Estadística Inferencial

Pamela E. Pairo

En la clase de hoy:

- Concepto de inferencia. Población y muestra.
- Parámetros y estimadores
- Distribuciones Muestrales
- Intervalos de confianza para la media con varianza conocida
- Ejercitación

Recreo ☕ 🧉

- Distribución t-student
- Intervalos de confianza para la media con varianza desconocida
- Ejercitación

Estadística Inferencial

Permite estimar parámetros poblacionales

Probar hipótesis formuladas por sobre una población

Construir modelos estadísticos y efectuar predicciones

Recordar:

Población: es el conjunto de todos los individuos de interés. Normalmente es demasiado grande para poder abarcarla toda = censo

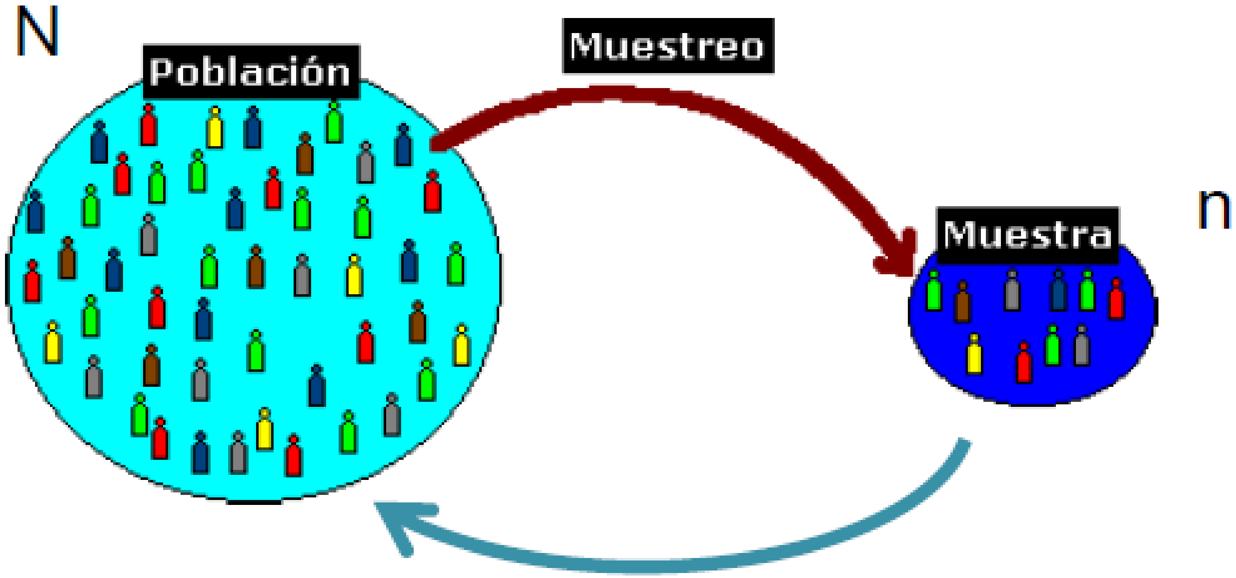
Muestra: es un subconjunto representativo de la población y es sobre el que realmente hacemos las observaciones

Variable: es la característica de interés que es medida en cada uno de los individuos

Observación o dato: es el valor particular que toma la variable en cada individuo

