Unidad III

Parte II

Estimación de parámetros:

Estimación puntual

Estimación por intervalos

Intervalo de confianza para la media de una población

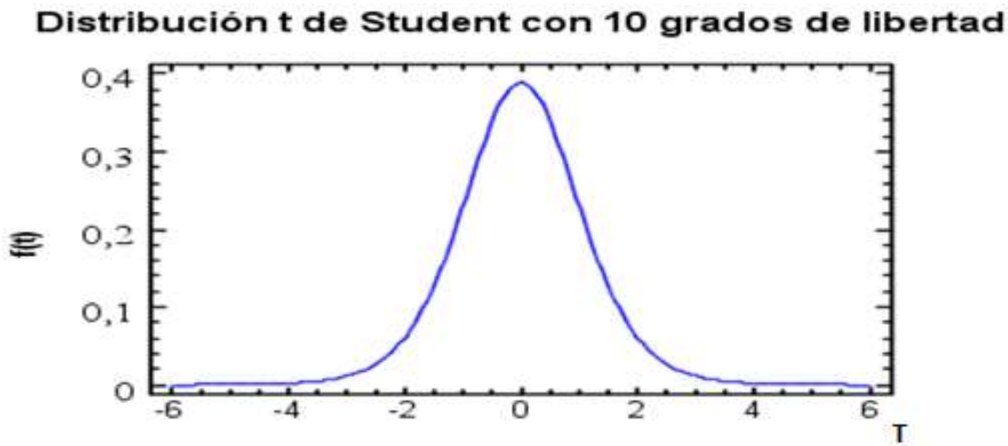
Caso II: **varianza σ^2 desconocida**

Dado que **X** $\sim$ N(μ,σ^2) con $\boldsymbol{\sigma^2}$ desconocida, se requiere de un intervalo donde no figure $\boldsymbol{\sigma^2}$. El valor de $\boldsymbol{\sigma^2}$ puede estimarse mediante el estimador puntual varianza muestral **S²**.

Como **X** $\sim$ **N(μ,σ^2),** el estadístico

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

(**tiene una distribución t -Student con n - 1 grados de libertad)**



Distribución t de Student

- Sea **X** una va **Normal** con media μ y varianza conocida σ^2. Sabemos que la v.a. media muestral $\overline{X}$, tiene distribución **Normal**, ie, $\overline{X} \sim N(\mu, \sigma^2/n)$, y

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

- Si la varianza σ^2 es **desconocida** ¿qué sucede con la distribución del estadístico si se reemplaza σ por **S**?

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño ***n*** de una v.a. **X ~** $N\left(\mu, \sigma^2\right)$

varianza σ^2 es desconocida, entonces la v.a. $T = \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}} \sim t_{\nu = n-1}$

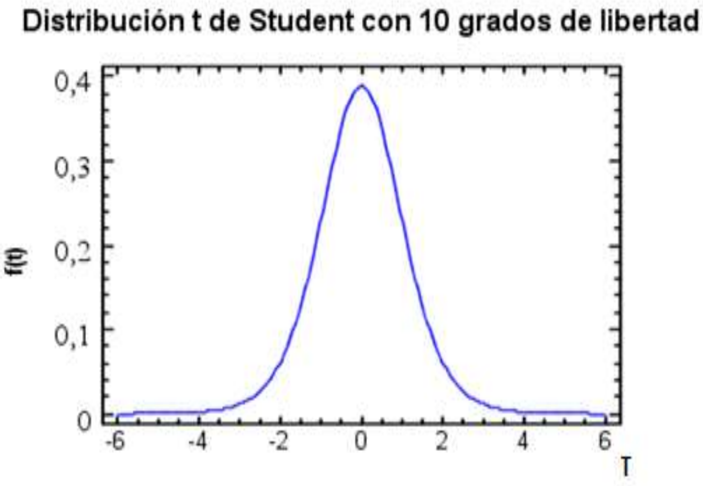
tiene una distribución **t de student con ν = n -1** grados de libertad.

Distribución t de Student

Si la v.a. T tiene una **distribución t de Student** de parámetro ν, se denota **T ~ t_ν** y una función de densidad está dada por:

$$f(t) = \frac{\Gamma\left(\frac{1+\nu}{2}\right)}{(\pi\,\nu)^{\frac{1}{2}}\;\Gamma\left(\frac{\nu}{2}\right)}\;\frac{1}{\left[\frac{t^2}{\nu}+1\right]^{\frac{\nu+1}{2}}} \qquad \text{si } -\infty < t < +\infty\,.$$

$$E(T) = 0 \qquad V(T) = \frac{\nu}{\nu-2}$$



Características de la función de densidad

•Esta caracterizada por un sólo parámetro ν, que recibe el nombre de *grados de libertad*.

•Es simétrica respecto del eje de ordenadas (recta **T = 0**).

