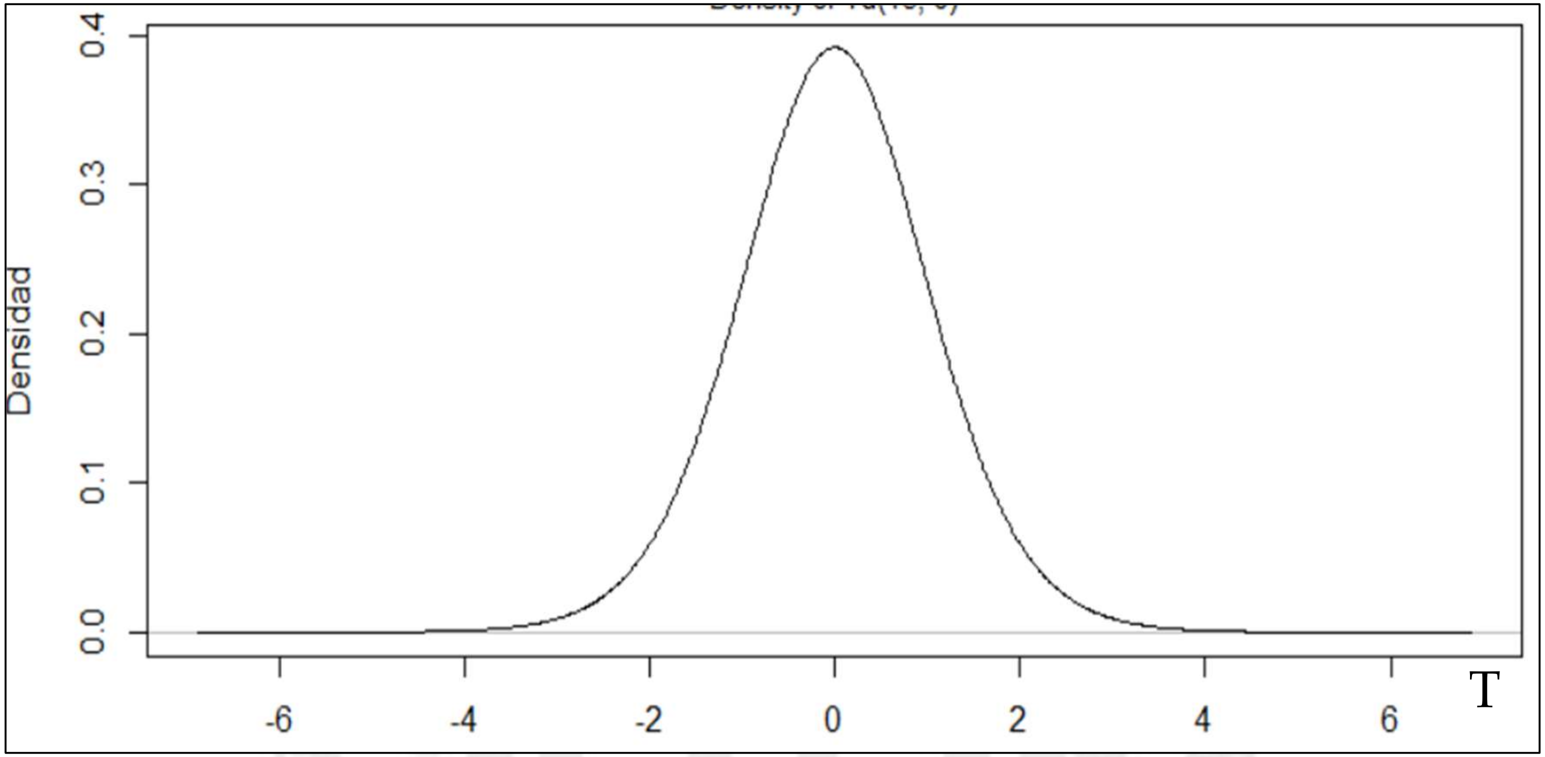
1. La muestra aleatoria se extrae de una población normal.

2. No se conoce la varianza de la población $\sigma^2(x)$, el cual debe ser estimado por medio de la varianza muestral $S^2(x)$.

Se emplea como variable pivotal la distribución *t-Student* con *n-1 grados de libertad*.