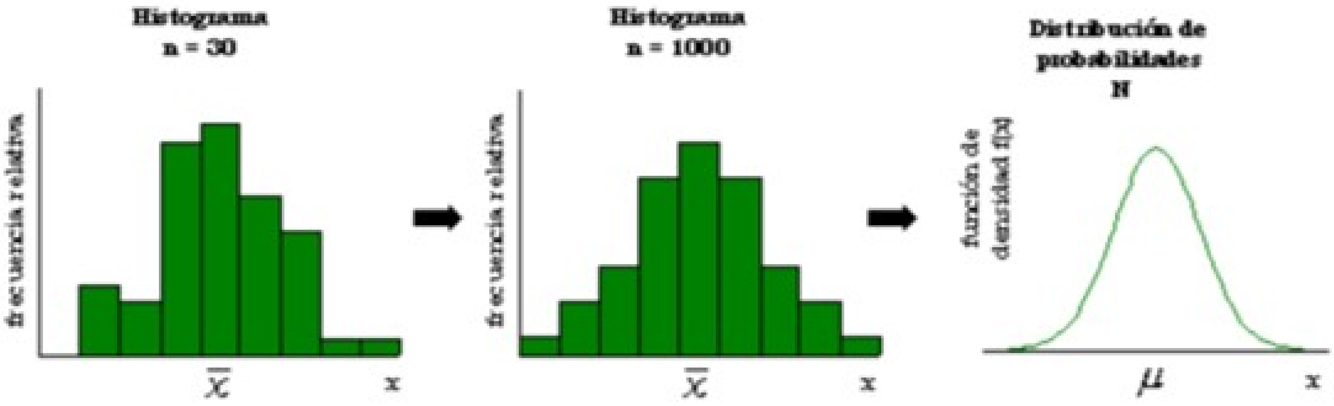
Unidad Experimental: es la menor unidad de la cual se obtiene una observación independiente

La inferencia estadística consiste en generalizar las conclusiones extraídas de una muestra sobra una población



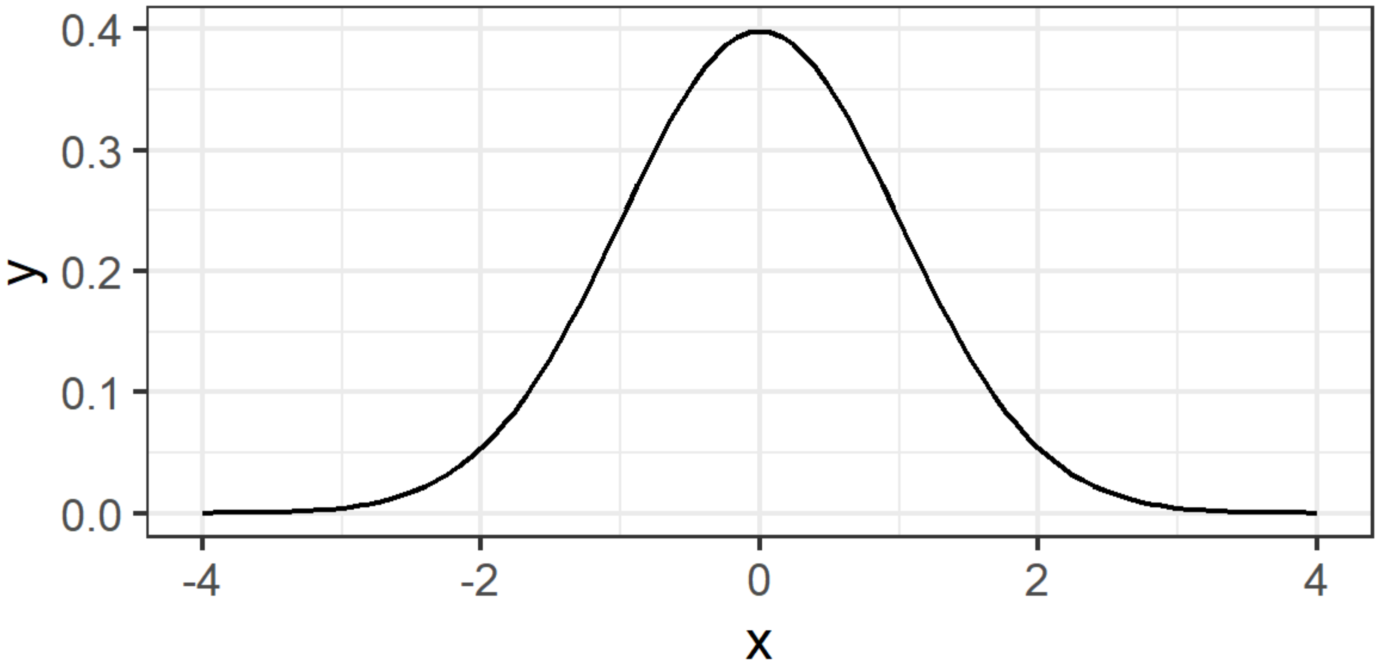
Distribución de Probabilidades

Función de densidad f(x) es una función que describe la distribución de probabilidades de la variable aleatoria continua x.