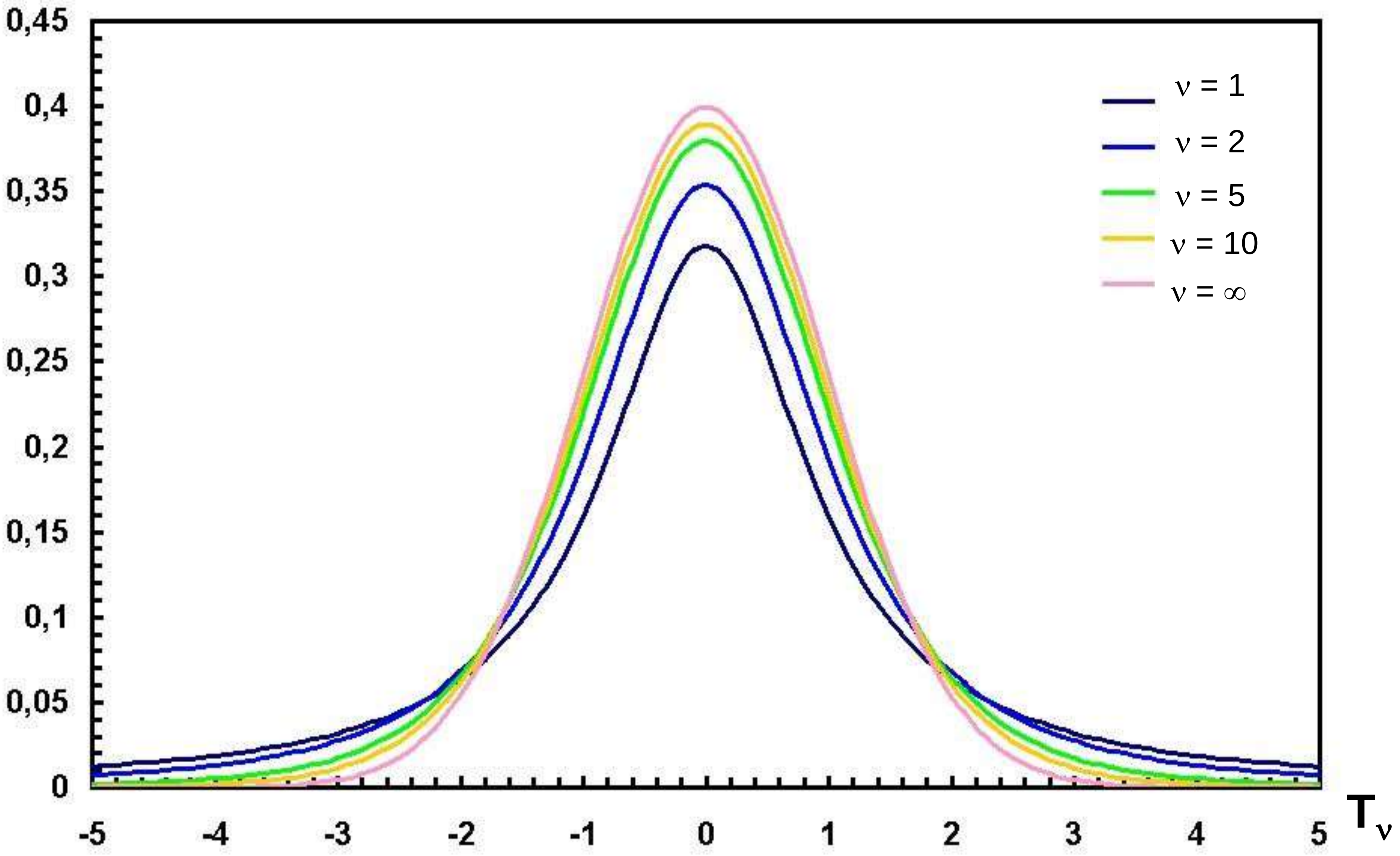
•El valor máximo de la función se alcanza en **μ = 0**.

•Está definida para toda la recta real.

•Conforme aumenta el valor de **ν** la distribución se aproxima a la distribución **Normal Estándar**.

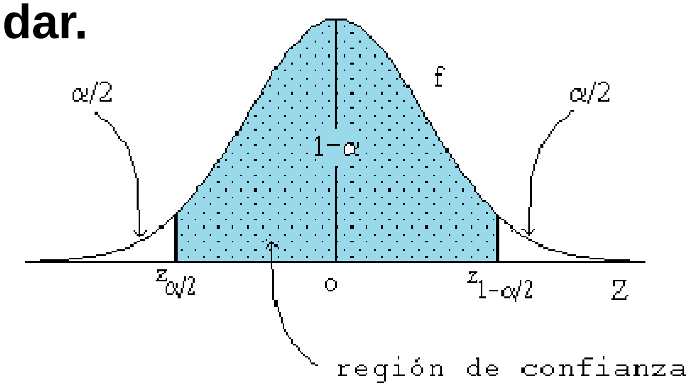
•Es más dispersa que la **Normal Estándar**.