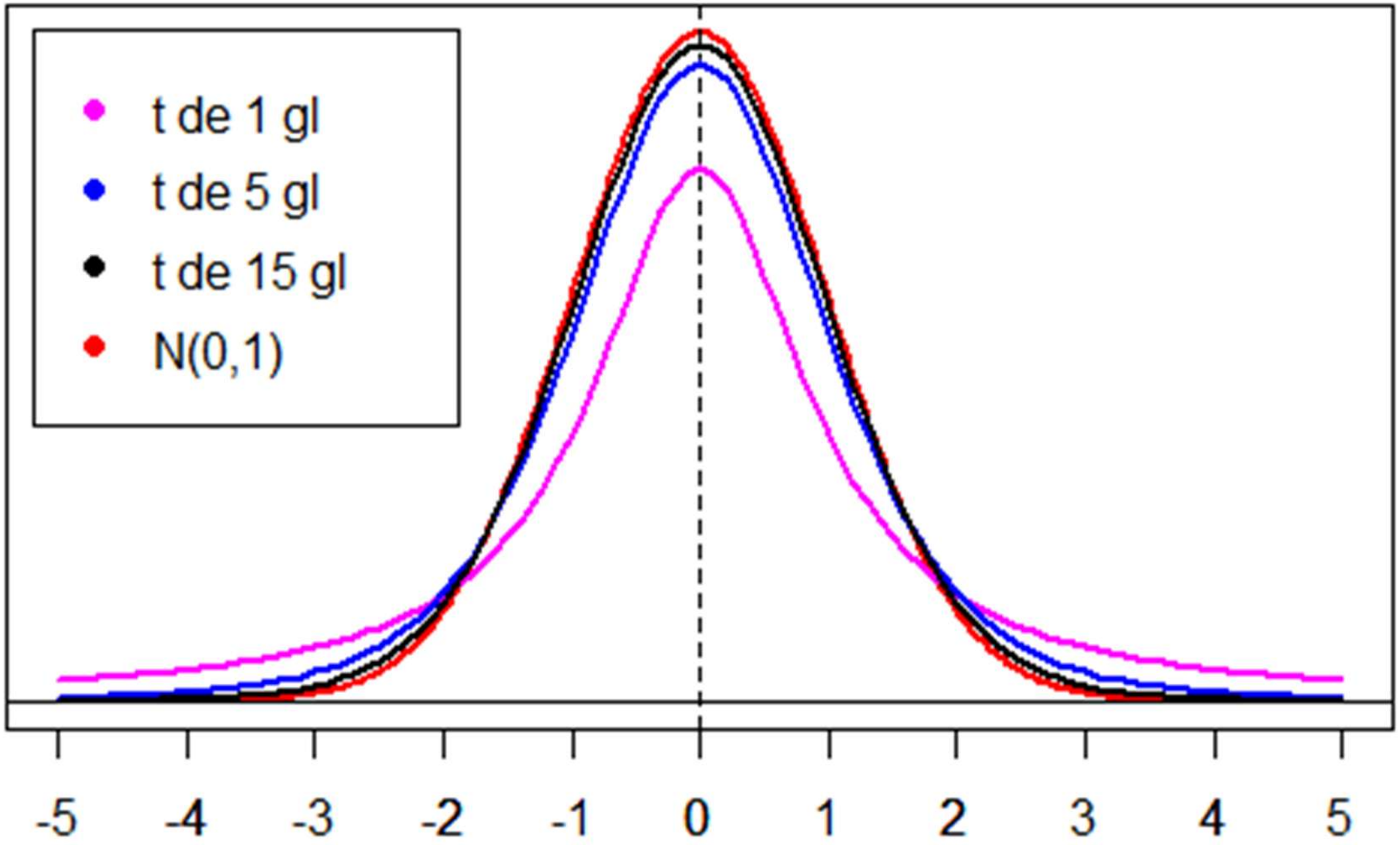
$$T = \frac{\bar{X} - \mu}{S(x) / \sqrt{n}} \sim t_{n-1}$$

Distribución t-Student



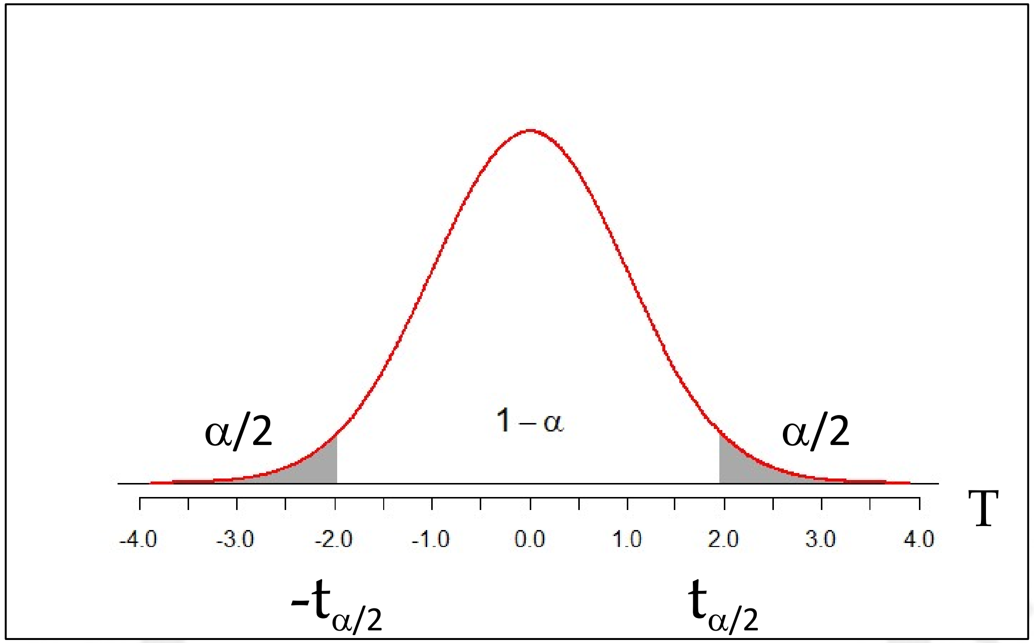
- La función de densidad de la variable T tiene forma de campana, es simétrica y está centrada en cero.
- Una distribución T en particular es especificada por su parámetro denominado "*grados de libertad* v".