Distribución Normal

$$X \sim \mathsf{N}(\mu, \sigma^2)$$



- Media, mediana y moda coinciden.
- Dominio de la variable -∞< x < ∞

Estandarización

- La variable de estudio es tranformada (reescalada) en una variable normal estándar.

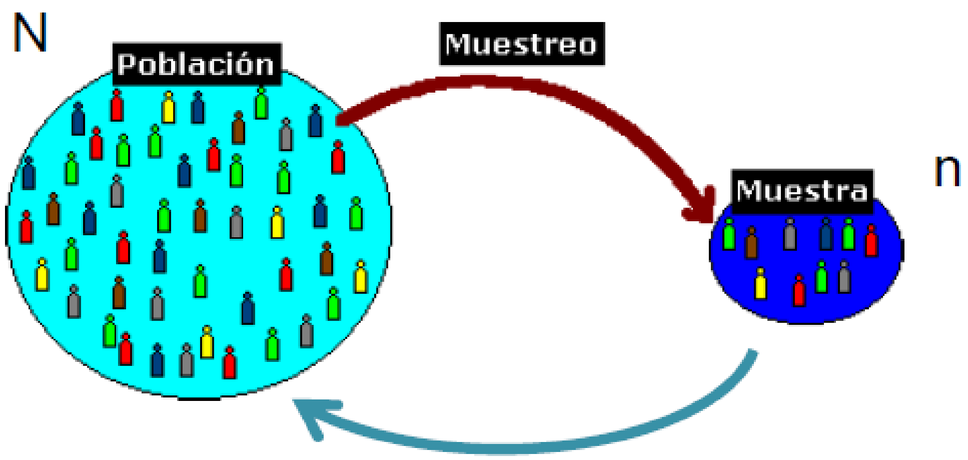
$$z = \frac{x-\mu}{\sigma}$$

- Cuanto mas grande sea el valor de Z, mas lejos estará el valor de la media.

Parámetro y Estimador

Parámetro: se calculan sobre los N valores de la población, por lo tanto no cambian a menos que cambie la población, son constantes

Estimador: se calculan sobra n valores muestrales, por lo tanto varían de muestra en muestra y por lo tanto son variables aleatorias.



Supongamos...

POBLACIÓN

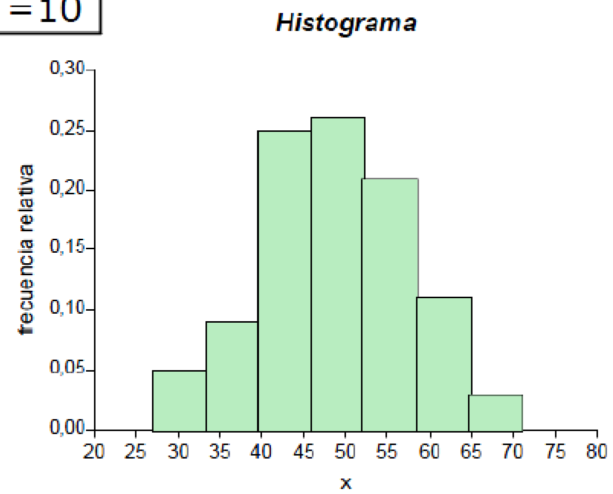
41	44	34	42	53
54	41	61	44	44
52	59	36	57	61
60	53	54	43	48
57	43	51	51	52
32	45	55	36	47
49	42	38	71	46
54	27	55	45	42
46	45	58	53	43
42	54	44	39	49
62	54	36	61	59
57	43	63	47	49
32	56	44	44	53
49	45	52	58	59
42	67	58	32	39
48	37	49	47	64
52	33	35	51	41
45	47	46	55	42
43	50	61	47	67
57	49	57	52	47

Población 100 individuos

Media de la población es 50

La variabilidad es 10

PROMEDIO μ =50
DESVÍO STD σ =10



Y si se repite el procedimiento muchas veces?