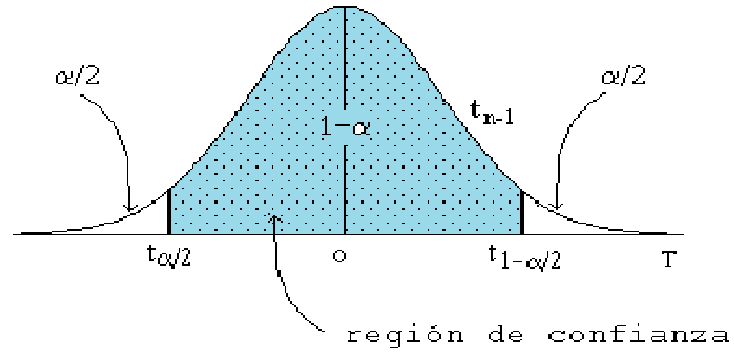
Distribución t de Student



Aproximación de la t-student a la Normal estándar

Conforme aumenta el valor de ν la distribución se aproxima a la **distribución normal estándar.**





Si el nivel de confianza = 1 - α = 0.95 entonces α = 0.05 y α**/2** = 0.025

ν	t_ν	z
30	2.04	1.95996
50	2.008	1.95996
100	1.9839	1.95996
1000	1.96234	1.95996
5000	1.96044	1.95996
10000	1.9602	1.95996

Manejo de la App t - student

I) Hallar:

(a) $P(t_{30} > 2.042)$
(b) $P(t_{23} \leq 3.485)$
(c) $P(t_{60} < 1.671)$

II) Para la distribución t-student hallar el valor de $t_{\nu, \alpha}$ de tal forma que:

(a) $P(t_{14} > t_{14, \alpha}) = 0.01$
(b) $P(t_8 < t_{8, \alpha}) = 0.975$
(c) $P(t_{40, \alpha} < t_{40} < 0.529) = 0.95.$

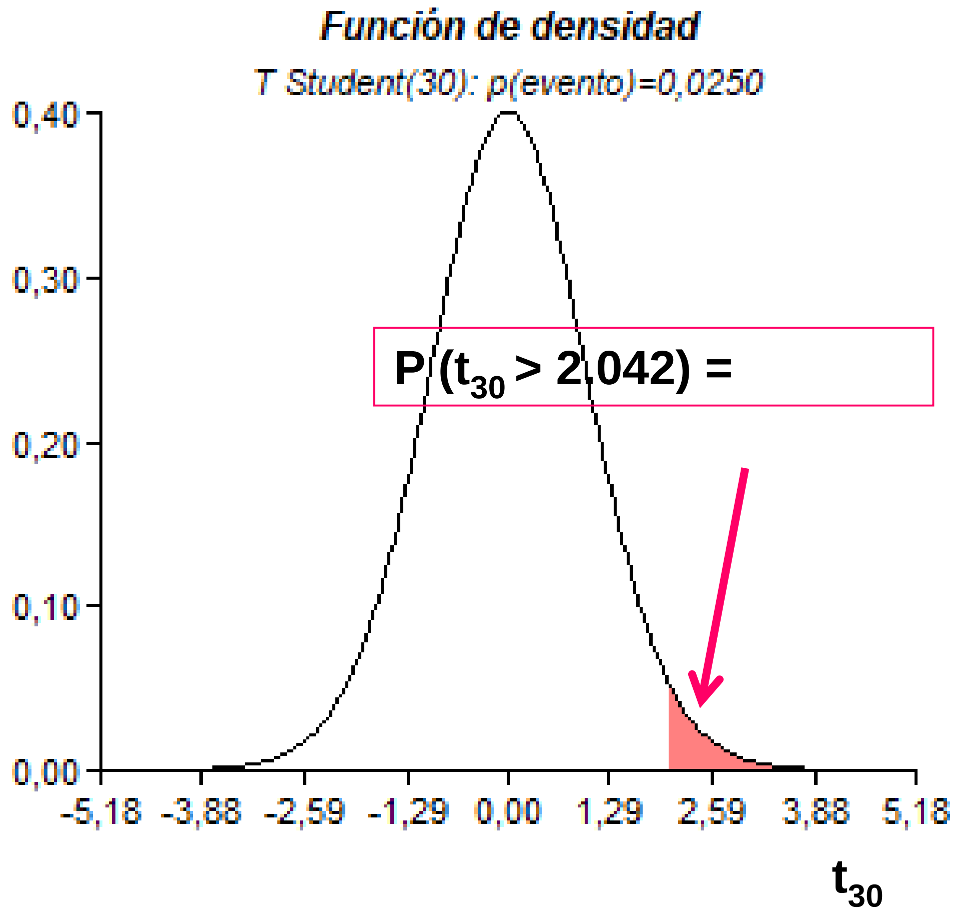
Manejo de la App t - student

I) Hallar:

(a) P (t_{30} > 2.042) ?

P (t_{30} > 2.042)