Distribución t-Student



- Al aumentar los *grados de libertad*, la dispersión de la distribución t disminuye.
- La distribución t converge a la $N(0,1)$ cuando $n \rightarrow \infty$

Intervalo para la media con σ^2 desconocida



PVP1

$$P\left(-t_{\alpha/2} \leq T \leq t_{\alpha/2}\right) = 1 - \alpha$$

PVP2

$$P\left(-t_{\alpha/2} \leq \frac{\overline{X} - \mu}{S(x)/\sqrt{n}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(-t_{\alpha/2}\, S(x)/\sqrt{n} \leq \overline{X} - \mu \leq t_{\alpha/2}\, S(x)/\sqrt{n}\right) = 1 - \alpha$$

Intervalo para la media con σ^2 desconocida

$$P\left(-\overline{X} - t_{\alpha/2}\, S(x)/\sqrt{n} \le -\mu \le -\overline{X} + t_{\alpha/2}\, S(x)/\sqrt{n}\right) = 1 - \alpha$$

$$P\left(\overline{X} + t_{\alpha/2}\, S(x)/\sqrt{n} \ge \mu \ge \overline{X} - t_{\alpha/2}\, S(x)/\sqrt{n}\right) = 1 - \alpha$$

$$P\left(\overline{X} - t_{\alpha/2}\frac{S(x)}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2}\frac{S(x)}{\sqrt{n}}\right) = 1 - \alpha$$

PVP3

estimación de μ por intervalo (σ^2 desconocida)

$$\overline{x_{ob}} - t_{\alpha/2}\frac{s(x)}{\sqrt{n}} \le \mu \le \overline{x}_{ob} + t_{\alpha/2}\frac{s(x)}{\sqrt{n}}$$

Intervalo para la media con σ^2 desconocida

$$P\left(\bar{X} - t_{\alpha/2}\frac{S(x)}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S(x)}{\sqrt{n}}\right) = 1 - \alpha$$

límite inferior (*li*)

media poblacional

límite superior (*ls*)

$$\bar{x}_{ob} - t_{\alpha/2}\frac{s(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + t_{\alpha/2}\frac{s(x)}{\sqrt{n}}$$

media muestral observada

valor de la variable t que de por abajo/arriba una probabilidad $\alpha/2$

desvío estándar muestral observado

Intervalo para la media con σ^2 desconocida

$$\bar{x}_{ob} - t_{\alpha/2}\frac{s(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + t_{\alpha/2}\frac{s(x)}{\sqrt{n}}$$

- El intervalo es simétrico y está centrado en el valor de la media muestral;
- En la medida que el tamaño n de la muestra crece, la amplitud del intervalo de confianza disminuye;
- En la medida que el nivel de confianza $(1\text{-}\alpha)$ crece, la amplitud del intervalo de confianza aumenta;
- La amplitud del intervalo de confianza depende de $S(x)$ que es variable aleatoria.

Ejemplo

Consideremos nuevamente el estudio de la densidad de los cristales de cloruro de sodio (en gr/cm3) en la formación Los Molles.

No se conoce la media μ ni el desvío estándar $\sigma(x)$ de esta población.

2,1	2,14	2,32	2,38	2,39	2,46	2,48	2,5	2,53	2,55
2,57	2,58	2,59	2,62	2,67	2,69	2,73	2,74	2,75	2,75
2,76	2,76	2,77	2,79	2,82	2,84	2,84	2,89	2,89	2,91
2,91	2,95	2,95	2,95	2,99	2,99	2,99	2,99	3,01	3,03
3,03	3,07	3,08	3,08	3,08	3,08	3,08	3,09	3,1	3,11
3,13	3,14	3,18	3,18	3,19	3,21	3,23	3,24	3,26	3,35

Utilizaremos el desvío estándar muestral $s(x)$, para obtener una estimación por intervalo de la media poblacional μ. Usaremos una confianza del 95%.

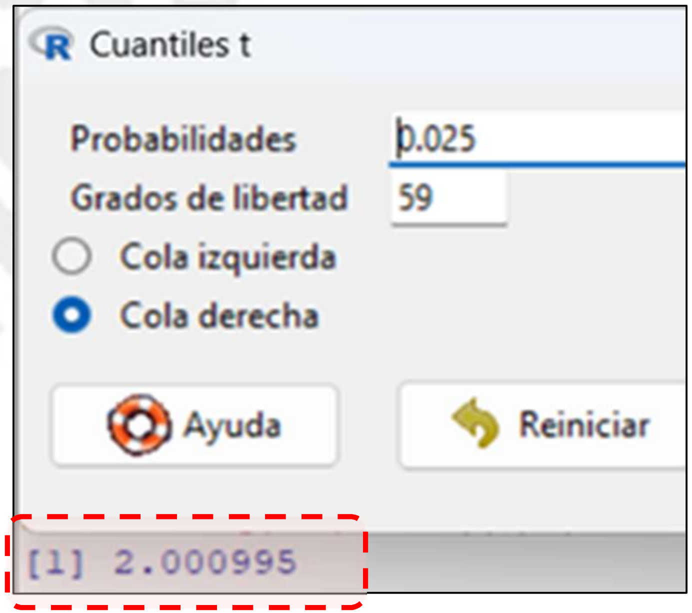
Condiciones

- Debemos suponer que la muestra se extrajo de una población con distribución normal, o debemos utilizar algún método estadístico y chequear que este supuesto es razonable.
- No se conoce la varianza poblacional $\sigma^2(x)$.