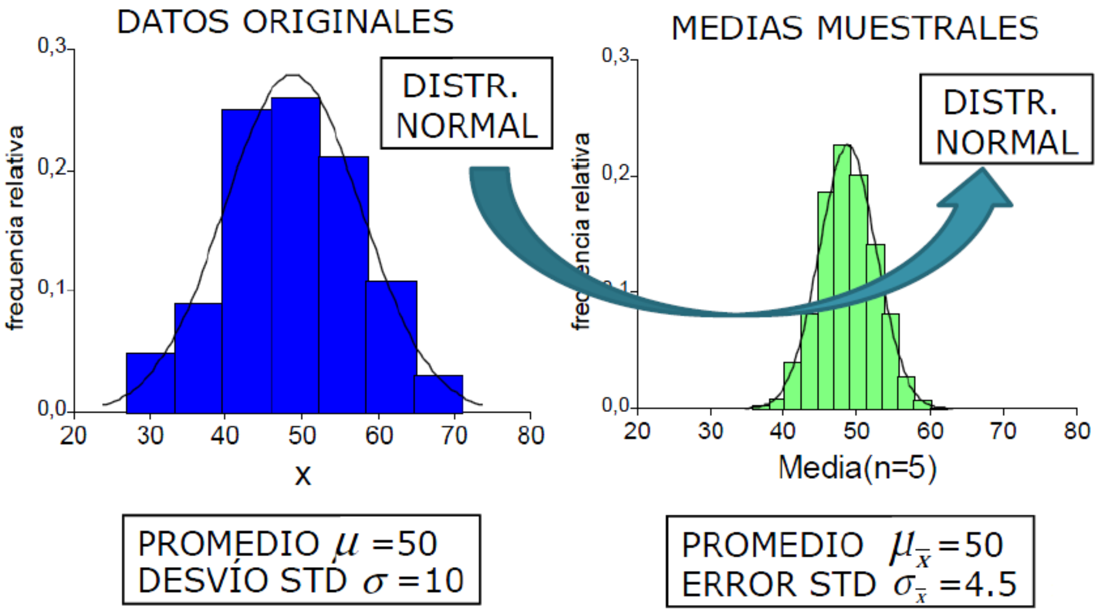
POBLACIÓN

41	44	34	42	53
54	41	61	44	44
52	59	36	57	61
60	53	54	43	48
57	43	51	51	52
32	45	55	36	47
49	42	38	71	46
54	27	55	45	42
46	45	58	53	43
42	54	44	39	49
62	54	36	61	59
57	43	63	47	49
32	56	44	44	53
49	45	52	58	59
42	67	58	32	39
48	37	49	47	64
52	33	35	51	41
45	47	46	55	42
43	50	61	47	67
57	49	57	52	47

MUESTRAS n = 5

44	52	47	33	42	$\overline{X}_1$
61	45	38	67	51	$\overline{X}_2$
51	54	50	33	71	$\overline{X}_3$
. . .					. . .
41	58	49	34	49	$\overline{X}_{75287520}$

Distribución muestral de $\overline{x}$



$$\mu_{\overline{x}} = \mu$$

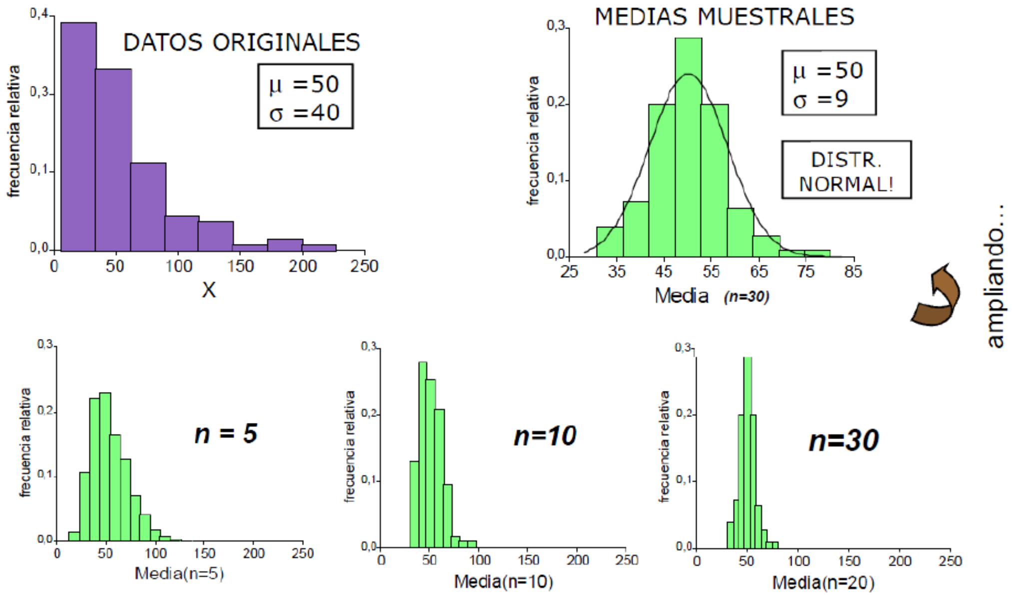
$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Distribución muestral