=



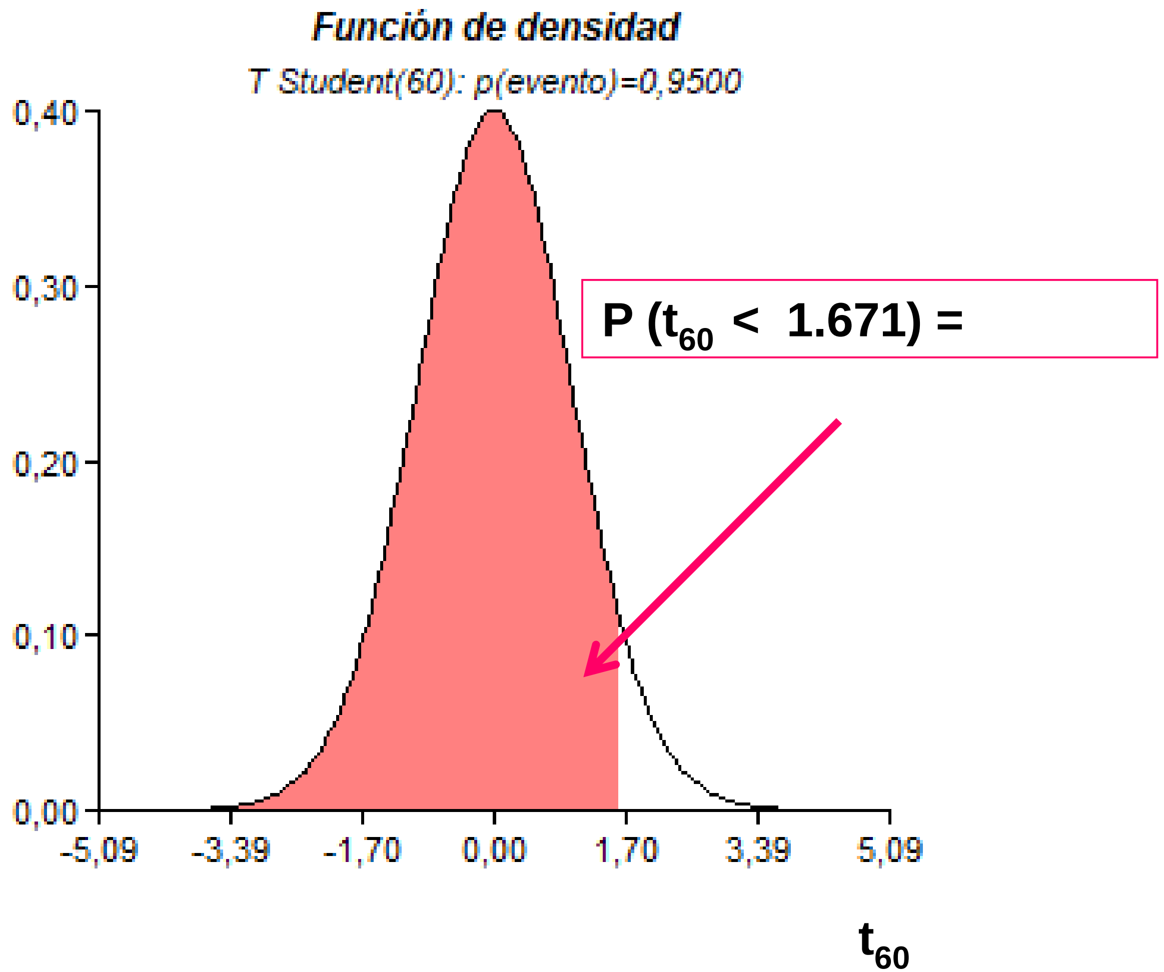
Manejo de la App t - student

I) Hallar:

(c) $P(t_{60} < 1.671)$

$P(t_{60} < 1.671) =$

$=$



$P(t_{60} < 1.671) =$

$=1 - P(t_{60} \geq 1.671) =$

$=$

II) Para la distribución t-student hallar el valor de $t_{\nu,\alpha}$ de tal forma que:

(b) $P(t_8 < t_{8,\alpha}) = 0.975$ $\cong t_{8,\alpha}$

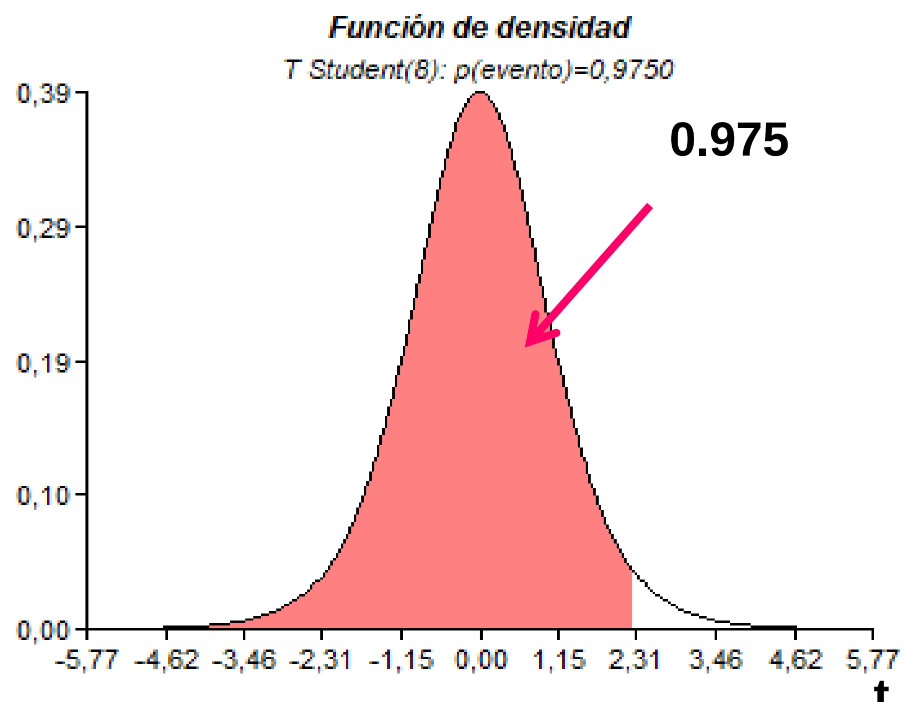
Otra forma de hacerlo:

$$0.975 = P(t_8 < t_{8,\alpha}) = 1 - P(t_8 \geq t_{8,\alpha})$$

$$1 - 0.975 = P(t_8 \geq t_{8,\alpha})$$

$$0.025 = P(t_8 \geq t_{8,\alpha})$$

$$\cong t_{8,\alpha}$$



Intervalo de confianza para la media de una población

Caso II: varianza σ^2 desconocida

Dado que X ~ N(μ,σ^2) con σ^2 desconocida, se requiere de un intervalo donde no figure σ^2. El valor de σ^2 puede estimarse mediante el estimador puntual varianza muestral S^2.