En la muestra observada resultó:

$\bar{x} = 2{,}858$ $\qquad s(x) = 0{,}286$

n=60



valor crítico de T que deja una probabilidad 0,025 por arriba.

Ejemplo

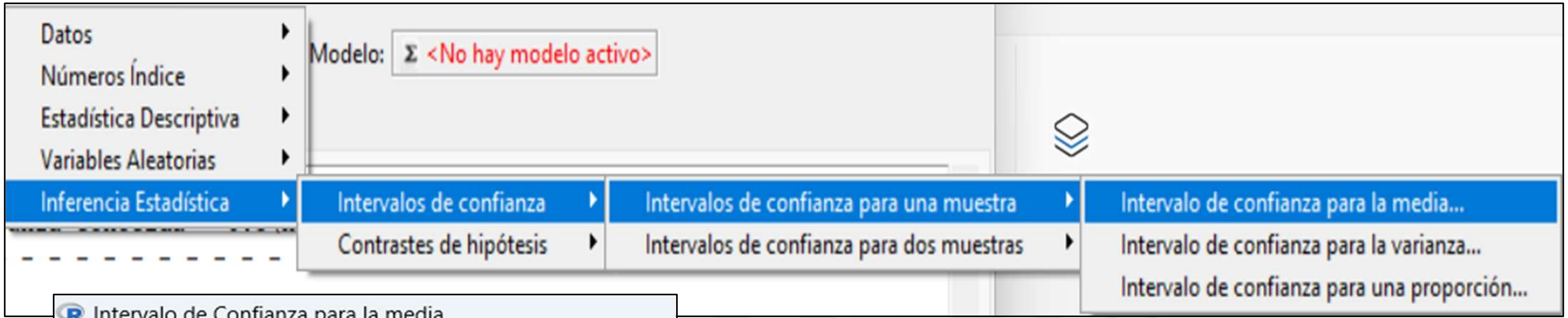
$$\bar{x}_{ob} - t_{\alpha/2}\frac{s(x)}{\sqrt{n}} \leq \mu \leq \bar{x}_{ob} + t_{\alpha/2}\frac{s(x)}{\sqrt{n}}$$

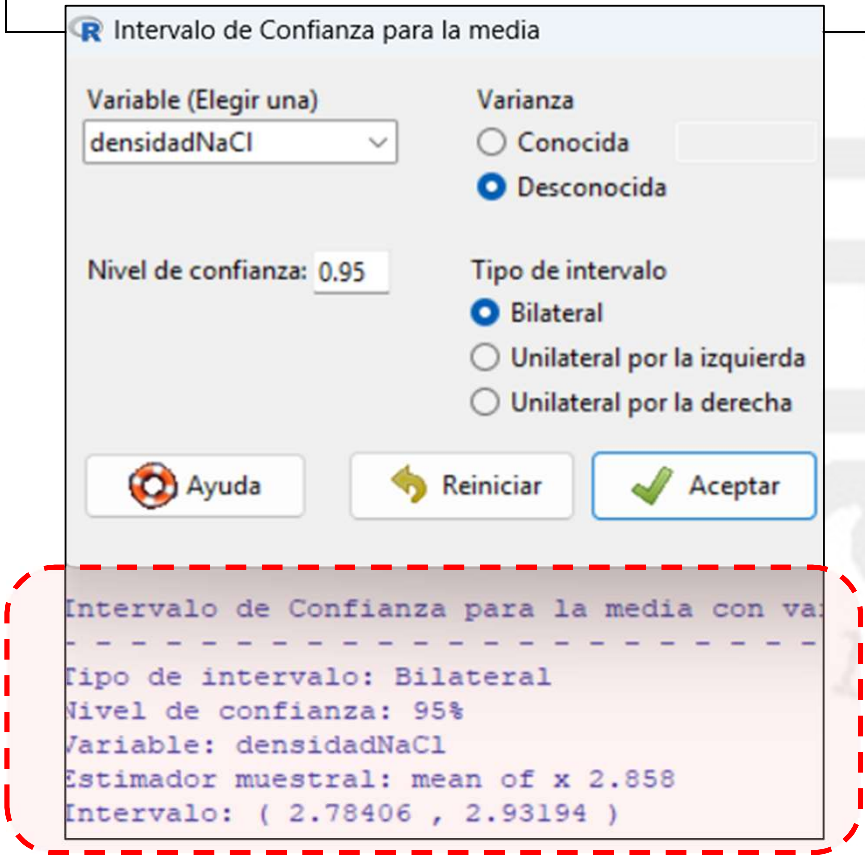
$$2{,}858 - 2\frac{0{,}286}{\sqrt{60}} \leq \mu \leq 2{,}858 + 2\frac{0{,}286}{\sqrt{60}}$$

$$\boxed{2{,}784 \leq \mu \leq 2{,}931}$$

Interpretación: hay un 95% de confianza que el intervalo (2,784 ; 2,931) contiene a la *media* poblacional de la densidad de cloruro de sodio.