La distribución muestral de un estimador es la distribución de probabilidades de todos los posibles valores de un estimador que se pueden obtener extrayendo infinitas muestras aleatorias de tamaño n de la población.

Teorema central del límite

Si de una población con distribución no normal o desconocida con media μ y desvío estándar σ se extraen infinitas muestras aleatorias de tamaño n y a cada una se ellas se le calcula el promedio X se demuestra que éste se comporta según una distribución normal si n es lo suficientemente grande.



Estimación puntual y por IC

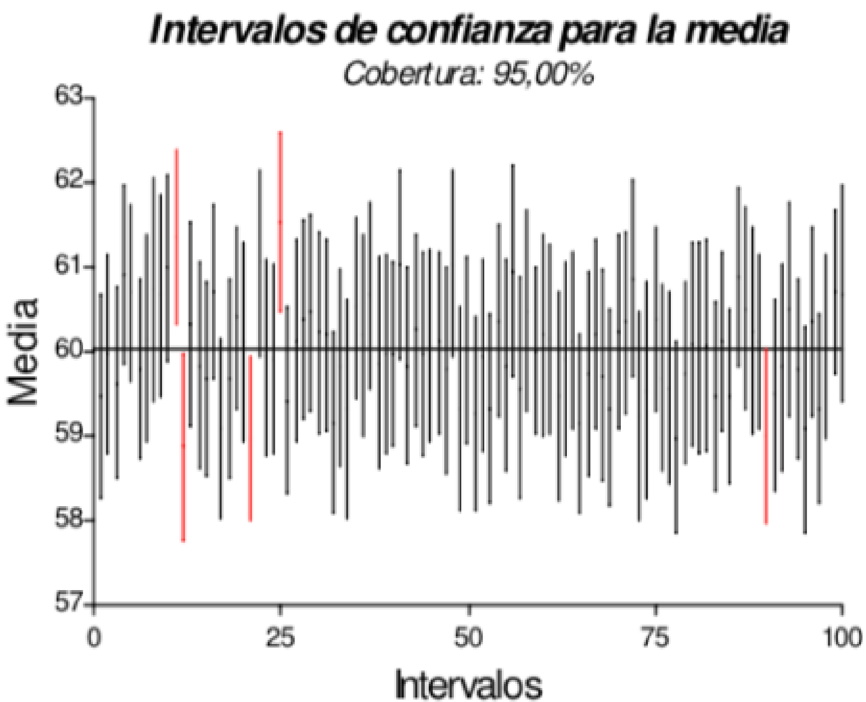
Estimación Puntual: se calcula un valor a partir de la muestra a fin de estimar el parámetro.

Intervalo de Confianza: se calculan dos números para crear un rango de valores que se espera que contengan al parámetro con una cierta probabilidad o nivel de confianza.