Como X ~ N(μ,σ^2), el estadístico

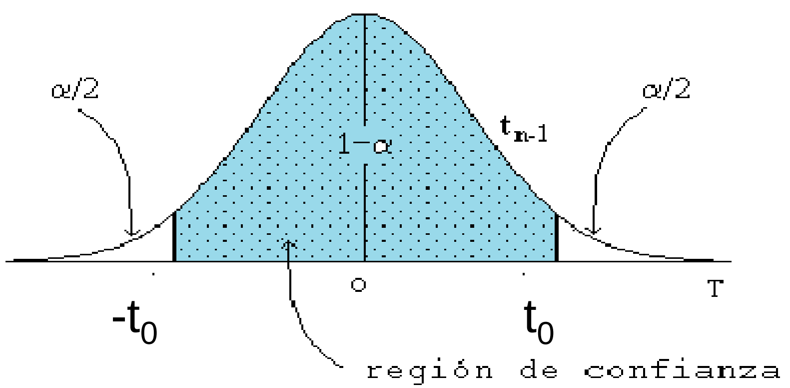
$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

(tiene una distribución t - Student con n - 1 grados de libertad)

Dado **1- α**, el nivel de confianza elegido para construir el intervalo, es posible encontrar t_0 tal que:

$$P(-t_0 < T < t_0) = 1 - \alpha$$

donde $t_0 = t_{\alpha/2}$ $\quad -t_0 = t_{1-\alpha/2}$



Intervalo de confianza para la media de una población

Si se reemplaza la v.a. T por su expresión, está igualdad resulta:

$$P\left(-t_0 < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_0\right) = 1-\alpha$$

Si se despeja el parámetro μ en la expresión dentro del paréntesis se obtiene:

$$P\left(\overline{X} - t_0 * S/\sqrt{n} < \mu < \overline{X} + t_0 * S/\sqrt{n}\right) = 1-\alpha$$

Por lo tanto, con una probabilidad **1 - α** el intervalo aleatorio

$$\left(\overline{X} - t_0 * S/\sqrt{n}, \overline{X} + t_0 * S/\sqrt{n}\right)$$

contendrá el parámetro μ.

Si se extrae una muestra de tamaño *n* y con las observaciones, x_1, x_2 ,..., x_n, se calculan los extremos del intervalo, se dispondrá del **intervalo de confianza**

$$(\overline{x} - t_0 * s/\sqrt{n}, \overline{x} + t_0 * s/\sqrt{n}) = \mathbf{(a, b)}$$

para el parámetro μ para un nivel de confianza de (1-α) 100 %.

Intervalo de confianza para la media de una población

Ejemplo 2 (Caso II)

El departamento de Atención al cliente de una empresa extranjera de cable desea estimar el tiempo promedio entre que se efectúa la solicitud del servicio y la conexión del mismo. Se consideraron las últimas 15 solicitudes de servicio de cable, y a continuación se detallan los tiempos (en días) transcurridos hasta la conexión del servicio (considerar estos valores como elementos de una muestra aleatoria):
El tiempo transcurrido hasta la conexión del servicio es una variable aleatoria con distribución normal.

11	**7**	**9**	**13**	**7**	**10**	**11**	**12**
7	**10**	**7**	**8**	**11**	**9**	**8**	

Estimar el verdadero tiempo promedio de conexión mediante un intervalo de confianza del 95%. Interpretar el resultado obtenido.

X =

X ~

Se desea estimar mediante un intervalo de confianza el parámetro μ:

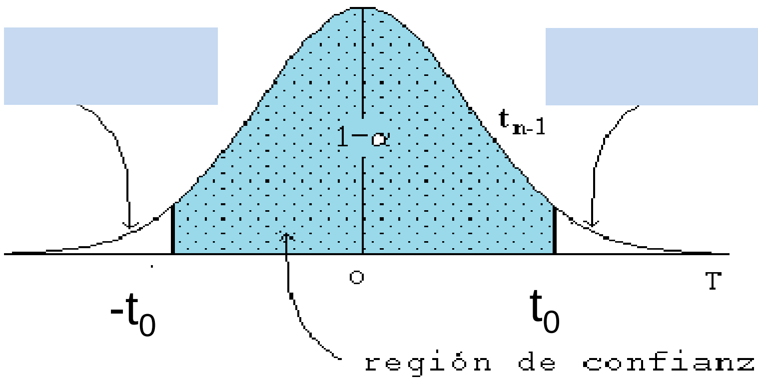
μ = tiempo promedio transcurrido desde que una persona efectúa la solicitud del servicio de cable hasta que la conexión se efectiviza”