Uso de RcmdrPlugin.TeachStat





Salida de R con un intervalo del 95% de confianza para la media de la densidad de cloruro de sodio (varianza desconocida)

Intervalo de confianza para la varianza

Condición:

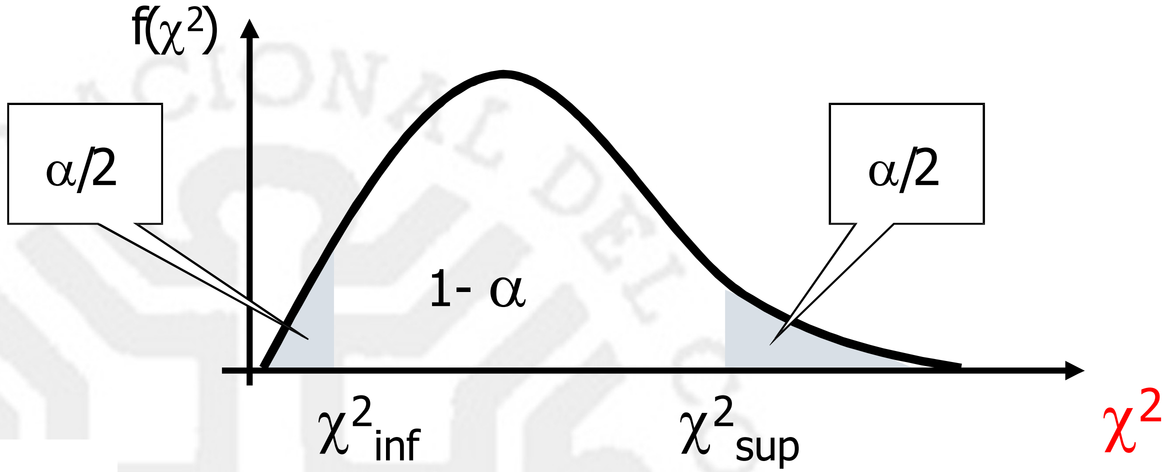
- La muestra aleatoria debe ser extraída de una población con distribución normal.

Para construir un intervalo de confianza para σ^2 se utiliza la distribución *Chi-Cuadrado* con *n-1 grados de libertad* y la siguiente variable pivotal:

$$\boxed{\frac{(n-1)S^2(x)}{\sigma^2(x)} \sim \chi^2_{n-1}}$$

La distribución *Chi-Cuadrado* define una familia de curvas, cada una de ellas para un número distinto de grados de libertad *v*.

Intervalo de confianza para la varianza



PVP1

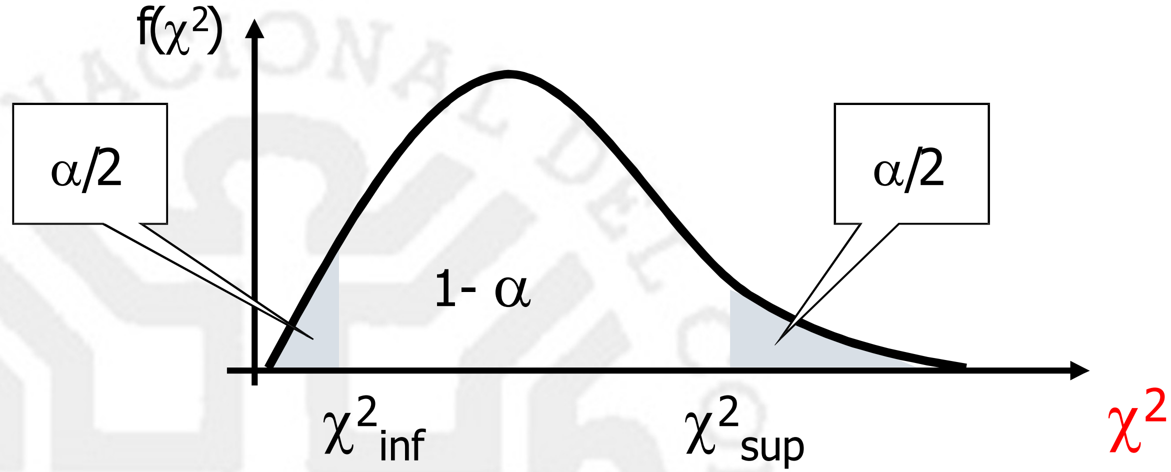
$$P\left(\chi^2_{\inf} \leq \chi^2 \leq \chi^2_{\sup}\right) = 1-\alpha$$

PVP2

$$P\left(\chi^2_{\inf} \leq \frac{(n-1)S^2(x)}{\sigma^2(x)} \leq \chi^2_{\sup}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2(x)}{\chi^2_{\sup}} \leq \sigma^2(x) \leq \frac{(n-1)S^2(x)}{\chi^2_{\inf}}\right) = 1-\alpha$$

Intervalo de confianza para la varianza



PVP3

$$\frac{(n-1)s^2_{ob}(x)}{\chi^2_{\sup}} \leq \sigma^2(x) \leq \frac{(n-1)s^2_{ob}(x)}{\chi^2_{\inf}}$$

Intervalo de confianza para la varianza σ^2 de una población normal.