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

Intervalo de confianza: Nivel de confianza

- Probabilidad de que el intervalo contenga al parámetro (1- α)
- Se fija a priori (0.9, 0.95, 0.99)
- α es la probabilidad de error (NO contener al parámetro) y se la denomina riesgo



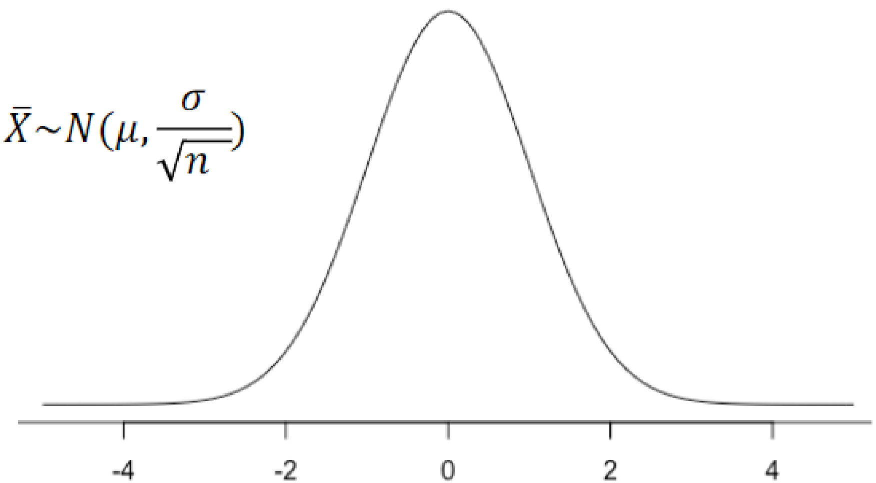
IC para la media poblacional de una v.a. normal con varianza poblacional conocida

1- ¿Qué parámetro vamos a estimar? En este caso μ

2- Estimador con distribución conocida

3- Definir un estadñistico de prueba con distribución conocida

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$



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

IC para la media poblacional de una v.a. normal con varianza poblacional conocida

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

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

$$P\left(\boxed{\bar{X} - Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}} < \mu < \boxed{\bar{X} + Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

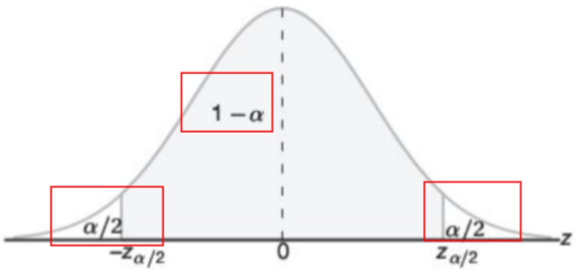


Figura 9.2 (Walpole 2012)

$$\overline{x} \pm Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

Ejercitación

Ejercicio 1- La Cámara de Comercio desea tener una estimación del valor medio del alquiler de un departamento de un ambiente. Una muestra aleatoria de 40 departamentos de estas características arrojó como resultado una media mensual de los alquileres (de esta muestra) es de $323. Se conoce en el mercado que la desviación estándar de la variable “alquiler mensual de un depto. de 1 amb” es $25.

a) Determine un intervalo de confianza de 99% para la media del alquiler de la población total de departamentos de un ambiente.

¿cuál es la variable x?¿Distribución?