Como el parámetro σ es desconocido (**uso caso II**)

Datos: n =, $\bar{x}$ = 9.33 días y s = 1.99 días

Nivel de confianza = 1 - α = entonces α = y **α/2** =

Intervalo a usar es: $(\bar{x} - t_0 * s/\sqrt{n}, \bar{x} + t_0 * s/\sqrt{n})$



$t_0 = t_{14,0.025} =$ $-t_0 = t_{14,0.975} =$

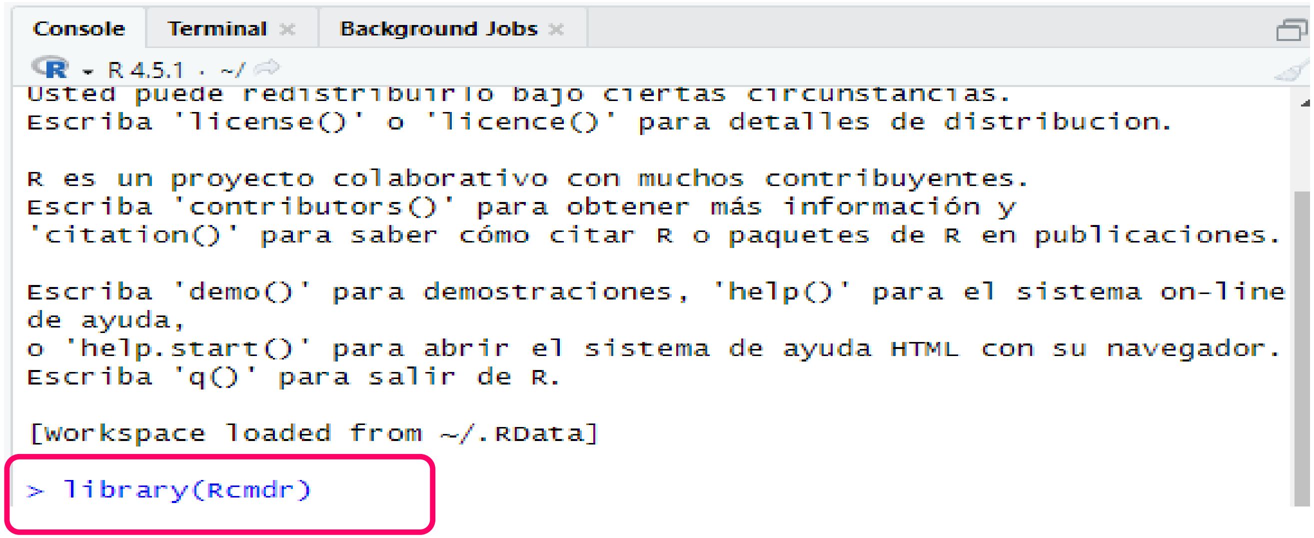
$$(\overline{x} - t_0 * s/\sqrt{n}, \overline{x} + t_0 * s/\sqrt{n})$$

Reemplazando los datos en:

$$(\quad , \quad) = \mathbf{(a, b)}$$

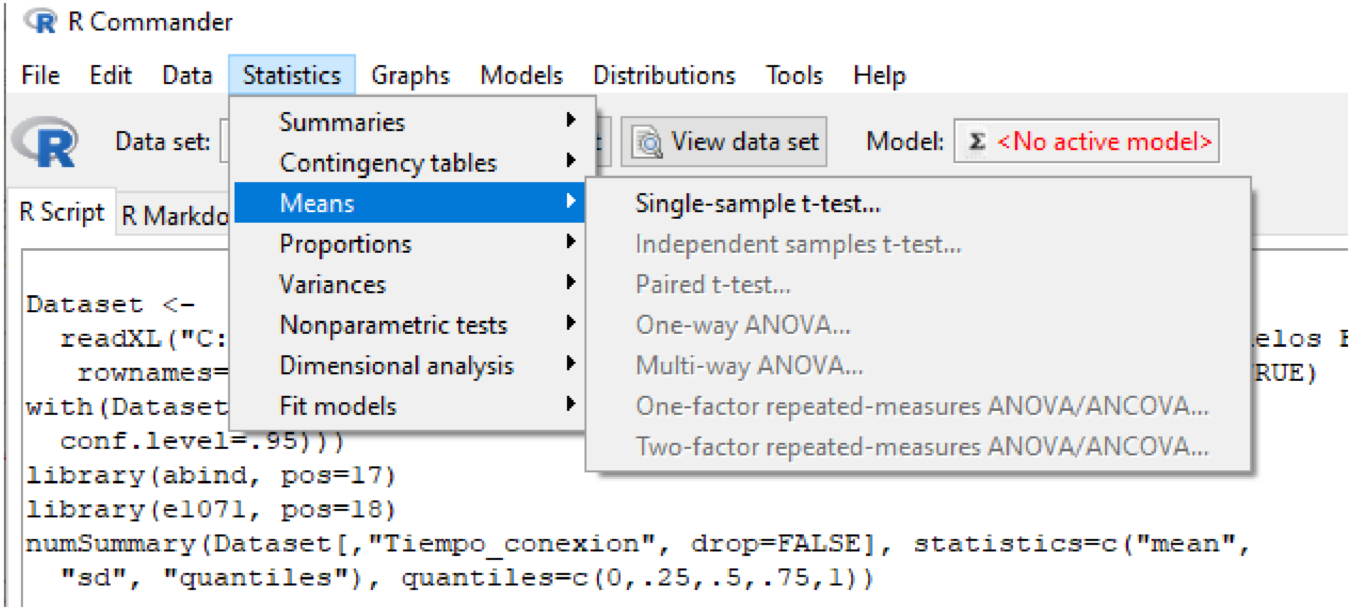
El tiempo promedio transcurrido entre que se efectúa la solicitud del servicio de cable y la conexión del mismo oscila entre , con un porcentaje de confianza del .