Intervalo de confianza para la varianza

$$\frac{(n-1)s_{ob}^2(x)}{\chi_{\sup}^2} \leq \sigma^2(x) \leq \frac{(n-1)s_{ob}^2(x)}{\chi_{\inf}^2}$$

- En la medida que el tamaño n de la muestra crece, la amplitud del intervalo de confianza disminuye, es decir aumenta la precisión.
- En la medida que el nivel de confianza crece, la amplitud del intervalo de confianza aumenta, es decir disminuye la precisión.

Ejemplo

Consideremos nuevamente el estudio de la densidad de los cristales de cloruro de sodio (en gr/cm3) en la formación Los Molles.

Deseamos estimar la varianza poblacional $\sigma^2(x)$ y el desvío estándar $\sigma(x)$ en esa formación mediante un intervalo del 95% de confianza..

- Debemos suponer que la muestra se extrajo de una población con distribución normal.
- Veremos más adelante métodos estadísticos para chequear que la suposición de normalidad sea razonable.

Ejemplo

En la muestra observada resultó: $s(x) = 0{,}286$ n=60

$$\frac{(n-1)s_{ob}^2(x)}{\chi^2_{\sup}} \le \sigma^2(x) \le \frac{(n-1)s_{ob}^2(x)}{\chi^2_{\inf}}$$

$$\frac{59 \cdot 0{,}081796}{82{,}11} \le \sigma^2(x) \le \frac{59 \cdot 0{,}081796}{39{,}66}$$

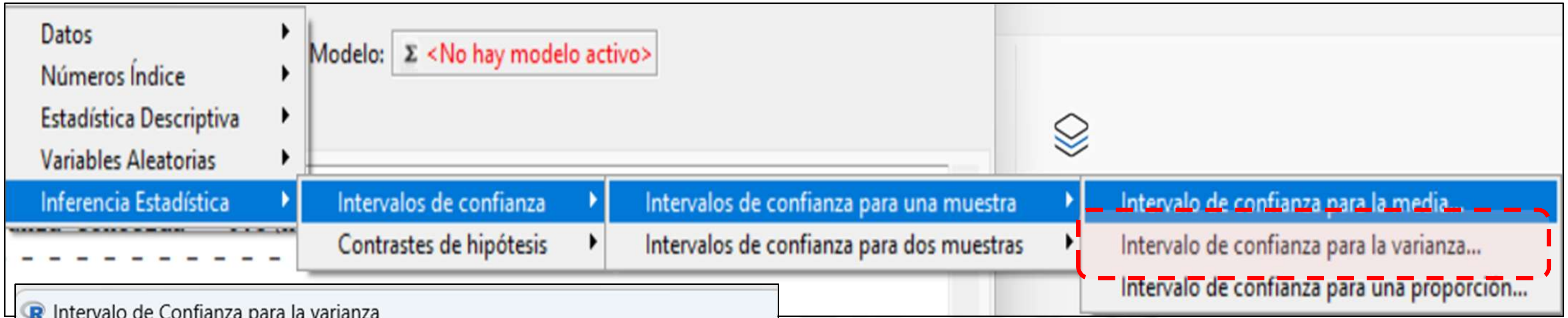
$$0{,}058861 \le \sigma^2(x) \le 0{,}121869$$