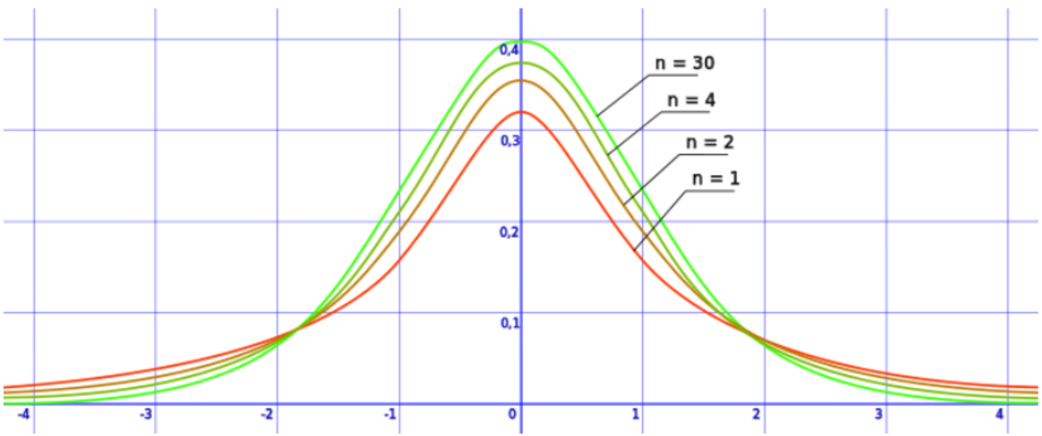
Ejercitacion

Ejercicio 2- La Asociación de Empresarios Gastronómicos reunió información sobre la cantidad de veces que los matrimonios jóvenes comen fuera de casa a la semana. Una encuesta de 60 parejas indicó que la cantidad media de veces que fuera de casa por semana fue 2.76 comidas semanales. Mediante estudios previos se pude afirmar que la desviación standard poblacional de la variable “cant de veces por semana que las parejas jóvenes comen afuera” es de 0.75. Construya el intervalo de confianza de 90 % de la media poblacional.

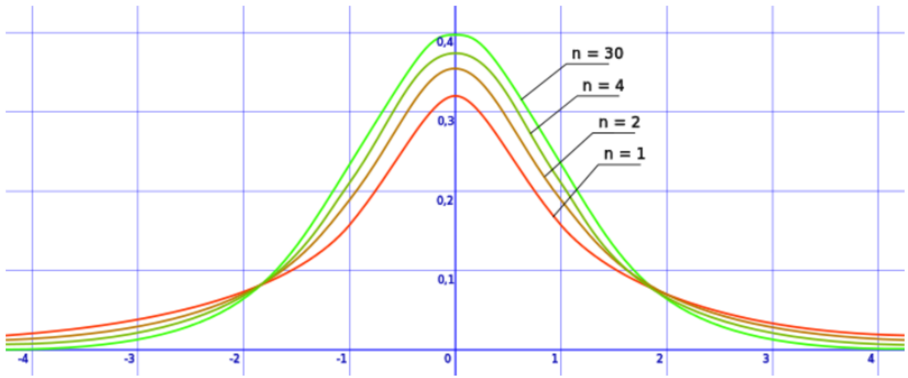
IC para la media poblacional de una v.a. normal con varianza poblacional desconocida

En la práctica es habitual que TODOS los parametros poblacionales son desconocidos, es decir que ni el promedio μ ni el desvío estándar σ son conocidos.

Se demuestra que la media muestral en estos casos ajusta a una distribución conocida como **t de Student**



Distribución de t-student

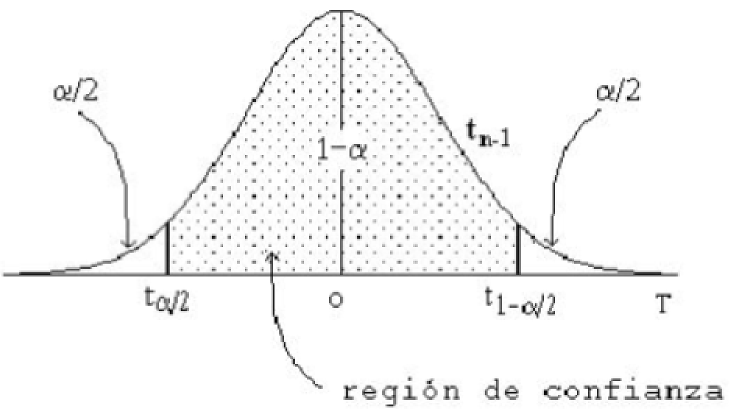


Los **GL indican la cantidad de datos independientes**, es decir el numero de observaciones de la varibal menos el número de restricciones que verifican

A medida que aumentan los GL mas se asemejan a la normal estándar (porque s converge a σ)

IC para la media poblacional de una v.a. normal con varianza poblacional desconocida

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



$$P\left(\bar{X} + t_{n-1;\,\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1;1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - t_{n-1;\,1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1;1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right) \approx 1 - \alpha$$

IC para la media poblacional de una v.a. normal con varianza poblacional desconocida

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

$$\overline{x} \pm t_{(n-1;1-\frac{\alpha}{2})} * \frac{s}{\sqrt{n}}$$

Ejercitación

Las ventas de una revista semanal han sido las siguientes (en miles) en las últimas cuatro semanas: 15,4 -18,5 - 16,3 - 19,2.

a) Estimar la venta media semanal con un 95% de con