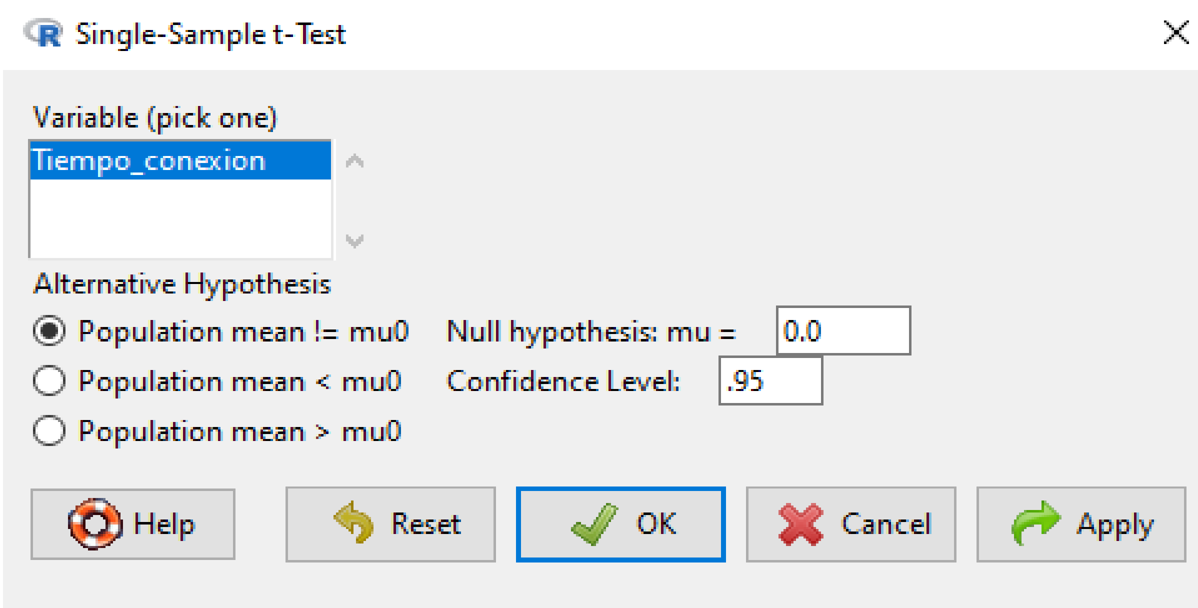
Rstudio: Rcmdr



Cargo la librería “Rcmdr”, una vez instalado el package “Rcmdr”

Clickeo “Data” para importar la base del Excel.


R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set:
View data set
Model:
<No active model>
R Script
R Markdo
Summaries
Contingency tables
Means
Proportions
Variances
Nonparametric tests
Dimensional analysis
Fit models
Single-sample t-test...
Independent samples t-test...
Paired t-test...
One-way ANOVA...
Multi-way ANOVA...
One-factor repeated-measures ANOVA/ANCOVA...
Two-factor repeated-measures ANOVA/ANCOVA...
Dataset <-
readXL("C:
rownames=
with(Dataset
conf.level=.95)))
library(abind, pos=17)
library(e1071, pos=18)
numSummary(Dataset[,"Tiempo_conexion", drop=FALSE], statistics=c("mean",
"sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))


Single-Sample t-Test
Variable (pick one)
Tiempo_conexion
Alternative Hypothesis
Population mean != mu0
Population mean < mu0
Population mean > mu0
Null hypothesis: mu =
0.0
Confidence Level:
.95
Help
Reset
OK
Cancel
Apply

Intervalo de confianza para la media de una población con desvío desconocido

```
        One Sample t-test

data:  Tiempo_conexion
t = 18.182, df = 14, p-value = 3.896e-11
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  8.232383 10.434284
sample estimates:
mean of x
 9.333333
```

Extremos del intervalo de confianza

Estimación puntual de la media

Medidas resumen:

media y desvío estándar muestrales

```
Rcmdr>  numSummary(Dataset[,"Tiempo_conexion", drop=FALSE], statistics
=c("mean",
Rcmdr+      "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
     mean      sd 0% 25% 50% 75% 100%  n
 9.333333 1.98806  7 7.5   9  11   13 15
```

Ejercicio

El administrador de una planta industrial generadora de energía desea estimar la cantidad promedio de carbón consumida semanalmente durante el año pasado. Para ello tomó una muestra de 10 semanas, obteniendo los siguientes valores de consumo:

10300	11250	11010	11100	11450
10980	10850	10870	11800	10990

Suponer que el consumo semanal de carbón es una v.a. normalmente distribuida.