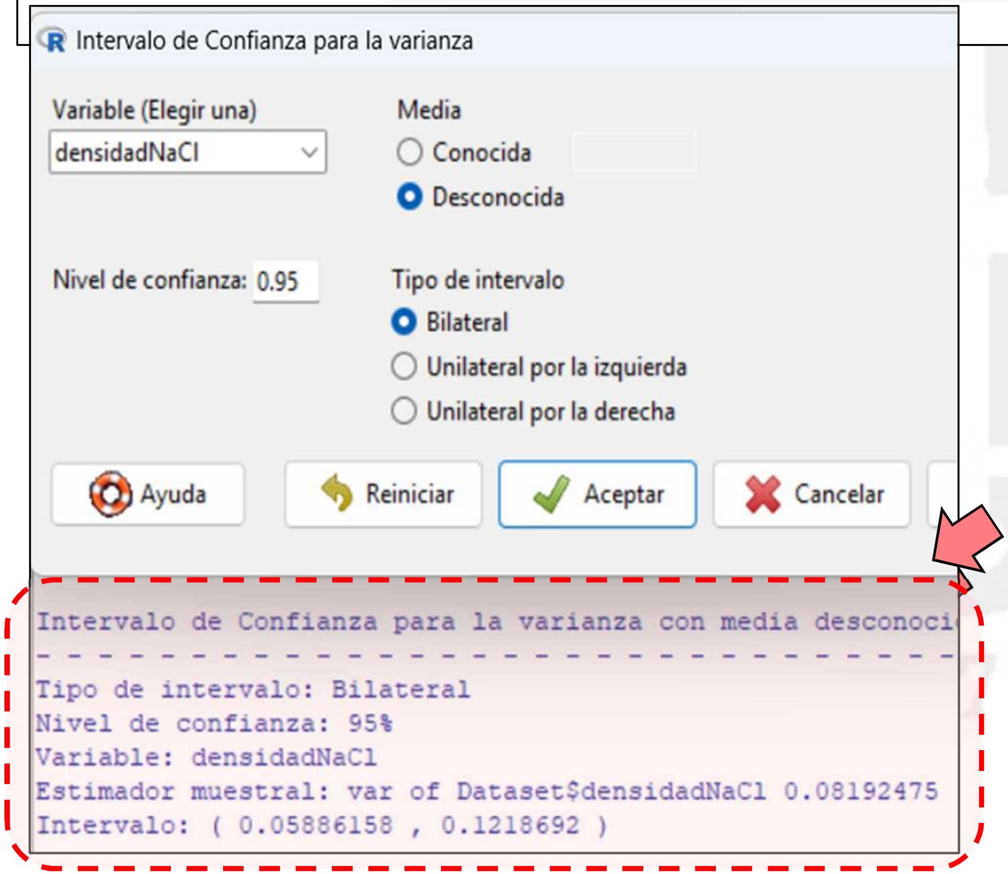
Aplicando raíz cuadrada, el intervalo de confianza para el *desvío estándar* es:

$$0{,}2426 \le \sigma(x) \le 0{,}3491$$

Hay un 95% de confianza que el intervalo (0,2426 ; 0,3491) contiene el *desvío* poblacional de la densidad de cloruro de sodio.

Uso de RcmdrPlugin.TeachStat





Salida de R con un intervalo del 95% de confianza para la varianza poblacional de concentración de cloruro de sodio.

Intervalo para la proporción poblacional

El estimador puntal de la proporción poblacional π es: $p = \frac{X}{n}$

X: cantidad de observaciones que pertenecen a la clase de interés en la muestra (éxitos)

$$X \sim B(n, \pi)$$

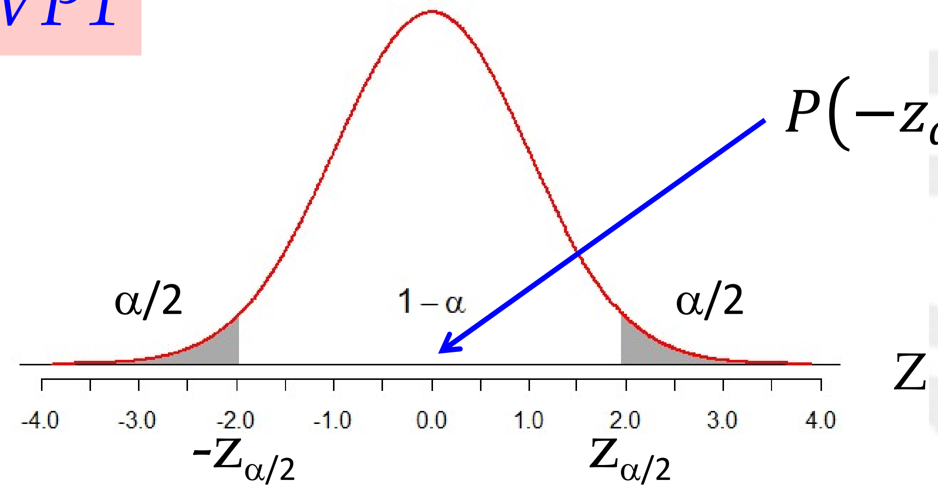
El TCL indica que si n es suficientemente grande ($n \geq 100$) $\Rightarrow$ p la proporción muestral de éxitos tiene una distribución *aproximadamente normal*.

$$p \xrightarrow{D} N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

Intervalo para la proporción poblacional

$$p \xrightarrow{D} N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right) \Rightarrow Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

PVP1



$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$$

PVP2

$$P\left(h - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq h + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

Intervalo para la proporción poblacional

$$P\left(p - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq p + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1-\alpha$$

$$P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

PVP3

$$\boxed{p_{ob} - z_{\alpha/2}\sqrt{\frac{p_{ob}(1-p_{ob})}{n}} \leq \pi \leq p_{ob} + z_{\alpha/2}\sqrt{\frac{p_{ob}(1-p_{ob})}{n}}}$$

Estimación por intervalo de la proporción poblacional π

Ejemplo: En una muestra de 500 personas de una región geográfica, se encontró que 285 de ellas reciclan plástico. Estimar el porcentaje de personas que reciclan plástico en esa región utilizando un intervalo del 95% de confianza.

n=500 X: Cantidad de personas que reciclan plástico

el valor crítico para 95% de confianza es: z = 1,96

$$p_{ob} = \frac{x_{ob}}{n} = \frac{285}{500} = 0{,}57$$

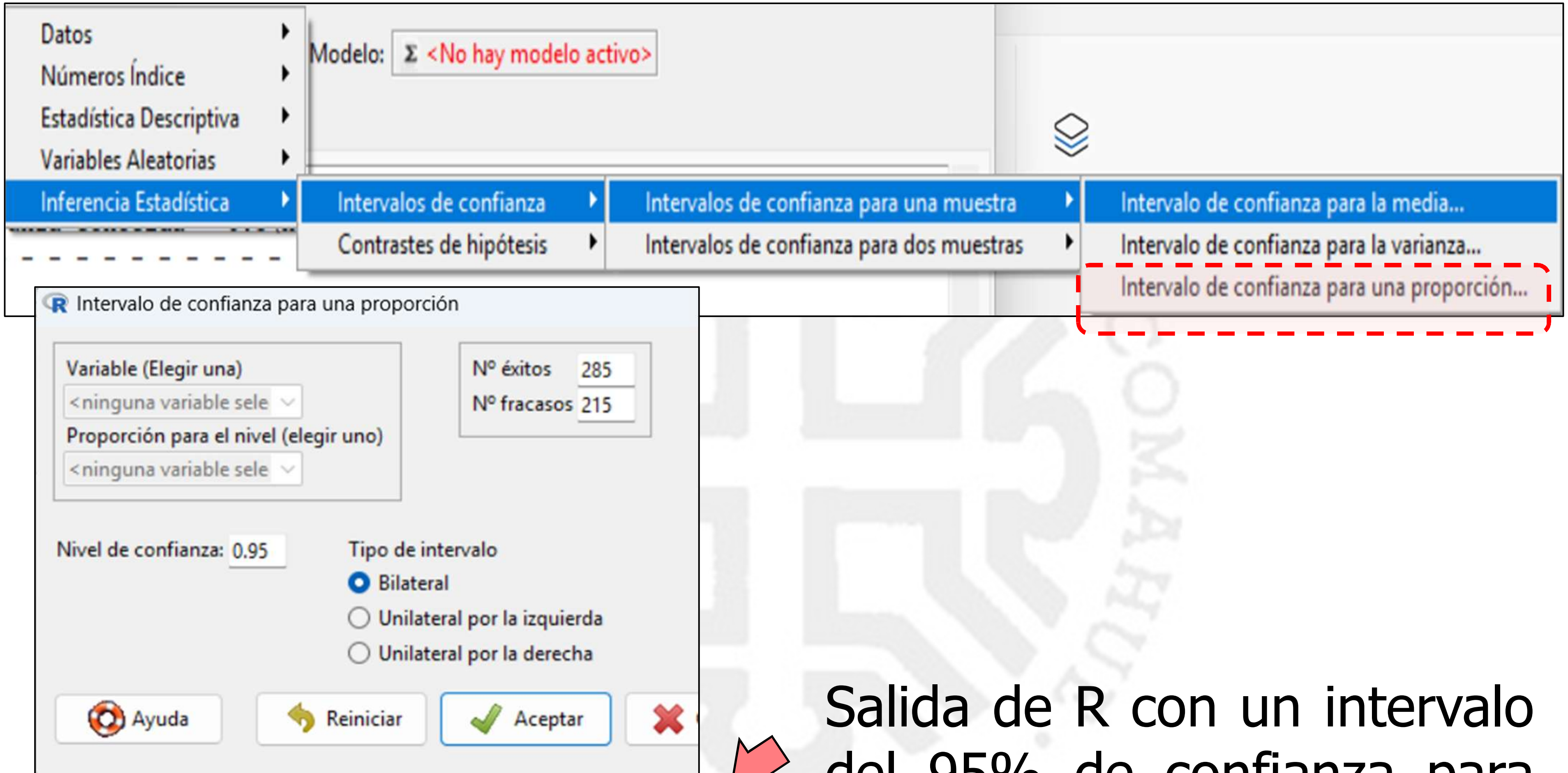
$$p_{ob} - z_{\alpha/2}\sqrt{\frac{p_{ob}(1-p_{ob})}{n}} \leq \pi \leq p_{ob} + z_{\alpha/2}\sqrt{\frac{p_{ob}(1-p_{ob})}{n}}$$

$$0{,}57 - 1{,}96\sqrt{\frac{0{,}57(1-0{,}57)}{500}} \leq \pi \leq 0{,}57 - 1{,}96\sqrt{\frac{0{,}57(1-0{,}57)}{500}}$$

$$0{,}5266 \leq \pi \leq 0{,}6134$$

En esa región, hay un 95% de confianza de que el intervalo (52,66%; 61,34%) contiene a la proporción de personas que reciclan plástico.

Uso de RcmdrPlugin.TeachStat




```
Intervalo de Confianza para una proporción
- - - - - - - - - - - - - - - - - - - - - - -
Tipo de intervalo: Bilateral
Nivel de confianza: 95%
Muestra: Nº éxitos = 285 -- Nº intentos = 500
Estimador muestral: proportion 0.57
Intervalo: ( 0.5266055 , 0.6133945 )
```

Salida de R con un intervalo del 95% de confianza para la proporción poblacional de rocas con cuarzo.