Estimar mediante un intervalo de confianza del 95% el consumo medio semanal de carbón durante el año pasado

Intervalo de confianza para la media de una población

Observación importante!!!!!

Los intervalos presentados anteriormente se construyeron sobre la base de una v.a. **X ~ N(μ,σ^2)** con varianza **σ^2 conocida o desconocida**.

- Si la v.a. X **no está normalmente** distribuida pero **σ^2** **conocida** y el tamaño de la muestra es ***n* ≥ 30**, el **teorema central del límite garantiza** que la v.a. media muestral tiene distribución normal y se estandariza: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z$

y por lo tanto el intervalo de confianza correspondiente a esta situación sería el presentado en el **caso I**:

$$(\bar{x} - z_0 * \sigma/\sqrt{n}, \bar{x} + z_0 * \sigma/\sqrt{n}) = \mathbf{(a, b)}$$

- Si la v.a. X **no está normalmente** distribuida pero **σ^2** **desconocida** y el tamaño de la muestra es ***n* ≥ 30**, la varianza muestral s^2 se acercará al verdadero valor de **σ^2** y entonces sigue siendo aplicable el teorema central de límite proporcionando el siguiente intervalo:

$$(\bar{x} - z_0 * s/\sqrt{n}, \bar{x} + z_0 * s/\sqrt{n}) = \mathbf{(a, b)}$$

Este intervalo se conoce como **intervalo de confianza de muestra grande**.

Ejemplo

Un investigador del área de arquitectura de computadoras está evaluando el rendimiento de un nuevo sistema de procesamiento paralelo con 8 núcleos físicos. Para ello, diseña una tarea computacional que simula cálculos matemáticos complejos y la ejecuta múltiples veces para medir el tiempo necesario hasta su finalización. Se realizan 80 ejecuciones independientes del proceso, todas bajo las mismas condiciones de hardware y sin otras cargas activas en el sistema. Estas arrojaron un tiempo promedio de 14.8 segundos y un desvío estándar de 1.6 segundos.

a) Estimar, con un nivel de confianza del 99%, el tiempo promedio de procesamiento de dicha tarea en el sistema paralelo. **Interpretar el resultado**.

b) Hallar un intervalo de confianza del 90% el tiempo promedio de procesamiento de dicha tarea en el sistema paralelo. **Interpretar el resultado.**

c) Comparar los resultados obtenidos en a) y b).

a) **Estimar, con un nivel de confianza del 99%, el tiempo promedio de procesamiento de dicha tarea en el sistema paralelo. Interpretar el resultado.**

X =   **X ~**

Se desea estimar mediante un intervalo de confianza el parámetro μ:

μ **=**

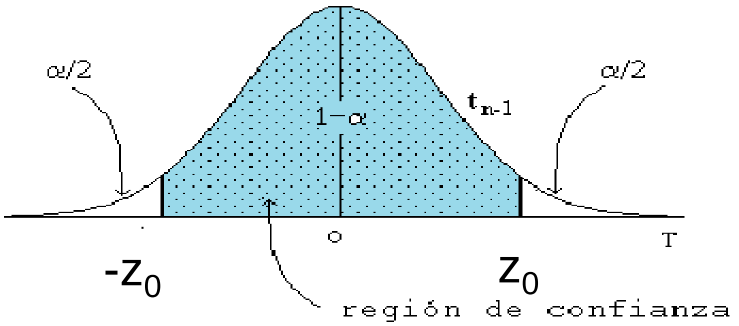
Como el parámetro σ es desconocido

Datos:

n = , $\overline{x}$ = y s =

Nivel de confianza = 1 - α = entonces α = y α**/2** =

Intervalo a usar es: $\left(\overline{x} - z_0 * \frac{s}{\sqrt{n}}, \overline{x} + z_0 * \frac{s}{\sqrt{n}}\right)$



$z_{\alpha/2} = -z_0 =$ $z_{,1-\alpha/2} = z_0 =$

$$\left(\bar{x} - z_0 * {}^{s}\!/\!\sqrt{n}\,,\ \bar{x} + z_0 * {}^{s}\!/\!\sqrt{n}\right)$$

Reemplazando los datos en:

$$(\qquad , \qquad) = \mathbf{(a,b)}$$

el tiempo promedio de procesamiento de dicha tarea en el sistema paralelo oscila entre

con un porcentaje de confianza del

Resumen: IC para estimar E(X) =μ

Si X ~ N(μ, σ²)

σ² es conocido $\overline{X} \sim N(\mu, \sigma^2/n)$

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

$$\left(\overline{x} - z_0 * \sigma/\sqrt{n},\ \overline{x} + z_0 * \sigma/\sqrt{n}\right) \quad \textbf{(1)}$$

σ² es desconocido $\overline{X} \sim N(\mu, \sigma^2/n\ \ ?)$

$$T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

$$\left(\overline{x} - t_0 * S/\sqrt{n},\ \overline{x} + t_0 * S/\sqrt{n}\right)$$

Si X ~ ????

n ≥ 30, por el Teo. Central de L.

σ² es conocido →

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

El **IC** es el **(1)**

σ² es desconocido →

$$Z \cong \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim N(0,1)$$

$$\left(\overline{x} - z_0 * S/\sqrt{n},\ \overline{x} + z_0 * S/\sqrt{n}\right)$$